Roadmap on deep learning for microscopy

ROADMAP • **OPEN ACCESS**

Roadmap on deep learning for microscopy

Journal of Physics: Photonics

OPEN ACCESS

RECEIVED
4 May 2023

REVISED
29 April 2025

ACCEPTED FOR PUBLICATION
2 October 2025

PUBLISHED
29 January 2026

ROADMAP

Roadmap on deep learning for microscopy

Giovanni Volpe[1,2,*], Carolina Wählby[3,*], Lei Tian[4,5,*], Michael Hecht[6,7,8], Artur Yakimovich[6,7,9,10,11,12], Kristina Monakhova[13], Laura Waller[13], Ivo F Sbalzarini[14,15], Christopher A Metzler[16], Mingyang Xie[16], Kevin Zhang[16], Isaac C D Lenton[17,18], Halina Rubinsztein-Dunlop[18,19,20], Daniel Brunner[21], Bijie Bai[22], Aydogan Ozcan[22], Daniel Midtvedt[1], Hao Wang[4], Tongyu Li[4], Nataša Sladoje[23], Joakim Lindblad[23], Jason T Smith[24,25], Marien Ochoa[23,26], Margarida Barroso[27], Xavier Intes[24], Tong Qiu[28], Li-Yu Yu[28], Sixian You[28], Yongtao Liu[29], Maxim A Ziatdinov[29,30], Sergei V Kalinin[31], Arlo Sheridan[32], Uri Manor[32,33,34], Elias Nehme[35,36], Ofri Goldenberg[36], Yoav Shechtman[36], Henrik K Moberg[37], Christoph Langhammer[37], Barbora Špačková[38], Saga Helgadottir[3,39], Benjamin Midtvedt[1], Aykut Argun[1], Tobias Thalheim[40], Frank Cichos[40], Stefano Bo[41,42], Lars Hubatsch[43], Jesus Pineda[1], Carlo Manzo[44,45], Harshith Bachimanchi[1], Erik Selander[46], Antoni Homs-Corbera[47,48], Martin Fränzl[40], Kevin de Haan[22], Yair Rivenson[22], Zofia Korczak[1], Caroline Beck Adiels[1], Mite Mijalkov[49], Dániel Veréb[49], Yu-Wei Chang[1], Joana B Pereira[49], Damian Matuszewski[23], Gustaf Kylberg[50], Ida-Maria Sintorn[23,50], Juan C Caicedo[51], Beth A Cimini[52], Muyinatu A Lediju Bell[53], Bruno M Saraiva[54], Guillaume Jacquemet[55,56,57,58], Ricardo Henriques[54,59], Wei Ouyang[60], Trang Le[61], Estibaliz Gómez-de-Mariscal[54,62,63], Daniel Sage[64], Arrate Muñoz-Barrutia[65,66], Ebba Josefson Lindqvist[67] and Johanna Bergman[67]

[1] Department of Physics, University of Gothenburg, 41296 Gothenburg, Sweden
[2] Science for Life Laboratory, Department of Physics, University of Gothenburg, 41296 Gothenburg, Sweden
[3] Department IT and SciLifeLab, Uppsala University, Uppsala, Sweden
[4] Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, United States of America
[5] Department of Biomedical Engineering, Boston University, Boston, MA 02215, United States of America
[6] Center for Advanced Systems Understanding (CASUS), Görlitz, Germany
[7] Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR), Dresden, Germany
[8] Mathematical Institute, University Wrocław, Wrocław, Poland
[9] Institute of Computer Science, University of Wrocław, Wrocław, Poland
[10] Bladder Infection and Immunity Group (BIIG), Department of Renal Medicine, Division of Medicine, University College London, Royal Free Hospital Campus, London, United Kingdom
[11] Artificial Intelligence for Life Sciences CIC, Dorset, United Kingdom
[12] Roche Pharma International Informatics, Roche Diagnostics GmbH, Mannheim, Germany
[13] Department of Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA, United States of America
[14] Technische Universität Dresden, Faculty of Computer Science, Dresden, Germany
[15] Center for Systems Biology Dresden, Dresden, Germany
[16] University of Maryland, College Park, United States of America
[17] Institute of Science and Technology Austria (ISTA), Klosterneuburg 3400, Austria
[18] School of Mathematics and Physics, The University of Queensland, Brisbane 4072, Australia
[19] Australian Research Council Centre of Excellence for Engineered Quantum Systems, The University of Queensland, Brisbane 4072, Australia
[20] ARC CoE in Quantum Biotechnology, Australia
[21] Département d'Optique P. M. Duffieux, Institut FEMTO-ST, Université Franche-Comté CNRS UMR 6174, Besançon, France
[22] Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, United States of America
[23] Department of Information Technology, Uppsala University, Uppsala, Sweden
[24] Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, United States of America
[25] ECS Federal, Fairfax, VA 22031, United States of America
[26] Department of Medical Physics, University of Wisconsin–Madison, Madison, WI 53705, United States of America
[27] Department of Molecular and Cellular Physiology, Albany Medical College, Albany, NY 12208, United States of America
[28] Research Laboratory of Electronics, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United States of America
[29] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States of America
[30] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States of America
[31] Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37996, United States of America
[32] Salk Institute for Biological Studies, La Jolla, CA, United States of America
[33] Department of Cell & Developmental Biology, University of California, San Diego, La Jolla, CA, United States of America
[34] Halıcıoğlu Data Science Institute, University of California, San Diego, CA 92093, United States of America
[35] Department of Electrical and Computer Engineering, Technion—IIT, Haifa, Israel
[36] Department of Biomedical Engineering and Lorry I. Lokey Interdisciplinary Center for Life Sciences & Engineering, Technion—IIT, Haifa, Israel

[37] Department of Physics, Chalmers University of Technology, Gothenburg, Sweden
[38] FZU, Institute of Physics of the Czech Academy of Sciences, Prague, Czechia
[39] Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden
[40] Peter Debye Institute for Soft Matter Physics, Molecular Nanophotonics Group, Leipzig University, Linnéstraße 5, 04103 Leipzig, Germany
[41] Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Straße 38, DE-01187 Dresden, Germany
[42] Department of Physics, King's College London, London WC2R 2LS, United Kingdom
[43] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany
[44] Facultat de Ciències, Tecnologia i Enginyeries, Universitat de Vic—Universitat Central de Catalunya (UVic-UCC), Vic, Spain
[45] Bioinformatics and Bioimaging, Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), Vic, Spain
[46] Department of Biology, Lund University, Lund, Sweden
[47] Research and Development Department, Cherry Biotech SAS, Paris, France
[48] OWL Lifesciences, Paris, France
[49] Department of Clinical Neuroscience, Division of Neuro, Karolinska Institutet, Stockholm, Sweden
[50] VINNOVA Sweden's Innovation Agency, Stockholm, Sweden
[51] Morgridge Institute for Research, University of Wisconsin–Madison, Madison, WI, United States of America
[52] Broad Institute of MIT and Harvard, Boston Cambridge, MA, United States of America
[53] Departments of Electrical & Computer Engineering, Biomedical Engineering, and Computer Science, Johns Hopkins University, Baltimore, MD 21218, United States of America
[54] Instituto Gulbenkian de Ciência, Oeiras, Portugal
[55] Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland
[56] Faculty of Science and Engineering, Cell Biology, Åbo Akademi University, Turku, Finland
[57] Turku Bioimaging, University of Turku and Åbo Akademi University, FI- 20520 Turku, Finland
[58] InFLAMES Research Flagship Center, Åbo Akademi University, FI- 20520 Turku, Finland
[59] MRC-Laboratory for Molecular Cell Biology, University College London, London, United Kingdom
[60] Science for Life Laboratory, Department of Applied Physics, KTH—Royal Institute of Technology, 171 65 Stockholm, Sweden
[61] Department of Bioengineering, Stanford University, Stanford, CA 94305, United States of America
[62] Gulbenkian Institute for Molecular Medicine, Oeiras, Portugal
[63] Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
[64] Biomedical Imaging Group and EPFL Center for Imaging, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne Switzerland
[65] Bioengineering Department, Universidad Carlos III de Madrid, Leganés, Spain
[66] Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[67] AI Sweden, Gothenburg, Sweden
* Authors to whom any correspondence should be addressed.

E-mail: giovanni.volpe@physics.gu.se, carolina.wahlby@it.uu.se and leitian@bu.edu

Keywords: deep learning, microscopy, imaging, AI

Abstract

Through digital imaging, microscopy has evolved from primarily being a means for visual observation of life at the micro- and nano-scale, to a quantitative tool with ever-increasing resolution and throughput. Artificial intelligence, deep neural networks, and machine learning (ML) are all niche terms describing computational methods that have gained a pivotal role in microscopy-based research over the past decade. This Roadmap encompasses key aspects of how ML is applied to microscopy image data, with the aim of gaining scientific knowledge by improved image quality, automated detection, segmentation, classification and tracking of objects, and efficient merging of information from multiple imaging modalities. We aim to give the reader an overview of the key developments and an understanding of possibilities and limitations of ML for microscopy. It will be of interest to a wide cross-disciplinary audience in the physical sciences and life sciences.

Contents

1. Introduction

The first attempts to automate the analysis of microscopy data preceded the availability of personal computers and computer screens. In 1956, an instrument for the automatic screening of cytological smears for cancer was presented [1]. This system was based on hard-wired analogue video processing circuits, aiming to quantify the size of cell nuclei to find precancerous lesions. The more widespread availability of computers in the 1960s and 70s made possible more ambitious and complex research projects, e.g. aimed at developing image analysis systems with a range of biomedical applications. There were however two major hurdles: On the one hand, even very small images of less than 100 kb would typically take up the whole working memory of a research computer of the early 1970s. On the other hand, no screens for viewing image data were available, making method development difficult. Image processing, where useful images were created as a result of computations, saw its first revolutionary success with the invention of computer tomography in the 1970s [2], and the advent of the IBM personal computer in 1981, with screens capable to display 640 × 200 binary monochrome graphics and up to four colors at 320 × 200 resolution. This sparked the mass-market gaming industry, which has been one of the major driving forces for the further development of computer and graphics power.

Automation of microscopy, including sample handling and microscope control, enables rapid collection of digital image data making microscopy one of the most data-rich scientific disciplines. Ideas about automated interpretation of image data using machine vision and artificial intelligence (AI) have been around since the 1970s, but it is not until recent years that increasing computing power and large amounts of annotated images of natural scenes are finally making these methods work in practice. We now see the fast emergence of approaches to image analysis where the computer learns the task at hand from examples and automatically exploits the input images for measurements or decisions. This is reflected by the rapid increase in the scientific community of methods combining microscopy with deep learning (DL) (figure 1).

This Roadmap article aims to provide a concise yet authoritative overview on the present and future of how machine learning (ML) can and will be used for microscopy-based scientific research, spanning from the development of tools and algorithms (sections 1–6), via enhanced microscopy techniques (sections 7–13), single molecule detection and tracking (sections 14–22), and biomedical applications (sections 23–31), to the establishment of software platforms and community resources (sections 32–36).

Overview

The thirty-six expert contributions of this Roadmap are arranged in five thematic blocks that map the current landscape of ML-enabled microscopy:

Tools and algorithms (sections 1–6)

This part discusses fundamental advances that underpin the field: the mathematical stability of inverse problems (section 1), physics-based learning for joint optics–algorithm design (section 2), interpretable models for large 3D data (section 3), plug-and-play (PnP) reconstruction frameworks (section 4), ML acceleration of optical-force simulations (section 5), and optical neural-network hardware for all-optical computing and imaging (section 6).

Enhanced microscopy (sections 7–13)

These chapters show how ML pushes the spatiotemporal and functional limits of imaging modalities: quantitative microscopy (section 7), computational phase microscopy (section 8), multimodal image registration (section 9), fluorescence-lifetime imaging (section 10), multi-modal nonlinear microscopy (section 11), automated scanning-probe microscopy (section 12), and DL-based restoration for scanning systems (section 13).

Single molecules and particle dynamics (sections 14–22)

This part presents cutting-edge methods for localizing, tracking and interpreting nano- to microscale entities: single-molecule localization (section 14), nanofluidic scattering microscopy (NSM) (section 15), particle tracking (section 16), single-shot self-supervised object detection (section 17), force-field calibration (section 18), diffusion characterization (section 19), motion analysis with MAGIK (section 20), quantifying sub-cellular dynamics from single-particle tracks (section 21), and plankton life-trajectory analysis (section 22).

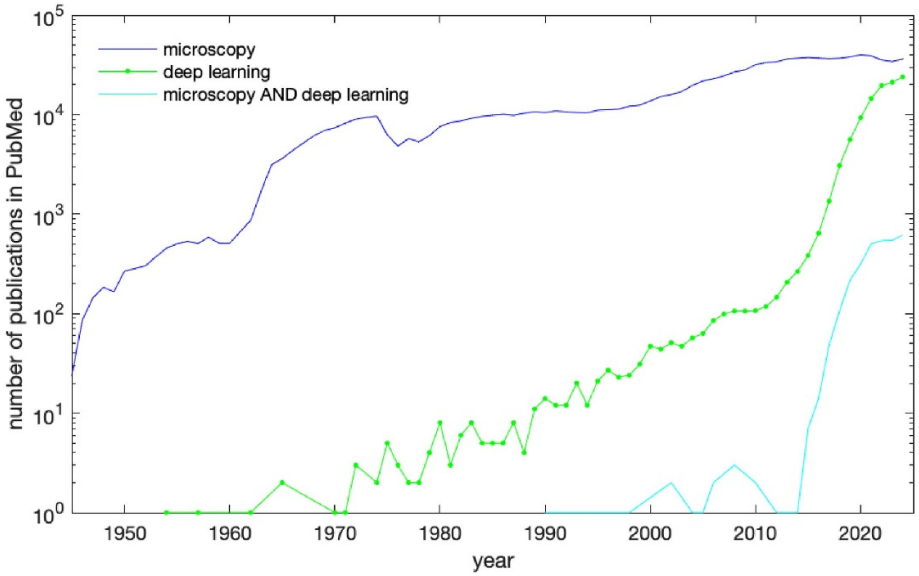

Figure 1. The exponential growth of microscopy with deep learning. A search for the terms 'microscopy', 'deep learning', and the combination 'microscopy AND deep learning' in the PubMed publication database clearly shows the fast development in the field. Between 2000 and 2023, the number of yearly publications concerning microscopy has doubled, while the number of yearly publications concerning artificial intelligence have increased by a factor 30, recently passing the number of yearly publications in microscopy. The number of yearly publications combining deep learning and microscopy have followed the same trend as deep learning, starting with 20 publications in 2002 and now reaching more than 600 publications per year.

Biological and biomedical applications (sections 23–31)
This part demonstrate with some examples how end-to-end (E2E) pipelines convert raw images into biological or clinical insight: micro-physiological systems (MPS) (section 23), self-learning thermofluidics (section 24), DL in digital pathology (section 25), virtual staining of histological tissue (section 26), cell-phenotype determination via virtual staining (section 27), neuro-imaging analysis (section 28), bio-analytical and diagnostic transmission electron microscopy (TEM) (section 29), high-content high-throughput screening (section 30), and ultrasound and photoacoustic image formation (section 31).

Software ecosystem and community platforms (sections 32–36)
This Roadmap closes with resources that democratize ML for microscopists: equitable access to DL solutions (section 32), cloud and containerized deployment strategies (section 33), DeepTrack 2 for user-friendly pipelines (section 34), DeepImageJ for ImageJ/Fiji integration (section 35), and hackathons that spur community innovation (section 36).

All contributions adhere to a common structure—status, current and future challenges, advances needed, and concluding remarks—to facilitate cross-comparison.

Glossary
To make the Roadmap equally accessible to DL specialists, microscopists, and policy-makers, we begin with a compact glossary that standardizes the technical vocabulary used throughout the 36 sections. Each entry gives a plain-language definition of a key concept—ranging from broad notions such as ML and inverse problem to modality-specific terms like FLIM and correlative light and electron microscopy (CLEM)—together with the meaning we adopt in this manuscript.

Term	Definition
Artificial intelligence (AI)	The broad field of creating machines that perform tasks normally requiring human intelligence; in this Roadmap the term is used generically, with a focus on its data-driven sub-domains (ML and DL).
Machine learning (ML)	A subset of AI in which computer programs improve their performance on a task through experience (data). It includes classical algorithms (e.g. SVM, random forests) and deep learning.
Deep learning (DL)	A family of ML methods based on artificial neural networks with many layers that learn hierarchical data representations; the main engine behind recent advances in microscopy image analysis.
Supervised learning	Training a model on paired input–output examples (e.g. raw/noisy image and ground-truth annotation) so it can predict outputs for new inputs.
Unsupervised learning	Training a model on unlabeled data to discover structure, clusters or latent variables (e.g. autoencoders, self-organizing maps).
Self-supervised learning	A form of unsupervised learning in which the data provide their own supervision through pretext tasks (e.g. predicting masked pixels); reduces the need for manual labeling.
Reinforcement learning (RL)	Learning through sequential decision-making where an agent interacts with an environment to maximize cumulative reward; emerging in 'smart' microscope control.
Physics-based/ physics-informed learning	DL or ML models that embed known physical laws (e.g. light propagation, Poisson noise) into the network architecture or loss, improving interpretability and data-efficiency.
CNN (convolutional neural network)	A DL architecture that uses spatially shared kernels; the backbone of most 2D and 3D microscopy image-processing networks (e.g. U-Net).
U-Net	An encoder–decoder CNN with skip connections, originally developed for biomedical segmentation; widely adopted for denoising, deconvolution and super-resolution.
GAN (generative adversarial network)	A framework with a generator and discriminator trained adversarially to produce realistic synthetic images (e.g. virtual staining, image-to-image translation).
Transformer	A DL architecture based on self-attention; it excels at capturing long-range context and is increasingly used for high-resolution microscopy data and multimodal fusion.
Diffusion model	A generative model that learns to reverse a gradual noise-adding process, enabling high-fidelity image synthesis or restoration.
Denoising	Computational removal of noise while preserving signal; DL methods (e.g. Noise2Void, CARE) can operate with or without clean targets.
Super-resolution (SR)	Computational or optical techniques that produce images with resolution beyond the diffraction-limited input; DL-SR networks up-sample low-resolution scans.
Inverse problem	Reconstructing sample properties (phase, structure, chemistry) from indirect or degraded measurements; solved via optimization or DL surrogates.
Transfer learning	Re-using a network pre-trained on one dataset or task as the starting point for a related task, reducing the need for large labeled datasets.
Domain adaptation	Techniques that align feature distributions between source (training) and target (test) domains to maintain performance across microscopes, labs or modalities.
Explainable AI (XAI)	Methods that provide human-interpretable reasons for a model's prediction—important for trust in automated biomedical analysis.
FAIR data principles	Guidelines that data should be Findable, Accessible, Interoperable and Re-usable.
Benchmark dataset	A publicly available, expert-curated set of images and annotations used to compare and validate algorithms under standard conditions.
Microscopy modality	A distinct contrast mechanism or instrument family (e.g. confocal, light-sheet, electron, coherent Raman, SHG/THG, AFM); many sections of this Roadmap discuss cross-modal fusion.
FLIM (fluorescence lifetime imaging microscopy)	Technique measuring the excited-state lifetime of fluorophores; DL accelerates lifetime extraction and noise suppression.
QPI (quantitative phase imaging)	Label-free measurement of optical phase shifts to map cellular mass/thickness; physics-informed DL improves phase retrieval.
CLEM (correlative light-and-electron microscopy)	Workflows that register fluorescence and EM images of the same specimen; DL aids alignment and hybrid segmentation.
Ptychography/FP (Fourier ptychography)	Computational imaging recovering phase and amplitude from multiplexed illuminations; DL helps accelerate reconstruction.
End-to-end pipeline	A single DL framework that maps raw microscope data to the final scientific read-out (e.g. segmentation masks, tracking results) without intermediate manual steps.
Active/adaptive imaging	Real-time ML-driven control of microscope hardware (e.g. laser power, scanning path) to optimize information content while minimizing photodamage.
Digital twin	A physics-accurate simulation of a microscope–sample system, used to generate synthetic training data or to test control policies safely.

2. Stability for inverse problems

Michael Hecht[1, 2, 3] *and Artur Yakimovich*[1, 2, 4, 5, 6]

[1] Center for Advanced Systems Understanding (CASUS), Görlitz, Germany
[2] Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR), Dresden, Germany
[3] Mathematical Institute, University Wrocław, Wrocław, Poland
[4] Bladder Infection and Immunity Group (BIIG), Department of Renal Medicine, Division of Medicine, University College London, Royal Free Hospital Campus, London, United Kingdom
[5] Artificial Intelligence for Life Sciences CIC, Dorset, United Kingdom
[6] Roche Pharma International Informatics, Roche Diagnostics GmbH, Mannheim, Germany

Status

The ability to experience the wonders of the microscopic world with one's own eyes has been fascinating researchers and enthusiasts for hundreds of years. This fascination, as well as the ability to explain the phenomena of the macroscopic scale through the occurrences in the micro-world, has led to the development of a plethora of techniques to visualize, probe, and reconstruct minute objects from the scale of the small animals to the scale of the atom (figure 2). These techniques include a variety of functional (labeled by a molecular dye) and label-free light (optical) microscopy modalities, including brightfield, epi- and confocal fluorescence, and lightsheet microscopy (reviewed in [3]). Attempts to overcome the limitations of the light diffraction limit have led to the development of electron microscopy (EM) techniques like scanning and transmission EM (SEM and TEM, respectively), featuring relatively complex sample preparation steps. The necessity to minimize the sample preparation artifacts and visualize the biological entities in their native state has in turn led to the development of cryo-EM techniques [4]. While the contribution of EM to our understanding of the microworld is difficult to overstate, the sheer complexity of sample preparation and the expensiveness of the equipment has sparked in recent years the development of super resolution microscopy (SRM). Further notable and up-and-coming techniques include x-ray microscopy, live time-lapse microscopy, and holographic microscopy as well as atomic force microscopy (AFM) which uses interaction force to map the microworld.

Remarkably, one common facilitator of this cambric explosion of microscopic techniques, which occurred mostly in the past half a century, is digital microscopy. Originating from micrography, the departure from the necessity to project the microscopic image on the microscopist retina and the ability to capture and describe the images digitally has also has turned microscopy into a quantitative discipline. Beyond facilitating the recording and storage of microscopy data, digitalization allowed improved image processing, denoising, and direct pattern recognition. This, in turn, paved the way for computer vision (CV) algorithms, including ML and DL, to facilitate further advances in microscopy, as further explored in section 34 on the development of user-friendly DL pipelines.

Current and future challenges

Given the immense diversity of the microscopic techniques (figure 2), it becomes obvious that the microscopy datasets are incredibly domain-specific. This represents a significant challenge for ML and DL efforts, as out-of-domain inference is far from trivial for the vast majority of CV algorithms. This is especially pronounced with quite different modalities, for example, EM and confocal fluorescence microscopy. Furthermore, conventional image augmentation approaches that work well for ImageNet work very poorly for microscopy datasets. Vendor-specific data formats are certainly not facilitating harmonization and transferability of datasets. Very often, models trained on images obtained using hardware of a specific vendor simply do not generalize to other vendors. Finally, while microscopy image data is gradually becoming available, high-quality annotations, especially those with a high level of consensus are still problematic to obtain. All these challenges are positioning ML and DL strategies for microscopy data into a low-data regime, dictating the choice of algorithms available to researchers.

In practice, a domain is defined not only by the sample class (e.g. HeLa cells vs. cryo-EM virus particles) but also by the imaging system (numerical aperture, sensor noise, vendor-specific preprocessing). A model that performs well on a single condition quadruplet—one sample, one microscope, one magnification, one preparation protocol—can easily fail once any of these factors drifts. This pitfall is often hidden because many proof-of-principle' papers employ highly restrictive training/test splits such as sparse MNIST-like data-sets or one-specimen demonstrations. Hence, the key question becomes: how much domain-specific training is acceptable before generalizability collapses? Empirically, increasing domain coverage (using cross-modal augmentations, vendor-agnostic raw data, or physics-based simulators) improves robustness but decreases peak performance on the original domain—echoing the

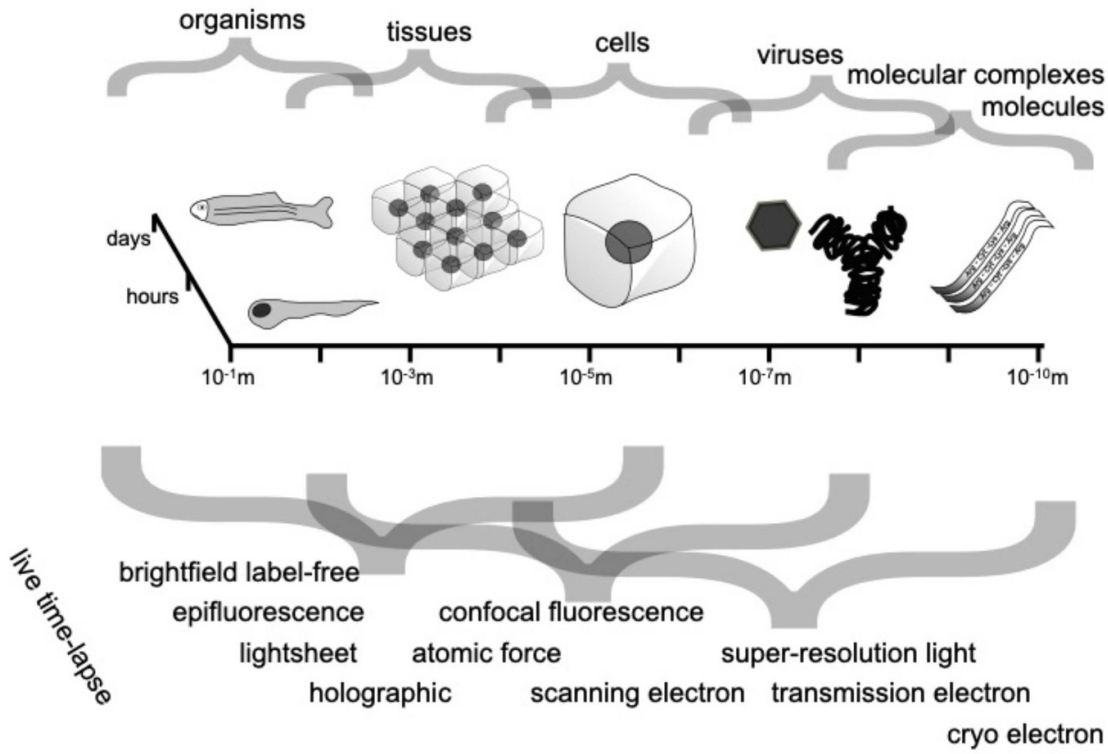

Figure 2. Spatio-temporal spectrum diagram of microscopy techniques and their applications. Here, some method keywords are aligned with a scale range, which is an approximation.

classical bias-variance dilemma and foreshadowing the stability–performance–generalizability trade-off explored below. Similar challenges regarding domain-specificity are encountered in multimodal image registration (section 9).

Advances in science and technology to meet challenges

Recent advances in ML methodology for microscopy have decisively demonstrated the ability of the established ML/DL algorithms to meet the pre-described challenges or their combinations [5], and provide solutions for image processing and analysis. Thus far, this has allowed to successfully address microscopy ML tasks, such as image reconstruction and superresolution, classification and generation, denoising, segmentation, cell tracking, feature selections [5–9] (see also sections 7–13). However, as discussed in reference [8], domain overfitting often hurts generalization. Conversely, attempts to generalize across modalities cause trade-offs [10]. Attempts to overcome these include physics-informed regularization [11]. The gravitation of the established algorithms towards the abovementioned low-data regime motivated Li *et al* [12] to investigate the techniques by asking the questions: 'How reliable are such algorithms when applied in the sciences?' and 'do AI-based algorithms have an unavoidable Achilles heel: instability?'

In mathematical terms, the mentioned tasks seek solutions of non-linear, inverse problems. However, the *Universal Instability Theorem* [13, 14] proves general inverse problems to be inherently unstable. The issue is illustrated in figure 3. Here, the forward model is given by x-projection of a 2D-curved plane onto a curved line. The loss of information in the forward direction (null space) can cause reconstruction distances of arbitrary close instances to become arbitrarily large. In other words: small noise perturbations of an observable x can cause a huge change in the reconstruction of y.

While stable solutions to linear inverse problems can be given by classic principal component analysis (PCA), the non-linearity present here demands the development of novel strategies resisting the instability phenomenon. Recently, approaches addressing the robustness for classification and denoising tasks proposed adversarial re-training, generative-adversarial-network-based denoising, and explicit kernel (null space) control to prevent this issue. These results were discussed and complemented by novel regression techniques in Li *et al* [12]. Furthermore, the current state of research suggests that, when following and incorporating the mathematical insights and delivered ML techniques, stability can be ensured for a general instance class of inverse problems relevant here. Altogether, this suggests that employing novel

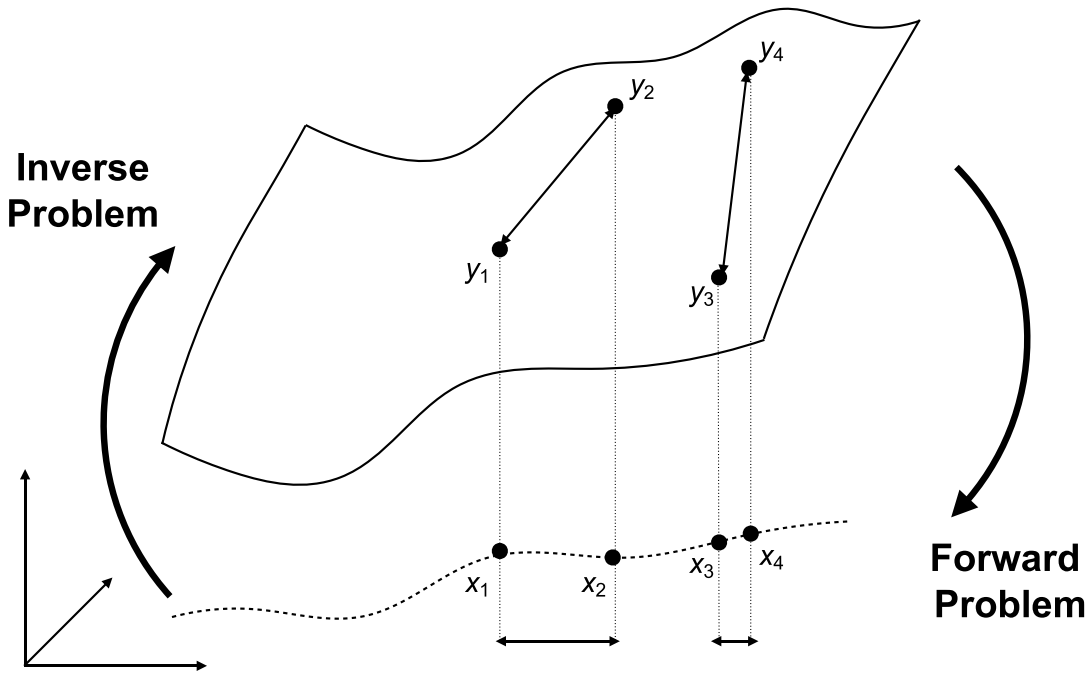

Figure 3. Geometric illustration of the instability for general inverse problems. Though the distance between x_3 and x_4 is small, the distance between the reconstructions y_3 and y_4 can become arbitrarily large [10–12].

more stable ML/DL algorithms may help avoid the inherent instability in ML/DL models for microscopy. Additionally, these approaches may facilitate generative algorithms alleviating low-data regimes and improving generalization.

For example, recent work suggests that these three objectives cannot be optimized simultaneously; improving any two degrades the third. Several strategies are emerging to balance this triangle. For example:

- Domain-adaptive fine-tuning—first train a stable backbone on large, heterogeneous data, then fine-tune lightweight heads for each new microscope or sample while freezing the core layers.
- Physics-informed regularizers—embed the forward model or conservation laws into the loss to anchor learning and reduce instability, even when the network is exposed to unseen domains.
- Uncertainty-quantification layers—Bayesian or ensemble heads that output pixel-wise confidence maps let users decide whether apparent performance is worth the risk of instability in unfamiliar regimes.
- Modular 'universal platforms' such as BioImage Model Zoo and MONAI use self-supervision, federated learning and model-card metadata to offer PnP models whose expected stability range is explicit to the end user (see also sections 32–35).

These trends indicate that rather than chasing a single 'best' model, future work will revolve around adaptive stacks that expose the trade-offs transparently and let practitioners dial-in the desired balance for a given task.

Concluding remarks

Recent advances in microscopy techniques and ML algorithms go hand in hand in powering a new generation of biomedical imaging methods. However, data scarcity and domain-specificity of microscopic datasets represent a major obstacle to this synergy. Furthermore, we identified a bottleneck for providing the strong reliability requested for scientific reasoning. That is the necessity of ensuring the stability of the ML methods addressing the (inverse) problems occurring for microscopy image processing tasks. While enriching the low-data regime may suppress the occurrence of instabilities, we pointed to the objectives hampering this strategy that are omnipresent in practice. Being aware of the mathematical

limitations, however, provides an exit strategy from the dilemma by incorporating techniques that deliver the needed resistance to the instability phenomenon in practice.

Acknowledgments

This work was partially funded by the Center of Advanced Systems Understanding (CASUS), financed by Germany's Federal Ministry of Education and Research (BMBF) and by the Saxon Ministry for Science, Culture and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament.

3. Physics-based learning

Kristina Monakhova and Laura Waller

Department of Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA, United States of America

Status

Computational imaging focuses on the co-design of imaging optics and algorithms to create better and more capable imagers that can see more than just 2D images. In astronomy, this co-design enabled the first pictures of a black hole. In photography, computational imaging allows us to take high-dynamic-range and portrait-mode photos with an extremely compact camera. Here, we focus on microscopy, where computational imaging has been used for super-resolution, single-shot 3D, and phase microscopy, with great potential to push new scientific discovery by allowing us to observe smaller, faster, invisible things in more dimensions.

Physics-based learning refers to any ML approach in which a differentiable physical model of light propagation is embedded directly in the training loop. In optics, this usually takes the form of a wave-optics forward model that enforces Maxwell's equations. In medical-physics applications, the same idea is applied to x-ray transport, ultrasound propagation, or radiative-transfer models for optical tomography. There are in fact parallels to analogous efforts in magnetic-resonance imaging, computed tomography (CT) and radiation-therapy planning, where physics-based learning is already accelerating both image reconstruction and scanner design.

Both the optical design and the algorithm design are critical for computational microscopes. Over the years, optical design has largely been based on heuristics, such as lens sharpness and hand-designed metrics for performance, which were not necessarily optimal given the reconstruction algorithms or imaging task. Similarly, the algorithms have largely been based on optimization approaches consisting of a data-fidelity term and hand-picked priors. These algorithms are often slow, taking many iterations to converge, relied on priors that were not necessarily optimal for the application, and could suffer from model-mismatch given any errors in the optical model. More recently, DL-based approaches have been introduced, which can more tightly couple the optical design with the algorithms [15] and improve algorithms through a data-driven approach [16].

DL-based reconstruction algorithms leverage trainable neural networks (NNs) and large datasets to learn better ways of solving imaging inverse problems. These methods have shown great promise for speeding up imaging inverse problems by multiple orders of magnitude, improving image quality [17], and providing better priors for underdetermined problems, such as compressive single-shot 3D microscopy [18] (see also the challenges of learning interpretable physical models in section 3). Two flavors of physics-based learning are now common. (i) Model-consistent learning keeps the optical hardware fixed and uses a differentiable forward model only to constrain the network during image reconstruction. (ii) E2E hardware-in-the-loop learning treats unknown optical parameters—phase masks, illumination patterns or even freeform lens surfaces—as trainable variables, updating them with the same back-propagation used for the NN. The latter is our focus here because it fundamentally changes how microscopes are designed.

By using differentiable physics-based optical models, the optical design can be represented as the first layer of the NN, and optimized E2E with the reconstruction algorithm [15, 19] (figure 4). This breaks the paradigm of traditional optical design, and opens the door for a new era of optical design where each element in a microscope is tailored specifically for a given algorithm and higher-level task, rather than for a sharpness metric.

Current and future challenges

Despite its promise, physics-based learning is not yet turn-key. Joint hardware-software optimization frequently yields freeform phase masks, multiplexed illumination codes or multilayer diffractive elements that require sub-wavelength lithography or multi-photon 3D printing. Even when printable, the resulting elements may demand sub-micron alignment tolerances or active axial positioning during use. Likewise, wavefront shaping with deformable mirrors introduces calibration overheads and feedback-control complexity that few biology labs possess. Finally, integrating a learned optical design into an existing commercial microscope often voids warranties and necessitates low-level firmware access; hence extensive engineering know-how is still a prerequisite.

Using DL-based techniques for computational microscopy often necessitates large, custom datasets during the training phase. For example, in order to learn the best optical 'encoder' and 'decoder' for

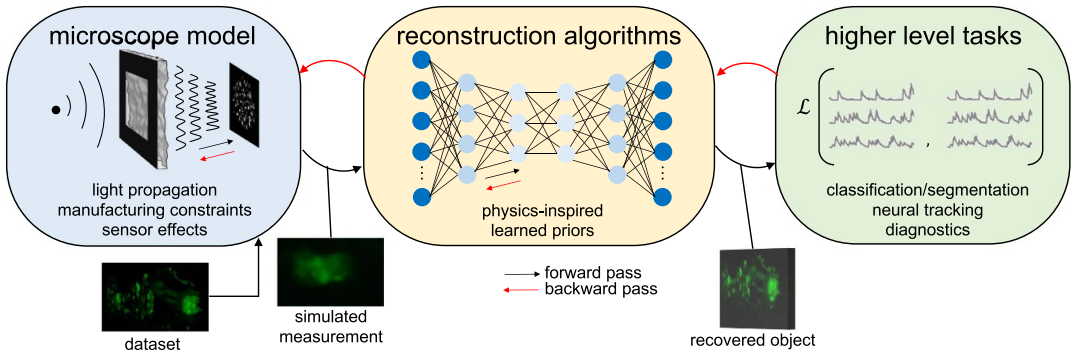

Figure 4. End-to-end design for computational microscopy. Traditionally, the optical hardware and inverse algorithms were heuristically designed. With differentiable optics, the optics can be represented as the first layer of a neural network and jointly learned with an inverse algorithm, and optionally a higher-level task. In this way, the optics can be optimized directly for a given task (here shown for neural activity tracking with 3D single-shot microscopy), rather than optimizing for the best image quality. Reprinted with permission from [103] © The Optical Society Reproduced from [293]. CC BY 4.0.

hyperspectral microscopy, there is a need for a large dataset of high-resolution hyperspectral images to use during training. For certain imaging modalities or tasks, such datasets do not exist, are too small, have insufficient resolution, or may be infeasible to acquire. For other areas of ML, such as self-driving vehicle research, high-quality physics-based simulators have shown great success in synthesizing realistic sensor data. Similarly, generating accurate physics-based simulators for microscopy (e.g. modeling complex light interactions, multiple-scatting, and complex biological samples) could enable computational microscopes to be trained using mostly synthetic datasets, eliminating the need for real experimental datasets. Alternatively, unsupervised learning approaches have the potential to leverage the structure of NNs, but without needing training data [20, 21]. In addition, building and maintaining the data-to-device pipeline is far from PnP: the optical design must be re-trained whenever the illumination source drifts, a component is replaced, or a specimen with different scattering properties is introduced.

Combining domain knowledge of optical physics with DL has the potential to improve the interpretability and performance of DL-based reconstructions [17, 19, 22]. Off-the-shelf networks, such as convolutional neural networks (CNNs), have no knowledge of optical physics and must therefore learn this information from scratch when used to solve an inverse problem, leading to the need for large datasets and lengthy training times. Physics-based networks, in contrast, use differentiable physical models to incorporate optical domain knowledge and create physics-informed networks that are more efficient. These physics-informed networks can also be used to calibrate optical systems, such as by learning how to synthesize and represent realistic camera noise [22], or potentially by learning other effects such as aberrations and non-linearities. This could be useful for synthesizing realistic datasets for a given optical system, or perhaps could be incorporated into the reconstruction pipeline as more accurate, nonlinear forward models.

Finally, computational imaging approaches excel in creating task-specific microscopes that are optimized for a specific higher-level task (e.g. cell counting, disease diagnosis). For many scientific and clinical applications, a high-resolution image is not needed to make a decision—there may be certain features within that image (e.g. polarization, wavelengths, phase) that are more important than others. Through E2E design, where the optics are learned together with the reconstruction and higher-level task, there is the potential to create better, faster, smaller, and more capable microscopes that are specially tailored to make decisions, or extract the most useful information from the world. Overcoming the practical bottlenecks of component fabrication, alignment and instrument control is therefore as critical as progress on the algorithmic side.

Advances in science and technology to meet challenges

Despite the promise of DL for computational imaging, major challenges include robustness and guarantees. A microscope that you cannot trust is a microscope that is unsuitable for scientific and clinical applications. For ML-based reconstruction algorithms, knowing when the algorithm works and when you can trust the reconstruction is difficult. The structure of the network and the training data used can impact the solution and potentially introduce artifacts that are indistinguishable from signal. Research in DL theory on uncertainty quantification [23] and robustness has the potential to resolve some of these challenges. Furthermore, fundamental research in signal processing applied to DL could bring some of the mathematical guarantees from classic signal processing, such as compressive sensing, to the realm

of DL [24]. Building in guarantees and quantifying uncertainty will be pivotal in the broad adoption of ML-based algorithms for scientific and medical computational microscopes. Furthermore, similar concerns about uncertainty quantification are highlighted in the context of particle tracking and force field calibration in sections 16–18.

Concluding remarks

Using DL-based methods to learn better optical designs and algorithms is fundamentally changing the way we design microscopes, helping push the limits of what we can observe, while potentially delivering devices that are smaller, cheaper, and more compact. Although this avenue has many exciting possibilities, there exist a number of scientific challenges that need to be addressed. Open problems include building in interpretability into learned optical designs and algorithms, robustness, quantifying uncertainty, and the need for high quality datasets. Building in domain-specific knowledge and known physics into NNs has promise in addressing some of these challenges. Moving forward, interdisciplinary collaborative research between ML theorists, optical physicists, and microscopists has the potential to further address these challenges and bring in a new era of more capable, ML-designed computational microscopes.

Acknowledgments

This work was supported by the David and Lucile Packard Foundation and the Gordon and Betty Moore Foundation Data-Driven Discovery Initiative Grant GBMF4562.

4. Learning interpretable physical models from large 3D images

Ivo F Sbalzarini[1,2,3]

[1] Dresden University of Technology, Faculty of Computer Science, Dresden, Germany
[2] Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
[3] Center for Systems Biology Dresden, Dresden, Germany

Status

With the advent of volumetric microscopy modalities in biology, such as light-sheet microscopy, it became possible to image entire developing tissues and embryos at sub-cellular resolution over the full time-course of development [25]. This accelerated developmental biology and our understanding of how cells form tissues by providing us with a direct means of observation. However, it also created a new problem: handling the terabytes of image data these microscopes produce at rates of one to five Gigabytes per second [26], depending on the number of cameras used. This is on par with the image sizes produced in astronomy and cosmology [27]. There, a host of approaches has been developed for image compression, sparse representation, and dictionary projection in order to save storage [28, 29]. A key difference with microscopy is that astronomy images are mostly 2D and composed of objects with known expected shape or appearance and slow (if any) time dynamics. The most important difference, however, is that microscopy experiments are typically based on applying a perturbation to the live sample and watching its response.

Since the response to a perturbation can be unpredictable, it is not always clear at the onset of an experiment what the important information in the images is going to be. One solution is to visualize the time-lapse images in real time, as they are being acquired, and only record and analyze the 'interesting' events. But unlike particle physics, where events of interest can be defined *a-priori* and detected algorithmically, biological interest depends on a human observer. For 3D + time microscopy images, human observation is best done using virtual reality (VR) displays. Originally, VR was developed for geometric data, such as triangulated surfaces and point clouds [30]. This can visualize the results of image segmentation, but not the raw voxel volume itself. Using volume rendering, VR has been extended to dense data, notably in medical imaging [31]. Compared to microscopy, however, the data volumes in medical imaging are relatively small. Scaling VR volume rendering to work at the required 120 frames/second on Terabyte-sized microscopy images is not trivial. Another difference between medical imaging and microscopy is that microscopy typically is time-resolved. This calls for real-time visualization while the microscope acquires the data, and it allows for interaction with the sample. Therefore, VR in microscopy is not limited to visualization but also enables new user interaction modalities with the microscope itself and with the sample [32]. An example for this is the open-source platform *scenery* [33]. With a rendering performance of several giga-voxels per second, *scenery* has enabled real-time VR microscopy on commodity hardware, even for the fastest microscopes. This has also enabled the use of natural user interfaces, such as eye tracking, for data analysis [34] and for actual physical perturbation of the sample [35]. Together, such methods help advance our intuition of the complex space-time organization of tissues and its role in disease [36].

Moving from observation to prediction, one would like to formalize this intuition in predictive mathematical models that are physically consistent. It has been shown that governing equations for the space-time dynamics of fluorescently labeled molecules can be algorithmically inferred from microscopy videos with sufficient robustness against imaging noise [37] and guaranteed physical consistency [38]. This extends pioneering works from applied mathematics and ML [39] to the noisy data of microscopy. So far, however, it has not been feasible to apply these ideas on the large 3D images produced by state-of-the-art volumetric microscopy. This is because the dimensionality of the resulting sparse regression problem is beyond the computationally feasible. A promising approach is to represent the raw images on a lower-dimensional data structure than a full regular grid of pixels. Approaches such as the adaptive particle representation (APR) of images [40], for example, improve the information-to-data ratio of the images by adaptively re-sampling them, storing intensity only where it contributes information to the image (figure 5). While the concept of adaptive sampling is classic in signal processing, the APR provides unprecedented approximation and optimality guarantees not present in previous approaches like supervoxels [41]. The APR can also directly be used for downstream image processing [42–44], which is not possible with, e.g. wavelet multi-resolution pyramids [45]. This enables E2E APR-native pipelines for images up to the Petabyte scale on consumer GPUs [43] and significantly reduces the compute needs of CNNs on large 3D images [44].

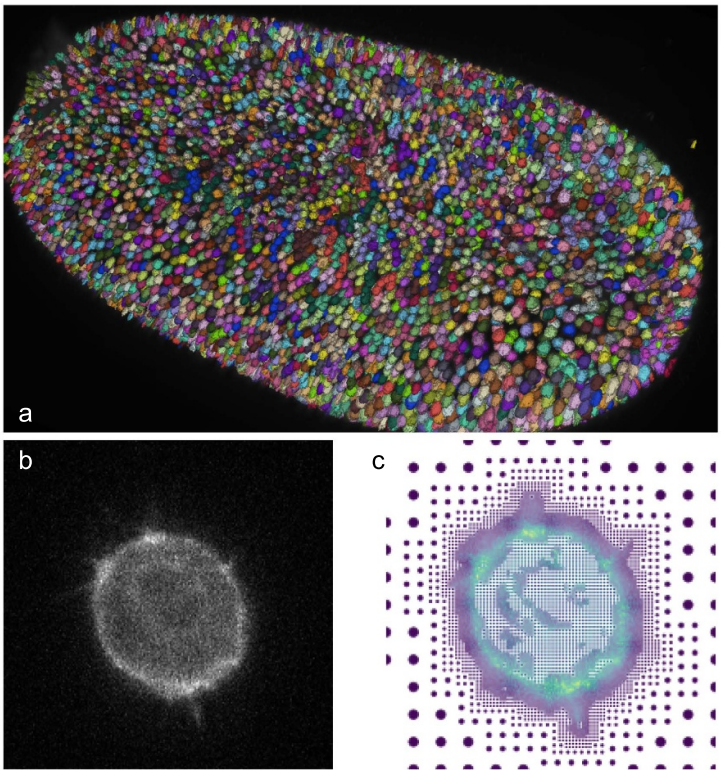

Figure 5. The adaptive particle representation (APR). The APR of images [40] provides an information-optimal sampling of an image at a fraction of the computational cost. (a) An embryo with cells segmented (original image: data set 'Fluo-N3DL-TRIF' from the publicly available ISBI cell tracking challenge; segmentation and visualization: Joel Jonsson, Sbalzarini lab). (b) A single A549 cell from the cell tracking challenge. (c) APR particles (shown as individual dots) of the cell from (b) with color encoding the intensity signal (APR and visualization by Joel Jonsson, Sbalzarini lab).

Current and future challenges

Current microscopy datasets are chronically under-used in large bio-imaging studies. Analyzing the images and turning them into scientific insight and knowledge currently is the rate-limiting step. We envision a future where volumetric microscopy data sets can be immersively experienced in VR environments, from within which the user can interact with the sample in real time using hand gestures to, e.g. apply optogenetic manipulations or laser ablation and thereby train ML modules running in-the-loop. This brings within reach human-augmented algorithmic inference of the physical laws and processes that characterize the dynamic self-organization of living matter. While the bits and pieces for making this vision a reality are becoming available, combining them remains a challenge.

It is, for example, challenging to combine the APR [40] with real-time VR rendering, because the performance of rendering algorithms on GPUs deteriorates on irregular, adaptive data structures. Another example is that, while learning interpretable mathematical models on limited and noisy microscopy image data is becoming possible [37], this does not currently scale to large 3D data.

Successfully scaling to large data, DL architectures have been adapted to biological imaging. This includes graph neural networks (GNNs) to estimate physical properties of biological systems directly from microscopy videos [46] and the variational autoencoder *Cytoself* to infer models of intra-cellular mesoscale organization from microscopy images [47], inspiring thoughts of a sequence-to-image generative model. Since they can be trained in a self-supervised way, not requiring manually annotated ground-truth data, autoencoders are a promising architecture for microscopy. In particular, they allow mapping images into a joint latent space with genetic and chemical states of cells, providing exciting prospects for drug screening [48] and cancer [49]. Other approaches are presented in sections 16–20.

The physical and biochemical interpretation of such DL models, however, remains a challenge, which could potentially be addressed by combining them with mathematical model inference. This could include endowing them with a notion of spatial interaction [50]. Currently, deep neural networks (DNNs) operate on the morphological appearance of objects in the image, but not on their spatial arrangement with respect to each other and with respect to reference structures, such as the cell nucleus or the plasma membrane. Concepts from spatial statistics could be re-cast in a data-driven framework to enable the next leap in understanding the spatial organization of living matter using DL.

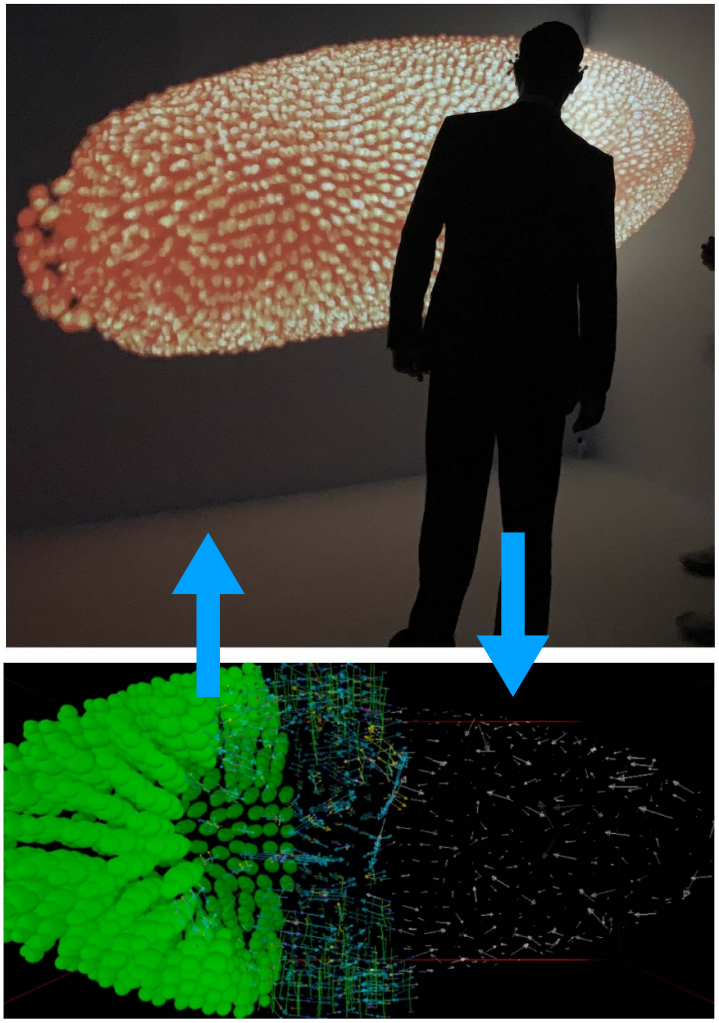

Figure 6. Vision of human-AI co-working in a virtual-reality space. The top panel shows a scientist interacting with a fruit fly embryo in the microscope in 3D virtual reality in real time (photo credit: Ulrik Günther, HZDR & Sbalzarini lab). The bottom panel shows output (synthetic image) of a simulation model [52] of the cells, the forces acting on them, and their velocity fields (image with permission by Vladimir Ulman, Technical University of Ostrava).

Advances in science and technology to meet challenges

Both scientific and technological advancements are needed to realize the vision of integrating immersive visualization and ML to infer interpretable physical models from large-scale 3D image data. Importantly, this needs to be linked with algorithms that extract physical fields, such as velocity fields of flows and deformations, from the images and use those to infer mathematical models that are simple enough to be physically interpreted, but complex enough to explain the dynamics observed in the data.

This link is likely going to come from modern ML architectures, such as transformer networks or GNNs (section 20). A key challenge in applying DL to microscopy data, however, is the limited availability of annotated ground-truth training data. Approaches such as self-supervised training, contrastive learning, geometric learning, and transfer learning from simulated synthetic data are hence going to be pivotal [51]. Combining this with data- and energy-efficient compute architectures, as well as immersive human-in-the-loop natural user interfaces, could define a new inference loop for future microscopy.

This inference loop potentially enables experiments in which scientists use hand gestures in a VR environment to interactively ablate a tissue in the sample, are able to directly observe the recoil and tissue rearrangement this causes, and within seconds get the estimated elastic constants of the tissue as well as a simulation results overlaid with the image (figure 6). Further interacting with the sample, or manually correcting estimation mistakes, the human re-trains the ML or simulation model in an active learning loop. An interesting question is how to visualize uncertainty in ML output to the users in order to focus their attention and prompt for additional perturbations that carry significant new information. Finally, bringing these advances in ML and bio-imaging into everyday laboratory usage requires

user-friendly, accessible, easily deployable, and robust software implementations with support for cloud computing.

Concluding remarks

Exciting times for biological live-cell microscopy! We are witnessing accelerated progress in the molecular markers and sensors available, the optics of the microscopes, and our ability to use light to perform quantitative measurements in living, growing, and deforming organs and organisms [53]. The challenge is to ensure that the computational tools for handling, storing, processing, and visualizing the images progress on par, and that we work toward making better use of the information contained in the images, e.g. by providing ML solutions that increase the throughput at which physically interpretable, mechanistic models can be inferred from the data. This ideally goes directly from images or videos to equations, models, and knowledge, bypassing the so-far common intermediate CV steps of segmentation and tracking. This will not only make large bio-imaging projects more insightful, but also potentially more accurate and less resource-demanding. Ultimately, a new ecosystem of biological imaging arises, combining approaches from VR, AI, and signal theory, which could all become common laboratory methods in a couple of years from now, just like single-particle tracking (SPT) has become a robust commodity [54].

Acknowledgments

I thank Vladimir Ulman (Technical University of Ostrava) and Ulrik Günther (HZDR & Sbalzarini lab) for kindly providing the two panels of figure 6 *and Joel Jonsson (Sbalzarini lab) for kindly providing* figure 5.

5. PnP learning-based computational imaging

Christopher A Metzler, Mingyang Xie and Kevin Zhang

University of Maryland, College Park, United States of America

Status

The performance of any computational imaging system is determined, in part, by the performance of its reconstruction algorithm. The role of a computational imaging reconstruction algorithm is to reconstruct a signal/image of interest x from measurements y of the form

$$y = A(x),$$

where $A(x)$ represents the system's forward model. This mathematical model is general and can represent almost any measurement process, from sampling K-space in magnetic resonance imaging (MRI) to capturing band-pass filtered images in Fourier Ptychographic microscopes.

Classically, computational imaging reconstruction has been performed by framing imaging as an optimization problem

$$\hat{x} = \arg\min_x f(x) + r(x),$$

where $f(x)$ is a data fidelity term, which ensures x is consistent with the measurements y, and $r(x)$ is a regularization penalty, which ensures x is consistent with our prior beliefs on how x should be structured. In the special case that $f(x) = -\ln p(y|x)$ and $r(x) = -\ln p(x)$, solving the above optimization problem provides a maximum *a posteriori* (MAP) estimate of x.

Assuming the optimization problem is sufficiently convex, one can minimize the objective, $f(x) + r(x)$, using any number of algorithms, such as proximal gradient descent:

$$v^{t+1} = x^t - \nabla_x f(x^t),$$

$$x^{t+1} = \arg\min_x \left\| x - v^{t+1} \right\|^2 + r(x).$$

The first line above takes a gradient step to minimize $f(x)$ while the latter, which is known as a proximal mapping, reduces $r(x)$.

Classical iterative algorithms such as proximal gradient descent offer several benefits: They are easy to interpret and, by changing the data fidelity term $f(x)$, can easily incorporate domain expertise about the measurement process $A(\cdot)$. Unfortunately, the performance of classical algorithms falls well behind that of purely DL-based methods, which, given a vast training set of paired examples $\{(x_l, y_l)\}_{l=1}^{L}$, can teach a NN to effectively map measurements y to images x. This performance generally comes at the cost of flexibility—a network is trained and specialized for a specific measurement process $A(\cdot)$, and if the measurement process changes the network is useless.

PnP optimization is a hybrid reconstruction framework which allows one to maintain the interpretability and flexibility of classical algorithms while taking full advantage of powerful data-driven priors [55]. The key idea behind PnP optimization is that one can interpret an off-the-shelf (learning-based) image denoiser $D(z)$ as a solution to

$$D(z) = \arg\min_x \left\{ \|x - z\|^2 + r(x) \right\},$$

for some implicit, data-driven prior $r(x)$. This interpretation allows one to 'plug-in' the denoiser into an existing iterative algorithm. For example, PnP proximal gradient descent can be written

$$v^{t+1} = x^t - \nabla_x f(x^t),$$

$$x^{t+1} = D\left(v^{t+1}\right).$$

This algorithm follows the same form as standard proximal gradient descent, but with the proximal mapping step replaced by a powerful image denoising algorithm.

PnP algorithms stand apart from purely data-driven and purely classical computational imaging reconstruction techniques in their unique ability to combine (1) domain expertise, in the form of the

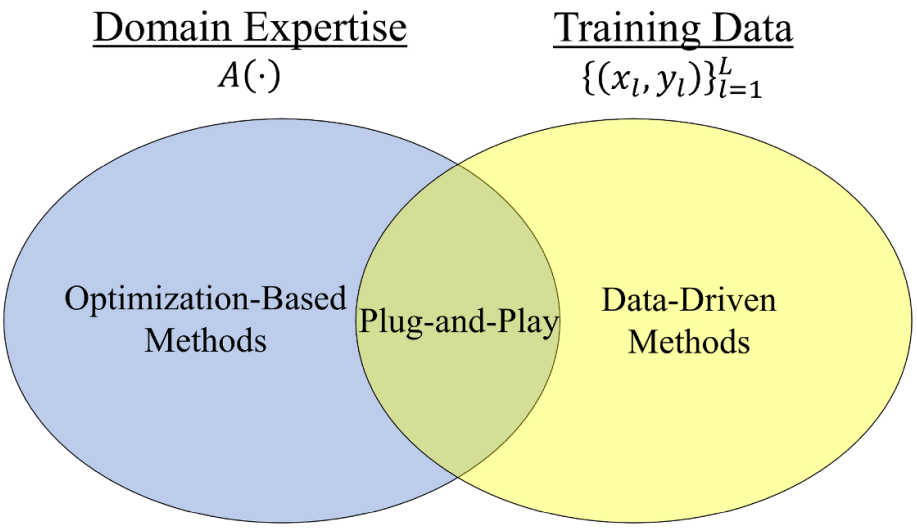

Figure 7. Plug-and-play methods. Plug-and-play methods are hybrid algorithms which can take advantage of both domain expertise about the measurement process and powerful data-driven learned priors. This combination makes them both adaptable and effective.

forward measurement operator $A(\cdot)$, and (2) data-driven-priors on the distribution of the dataset, in the form of an image denoiser trained using a vast set of training examples $\{x_l\}_{l=1}^{L}$ (figure 7). PnP algorithms represent the current state-of-the-art in computational imaging reconstruction [56]. Related hybrid strategies combining domain knowledge and learned priors are also employed in physics-based computational microscopy (section 2). Unsupervised neural-network-based methods can also make use of domain expertise about the forward process, but they are generally incapable of using training data; accordingly, the resulting priors are far weaker.

Current and future challenges

PnP algorithms face several hurdles which can make them challenging to apply in practice.

Forward model mismatch: Like almost any other reconstruction method, PnP algorithms rely on accurate knowledge of the forward measurement model $A(\cdot)$. If this forward model is mis-specified severe artifacts can appear in the reconstructions.

Denoiser model mismatch: PnP algorithms typically use off-the-shelf denoising algorithms, which often assume the noise they are removing follows an additive white Gaussian distribution. In general, however, the 'effective noise', i.e. the difference between the true signal and the intermediate solution fed into the denoiser, is neither Gaussian nor white, making existing denoising algorithms suboptimal. In addition, denoisers are trained on a specific class of data (e.g. natural images) and their performance can suffer when applied to other distributions (e.g. MRI).

Parameter tuning: Most PnP algorithms have multiple hyperparameters (step sizes, regularization strengths, etc). These parameters can vary iteration to iteration and setting them correctly is critical to getting the best performance. One solution to the parameter tuning problem is to 'unroll' an iterative PnP algorithm into a feedforward NN (figure 8), whose parameters can be automatically tuned with training data and backpropagation [57]. However, unrolling gives up flexibility, the network is now specific to a particular forward operator $A(\cdot)$, and comes with its own set of limitations.

Memory usage: To train an unrolled algorithm, one needs to back-propagate errors through multiple copies of the denoiser, which requires storing many intermediate variables. When dealing with high-dimensional data, e.g. time-varying volumetric data, the memory costs associated with storing all these intermediate variables can become prohibitive (see also section 3).

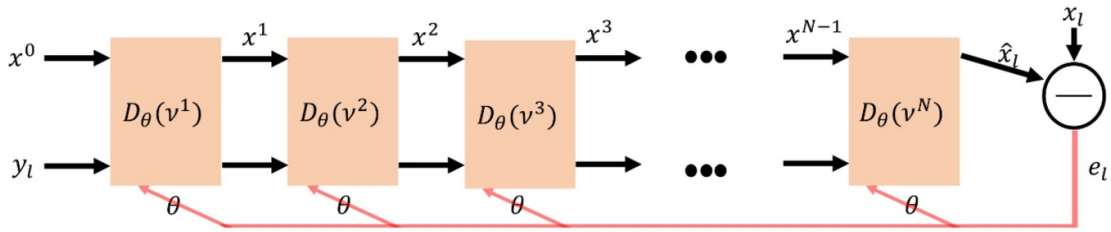

Figure 8. The unrolling process. One can 'unroll' N iterations of a PnP algorithm into an N-layer feedforward neural network. Unrolling allows one to use training data to tune and optimize its many parameters but comes at the cost of flexibility: The resulting network becomes specialized for a particular forward measurement process.

Advances in science and technology to meet challenges

A host of solutions have been put forward to address the above challenges.

Forward model mismatch: If one is dealing with a parametric forward model $A_\gamma(\cdot)$, one can jointly recover the target image x and the forward model parameters γ through alternating minimization. Specifically, in addition to the usual PnP steps, one can also minimize the objective function with respect to the forward model parameters. Variations on this idea have been applied successfully to x-ray CT [58], MRI [59], and holography [60].

Denoiser model mismatch: Denoiser model mismatch comes in the form of mismatched noise distributions and mismatched data distributions. The former problem can be addressed through unrolling [57] or by careful characterization of the per iteration noise distribution [61]. The latter problem is largely still open, though denoisers tend to be relatively robust to this mismatch in practice.

Parameter tuning: While unrolling is the most straightforward way to perform parameter tuning [57], unrolled algorithms largely (though not entirely [59]) give up the ability to adapt to new measurement operators. Alternative parameter tuning methods, such as reinforcement learning [62], have recently been put forward to avoid this trade-off.

Memory usage: One can reduce the massive memory costs associated with training an unrolled algorithm by utilizing invertible network architectures, which allow one to recompute intermediate activation from a network's output [63]. Alternatively, one can improve memory usage by relying upon stochastic approximations of the data fidelity update step [64].

Concluding remarks

PnP algorithms provide a convenient framework to combine the adaptability and interpretability of classical optimization-based computational imaging reconstruction methods with the powerful priors and efficiency of DL. Here, we have highlighted some of the recent developments in PnP reconstruction and have described some of the pitfalls to avoid when applying learning-based PnP algorithms.

Acknowledgments

This work was supported by the AFOSR Young Investigator Program Award FA9550-22-1-0208.

6. Accelerating simulations of optical forces

Isaac CD Lenton[1,2] *and Halina Rubinsztein-Dunlop*[2,3,4]

[1] Institute of Science and Technology Austria (ISTA), Klosterneuburg 3400, Austria
[2] School of Mathematics and Physics, The University of Queensland, Brisbane 4072, Australia
[3] Australian Research Council Centre of Excellence for Engineered Quantum Systems, The University of Queensland, Brisbane 4072, Australia
[4] ARC CoE in Quantum Biotechnology, Australia

Status

Optical tweezers (OTs) are a powerful tool for manipulating and handling microscopic samples, whether in liquids, in gases, or in vacuum [65]. Since their invention by Arthur Ashkin [66], OT have developed as a robust tool that is routinely employed in multiple disciplines [67]. OT use the optical force of light to trap and manipulate objects; in their simplest form OT can be thought of as an *optical harmonic spring*, which, under the right circumstances, can allow a particle to be held in all three dimensions and moved around by simply moving the optical beam focus. The strength of the spring is proportional to the beam's optical power. When the optical trap is well characterized, i.e. the stiffness of the spring is well known, OTs become a powerful method for measuring piconewton scale forces, nanometer displacements, and torques by directly monitoring how the particle moves while held in the trap.

Simulations and models of OT have often aided their development, and can be useful for verifying and understanding experimental observations [68]. For instance, when trying to understand the observed dynamics of an optically trapped motile bacterium, a thorough understanding of the optical forces can help to shed light on the non-optical forces involved in the bacterium's motility. Depending on the size and shape of the particle, different descriptions of the trap are often needed, ranging from simple methods analogous to one or more harmonic springs, through to methods solving Maxwell's equations. The main limitation has often been the ability to use these models to make predictions about experiments. This can be either due to computational constraints, when advances in computing power are needed to evaluate models [69] or due to how easy-to-use and well-documented available codes are. Complex models often have many parameters and verifying that these models accurately describe a given experiment can be difficult.

ML, which encompasses a range of techniques including DL and computer-enabled statistical inference, has seen a huge growth in popularity in recent years. Aided by a rise in computational power and simultaneously an increase in the number of easily accessible algorithms, ML has been adopted into many fields. In the past few years, we have seen ML techniques applied to OTs related problems, including modeling optical forces [70, 71] and extracting relevant information from experiments [72, 73]. These early works are very promising and suggest ML could significantly accelerate OT simulations, while simultaneously making accurate models more accessible for researchers wanting to model and understand their experiments.

Current and future challenges

Multiple challenges limit the application of accurate OT models including available computational power, verifying models accurately represent reality, and accessibility of models to researchers.

Accurately modeling large, complex particles, and structured light fields can be difficult. Although computers have significantly improved in speed and memory capacity (extending the range of cases that can be modelled), for many researchers it is still difficult to model even large spherical particles using available hardware[68], let alone the more complex geometries of bacteria or deformable cells often studied using OT. To simulate the dynamics of these particles in OT, the simulations must be sufficiently fast. One solution to modeling large and complex particles/beams is to pre-compute the forces using a more powerful computer or computer cluster. Interpolation can then be used to approximate the forces at intermediate points not included in the data set. However, this simply transfers the problem from being a processor-limited problem, to a memory/storage problem. An alternative option is to use ML to encode all this information in a much smaller representation. One of the main challenges in using interpolation or this kind of ML is generating sufficient training data spanning a representative range of parameters (beam shape, particle size, etc).

[68] We were only able to simulate particles up to about 100 μm and 1 mm with implementations of generalized Lorentz-Mie theory and geometric optics via the optical tweezers toolbox and optical tweezers simulations software packages. We only tested if these methods would run in a feasible amount of time for these sizes. No checks were made to see if the results for these sizes were accurate.

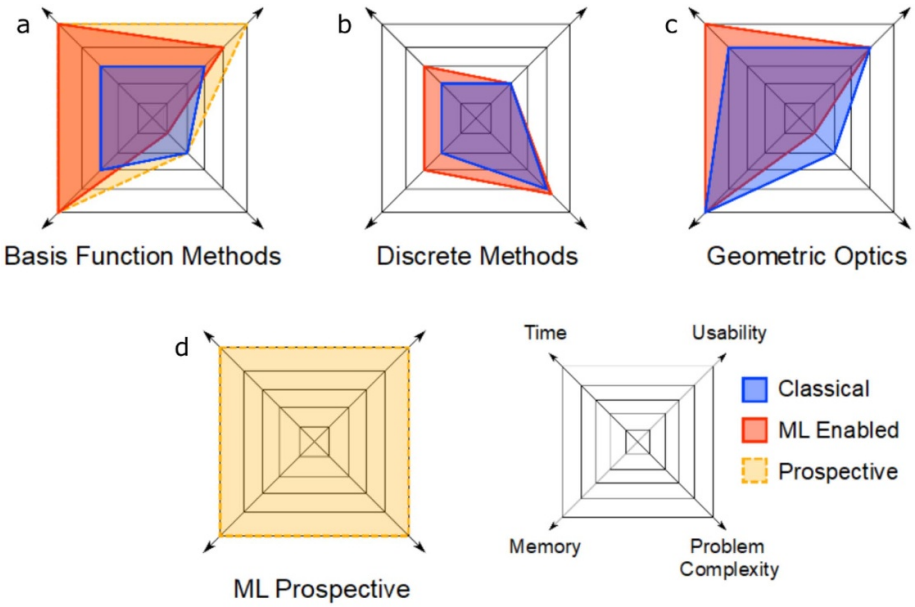

Figure 9. Comparison of the current classical, machine learning enabled and prospective capabilities of different methods used to model optical tweezers. Shaded regions indicate how well the methods work in terms of time, usability, memory, and problem complexity: larger regions correspond to better performance/usability. Plots broadly group different methodical approaches to modeling optical tweezers: (a) basis function methods (such as T-matrix and dipole approximation based methods), (b) discrete methods (including discrete dipole approximation and finite difference based methods), (c) geometric optics, and (d) the prospective advancements enabled by a combined ML approach (as described in the text).

One of the main difficulties with developing computational models for OT is *determining whether these models accurately depict reality*. Not all models work well on different size scales: for instance, the dipole approximation can be very effective for dipole-sized particles but give unreliable results for larger particles; similarly, geometric optics can be great for large particles, but produce erroneous results for small particles. Sometimes it can be necessary to combine multiple models, and care must be taken for when a particular model should be used/is valid. Often it is up to the user to assess if the model is suitable/accurate for the particular experiment they are trying to model. Care must be taken when using both traditional models and ML enabled models that the model predictions accurately depict reality.

The final challenge is *usability and accessibility*. There are many excellent models for optically trapped particles, however many of them are not the easiest to use or access. In some cases, users need specific operating systems or familiarity with certain programming languages to use/run codes (some of which may be proprietary). Once the user has a running model, they need to learn how to use it and assess its validity. For example, the discrete dipole approximation is extremely useful in modeling a range of particles but the choice of voxel placement, size, and spacing can have huge effects on the simulation results and care must be taken to choose the appropriate parameters and verify results. This requires good documentation including good release notes highlighting differences between versions of the code. This extra work sometimes creates a barrier between researchers who develop codes and users wanting to apply these codes to their experiments. Further still, in some cases researchers are hesitant to release code in the fear that it might be used incorrectly.

Advances in science and technology to meet challenges

There is no shortage of good algorithms for modeling OT. While there are still plenty of active challenges (for example, see sections 7–11 of [74]), for many researchers working with OT the problems are rather in accessibility/usability. In some cases this is due to computational limits, uncertainties about applicability to a particular problem, or simply how easy the method is to download/install. To summarize this problem, figure 9 shows a current version of the methods' comparison in Bui *et al* [68] with the addition of a usability axis. We have grouped methods into three broad categories. Tractable problem complexity is represented by a single axis (somewhat representing minimum/maximum size and beam/-particle complexity).

ML has already pushed the boundaries of what traditional techniques can achieve. For example, in our work [70], we showed how an artificial neural network (ANN) (i.e. a specific ML technique) depicted in figure 10(a) can be used to significantly reduce simulation memory and time requirements. This

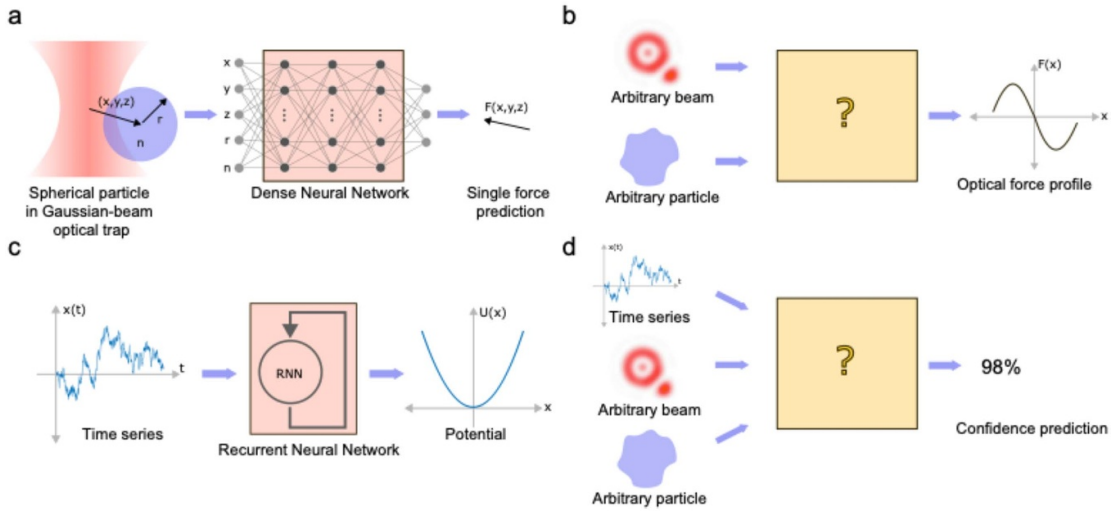

Figure 10. Graphical depictions of different machine learning models for optical tweezers simulation and classification. (a) A 5-degree-of-freedom network for calculating forces on a spherical particle in Gaussian beam optical tweezers (Reproduced from [70]. CC BY 4.0.). (b) A more general target for ML where the network transforms any beam, particle distribution into a force field. (c) Example of a recurrent network for extracting optical potentials from experimental time-series measurements (Reproduced from [73]. CC BY 4.0. CC BY 4.0.). (d) A network which estimates if the given beam/particle are good predictors of an observed trajectory.

network is trained on representative force values for a particular beam/particle combination and can rapidly predict values at other locations while using significantly less memory compared to interpolation. We demonstrated the technique for the T-matrix method and a similar approach has since been demonstrated for accelerating geometric optics simulations [71]. The technique has the potential to increase accessibility/usability since the resulting networks are often small and fast enough to be embedded in an interactive web-browser based simulation. The main shortcoming of these kinds of models is generalizability: i.e. a model trained on spherical particles is not well suited to predicting forces and torques on ellipsoidal particles. Ideally, we would want a model that looks more like figure 10(b): taking a beam and particle as input and predicting the corresponding force field[69]. This would be a significant undertaking, with many challenges including the generation of sufficient training data, development of efficient ways to represent the problem and identification of the current limitations in numerical techniques for modeling OT. For this to be achieved, sufficient training data must be generated and collated. This will likely involve choosing a network architecture which generalizes well to learn features with much less training data. For instance, a convolutional network might perform better when given images of the optical field and particle as compared to the network in figure 10(b). Other techniques such as transfer learning or model re-training might also be key to reducing the required volume of training data. While it is unlikely that ML used in this context will identify new physics, it may help to reveal emergent behavior from the dynamics of particles moving in OT—behaviors that are difficult to predict with traditional models due to computational limitations.

ML also offers a potential solution to validating models correctly describe experiments. ML techniques have been used to analyze experimental measurements, and can be used, for instance, to extract potentials from particle trajectories [72] (see also sections 16–20). Figure 10(c) shows a depiction of a recurrent neural network (RNN) (another ML technique) which converts a time series measurement into a potential. By comparing these potentials with model predicted optical forces, we can verify the accuracy of our models—while this verification could conceivably be performed without ML, the use of ML significantly simplifies the process. A extension to this idea is depicted in figure 10(d), showing a network that discriminates between if a model (beam/particle description) is sufficient to describe a particular experimental observation with some confidence level. One advantage of this problem formulation is that we can apply the same optimization functions we used to train the network also for optimizing the model parameters (beam/particle shape).

As a final note, the ML boom has coincided with many other advances in available tools for sharing and collaborating on code and ML model training. Particular tools of note include:

[69] For non-spherical particles, this would be a 6-dimensional space spanning possible positions and orientations of the particle.

1. Version control software (such as Git and SVN) and related hosting sites (such as GitHub and GitLab) which provide tools for tracking both code and known issues/features/developments.
2. Papers with Code: a website where anyone can share implementations of algorithms described in papers.
3. Containers (such as Docker and Singularity) offer a way to package not just the code but the entire working environment (i.e. specific software versions and dependencies).
4. Read the Docs: an online documentation hosting site which can be setup to automatically hook into popular source distribution tools and automatically generate and publish documentation when a new version of the code is released.

The availability of easy-to-use tools is only part of the solution. In addition, there needs to be a cultural shift to encourage the publication of ML OT models and acknowledgment of the additional work required developing documentation, maintaining models and responding to user requests/feedback.

Concluding remarks

We have already seen that ML can improve access to powerful OT models without users needing super-powerful computers. We have also seen knock-on effects of the ML boom improving software tools for better documentation and collaboration. Improvements in particle tracking and OT calibration, enabled by ML, have the capabilities to ensure the models we generate are accurate. The advances described above are summarized in figure 9. We currently envision these being extended further, one of the main limitations is resources including time to generate large data sets, train models, and explore alternative model architectures. We should expect to see new ML models pushing the bounds of what is possible in OT simulation. We might even envisage a universal OT simulator that spans and unifies the capabilities of our present numerical models for a range of cases that would be useful to the OT community. This would require accurate datasets spanning a wide range of beams and particles. In order to achieve this goal, we need to make the existing models more accessible. While we have focused on OT here, many of these techniques could be applied to simulation of other systems such as acoustic tweezers, electrodynamic traps and optical imaging.

Acknowledgments

H.R-D received funding from the Australian Research Council Discovery Project DP180101002; and the Australian Research Council Centre of Excellence for Engineered Quantum Systems (EQUS, CE170100009).

7. Optical networks for all-optical computing and imaging

Daniel Brunner[1], *Bijie Bai*[2] *and Aydogan Ozcan*[2]

[1] Département d'Optique P. M. Duffieux, Institut FEMTO-ST, Université Bourgogne-Franche-Comté CNRS UMR 6174, Besançon, France
[2] Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, United States of America

Status

Optical NNs have a long history. It started in 1985 in seminal experiments that implemented an optical version of Hopfield networks [75]. The prime motivation, which remains unchanged, was to exploit attributes inherently linked to optics: massive parallelism and the resulting advantage for the *connectionist* NN computing concepts. And this promise actually materialized, as real-time face recognition was demonstrated in an experiment that holographically stored weights optimized during a training procedure [76].

Since then, AI went through numerous hype and bust cycles, until it became the indispensable tool of today. Today's high-performance electronics can implement NN topologies with ever-increasing complexity and capabilities. Yet, the fundamental mismatch between computational hardware and NN topology creates enormous challenges in terms of scalability, energy consumption and speed. Similar motivations to overcome hardware limitations through novel learning-based designs are also driving developments in PnP optimization frameworks, as discussed in section 4. The result is that adding NN-functionality to other devices comes associated with a substantial power budget, and that real-time processing of high-volume data is unattainable. Employing NNs for either real-time high-resolution wide-field microscopy or small, non-lab-based systems remains challenging. Optical NN, therefore, remains of interest, and recently, attention towards such unconventional hardware exploded.

Current and future challenges

Currently, efforts toward next-generation photonic hardware for NNs include integrated photonic solutions [77] and free-space [78–80], but also efforts towards addressing the fundamental challenge when integrating NNs using 3D photonic integration [81] or unlocking the ultra-fast potential of fully parallel and autonomous photonic NN [82].

As a free-space optical computing platform, diffractive deep neural networks (D^2NNs) have attracted growing interest as they compute a given task by engineering light diffraction through a series of complex, spatially structured surfaces, one layer following another. Given a targeted task, these layers are optimized using DL concepts to minimize a task-specific loss function. The resulting diffractive layers collectively form an all-optical computer, performing the desired transformation between the input and the output fields-of-view. Once the DL-based training is complete, the diffractive layers can be fabricated through e.g. 3D printing or lithography, creating a physical network that performs the designed computational task at the speed of light propagation without requiring any external power except for the illumination light.

Since its first demonstration in 2018 [78], diffractive NNs have been successfully employed for various applications, for example, holography, quantitative phase imaging (QPI) [83] and class-specific imaging [83]. Compared to using electronic digital processors, D^2NN inference in computational imaging offers some unique advantages since the optical information of a scene is directly accessible to a diffractive network. This shares conceptual parallels with physics-based learning approaches to direct phase retrieval without heavy pre-processing, highlighted in section 2. Therefore, no digitization or complicated pre-processing steps (such as phase retrieval) are needed, enabling computer-free 'computational imaging'. Using a trained D^2NN, the spatial information of any unknown objects can be instantly retrieved from raw holograms at the speed of light propagation without requiring any digital processors [84], or can be used to directly convert the phase information of an arbitrary input scene into an intensity distribution at its output plane, creating an all-optical QPI system (figure 11(a)). It was demonstrated that this QPI diffractive network could generalize to unseen, entirely new phase objects.

D^2NNs have also been applied to imaging through diffusive media, which is crucial across various fields such as biomedical optics, atmospheric physics, astronomy, oceanography, and autonomous systems. A trained D^2NN was demonstrated to all-optically recover the images of arbitrary objects completely distorted by unknown, random phase diffusers, presenting a real-time, computer-free, and power-efficient solution to imaging through random and unknown diffusers (figure 11(b)).

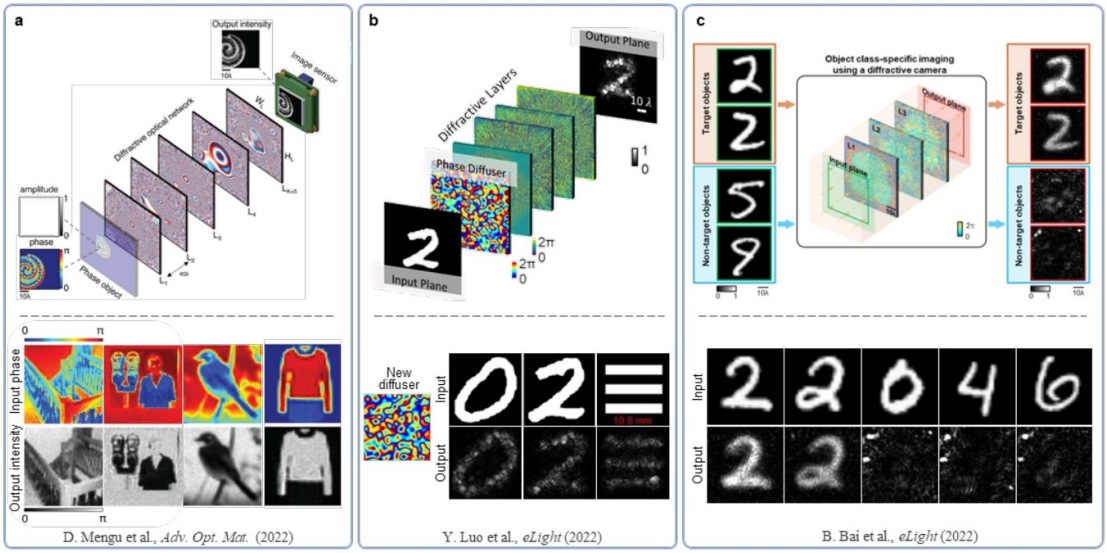

Figure 11. Examples of D^{2}NN applications for computational imaging. (a) Using D^{2}NNs for quantitative phase imaging (QPI) [83]. John Wiley & Sons. © 2022 Wiley-VCH GmbH. (b) Using D^{2}NNs for imaging through random, unknown diffusers. Reproduced from [85]. CC BY 4.0. (c) Using D^{2}NNs to design class-specific cameras. Reproduced from [86]. CC BY 4.0.

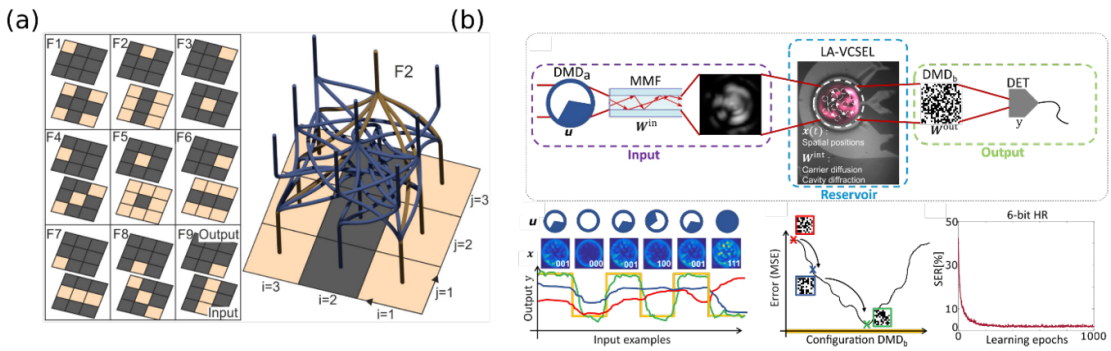

Figure 12. (a) Schematic illustration of three-dimensional (3D) integrating nine Haar filters (F1–F9). A 3D-printed waveguide architecture realizing all nine filters in parallel is shown on the right, with filter F2 highlighted in orange. Reprinted with permission from [81] © The Optical Society. (b) Implementation of a recurrent neural network inside a multimode large-aperture vertical-cavity surface-emitting laser (LA-VCSEL). The optical weights are written by a digital micromirror device (DMD) through a multimode fiber (MMF). The modulated output is read by a high-speed detector (DET), enabling fully parallel linear and nonlinear operations at ≈20 GHz bandwidth. Reproduced from [82]. CC BY 4.0.

As another example, D^{2}NNs also present unique opportunities for designing novel computational camera systems with customized functions that cannot be implemented through standard optical designs. As an example, a new type of privacy-preserving imager was designed based on D^{2}NNs, which images only certain desired types of objects, while all-optically and instantaneously erasing other undesired types of objects from its output images (figure 11(c)).

Advances in science and technology to meet challenges

There exist, therefore, a wide range of high-potential applications. An important challenge remains to physically integrate the many values that define a NN's topology, i.e. the weights to its connections. In two dimensions, such physical implementations do not scale, and recent attention has shifted towards full-scale integration in 3D. Moughames *et al* [81] demonstrated optical convolutional filters integrated via 3D photonic waveguides (figure 12(a)), while some of the advances in the associated fabrication technology now allow for low-loss and large-scale 3D photonic integration of optical couplers [87, 88].

Finally, substantially improving the speed of NN computation requires implementing all the involved processes in parallel hardware, abolishing the slow-down induced through the large-scale serial communication used in schemes that do not leverage full parallelism on each computation stage. This was recently demonstrated by implementing an optical NN in the high-dimensional space of a multi-mode semiconductor laser diode, while trainable network weights were realized via a spatial light modulator.

As a consequence, both the linear operations through optimized weights as well as the nonlinear transformations happened in parallel and with approximately 20 GHz bandwidth [82], see figure 12(b).

Concluding remarks

In summary, using fully parallel optical NNs, computational imaging tasks can be executed at the speed of light propagation through compact, and potentially integrated photonic systems, establishing low-power and fast all-optical computing platforms beyond what existing electronic systems can offer. Free-space diffractive computing framework is scalable to different parts of the electromagnetic spectrum, including the visible and infrared wavelengths, which could find wide-ranging applications for high-throughput computing tasks and inspire the design of intelligent optical front-ends for advanced machine vision and communication systems. Three-dimensional (3D) photonic integration, on the other hand, provides the tools and techniques to enable scalable photonic integration of such concepts, while the use of high-dimensional nonlinear optical media such as semiconductor lasers unlocks another potential of photonics for computing: ultra-high speed. Despite these exciting developments and the unique advantages of optical networks, some challenges also remain that need further work and advances; for example, *in situ* training methods that are resilient to hardware and fabrication imperfections and potential misalignments and temporal variations in the optical set-up are highly sought with several innovative approaches emerging as potential solutions [89]. Similarly, the experimental data throughput and learning speed of physical systems based on optical hardware also need further advances to enable fast and periodic training for performance assurance (i.e. quality control) of optical network inference in the autonomous deployment stage of the system. These needs require a blend of innovations in both optical devices and algorithms, working hand in hand to address some of these emerging challenges.

Acknowledgments

The authors acknowledge the support of the Volkswagen Foundation, the region of Bourgogne Franche-Comte, the European Commission, and the US Army Research Office (ARO) and the National Science Foundation.

8. Quantitative microscopy

Daniel Midtvedt

Department of Physics, University of Gothenburg, Sweden

Status

Optical microscopy has had a tremendous impact on our understanding of the microscopic world and is widely utilized across many disciplines. The traditional workflow consists of (1) data acquisition, relying on manual adjustments of condenser lenses, illumination intensity and camera exposure time, (2) ocular inspection of the acquired microscopy images followed by (3) a qualitative assessment of the sample. The lack of standardization in this workflow makes optical microscopy a qualitative, as opposed to quantitative, technique.

To resolve this limitation, we first note that quantitative microscopy requires that uncontrollable and/or unmeasurable parameters of the optical system do not influence the measurement result. This restricts the measurable quantities to those that are agnostic to the parameters of the optical system. The set of such measurable quantities depends on the choice of microscopy method. The development of microscopy techniques and data analysis techniques go hand in hand: new microscopy techniques can extend the set of parameter-agnostic quantities, calling for new analysis techniques capable of quantifying them. In this context, DL-powered analysis has emerged as an alternative for fast, accurate, and automatized analysis and quantification of microscopy data, in alignment with broader efforts in quantitative computational imaging, such as E2E differentiable approaches discussed in section 2. As an example, the positions, and motion, of objects within the sample do not depend on the parameters of the illumination or the optical system. Particle tracking is, as such, a prime example of a parameter-agnostic technique for most microscopy techniques. It is therefore not surprising that large effort has been put into developing accurate and efficient tracking techniques. Here, DL-powered analysis has been demonstrated to surpass algorithmic techniques in terms of accuracy and speed while being fully automatized and objective [72].

Beyond this, interferometric techniques such as holography, interferometric scattering microscopy (iSCAT), and optical coherence tomography, have the potential to provide a quantification of the amount, and in some cases the angular distribution and/or the optical phase shift, of scattered light from objects. This can in turn be related to physical parameters of the objects in the sample such as their mass [90], size and refractive index [91]. Again, DL powered techniques have demonstrated superior accuracy and speed in quantifying such parameters, in particular at low signal-to-noise ratios (SNRs) [92].

Finally, multi- or hyperspectral imaging modalities, such as Raman imaging or Brillouin imaging, record a spectrum of optical responses in each pixel, generating enormous datasets. In this context, DL has provided a powerful method for automatized extraction of relevant data from such datasets [93].

Current and future challenges

Arguably, the holy grail within the field of quantitative microscopy is the accurate quantification of physical properties of specific components in a sample. Generally speaking, achieving this goal requires distinguishing the signal of the components of interest from that of other components (see figure 13 for a possible work flow). Doing this in a quantitative manner requires, in turn, that (**1**) sufficiently many physical parameters can be quantified, on an individual object level, to classify objects in the sample. In complex environments, as is often the case when considering biological samples, multiple components are often in close proximity and as a result the signals from such components are intertwined and not readily distinguishable. Therefore, in order to be useful, such specific imaging needs (**2**) to be robust to noise and (**3**) to feature a high spatial resolution (preferably sub-diffraction limited). Further, since components are often dynamic, it is also crucial (**4**) to have a high temporal resolution (1–10 ms).

More specifically, interferometric techniques typically display a high temporal resolution, meeting the requirement (**4**) above. It has also been demonstrated that DL-powered analysis techniques have the potential to quantify physical parameters of individual objects at high noise levels (**2**) and short length scales (**3**) (comparable to the diffraction limit) [92], much faster than traditional methods [91]. The set of parameters that can be quantified from individual nanoscale objects depends on the object size and imaging modality, and is currently limited to mass (for Rayleigh scatterers) [90], and size and refractive index (for particles larger than the diffraction limit) [92]. While these parameters arguably are key physical parameters in many cases, they may not be sufficient to do a robust classification and to perform

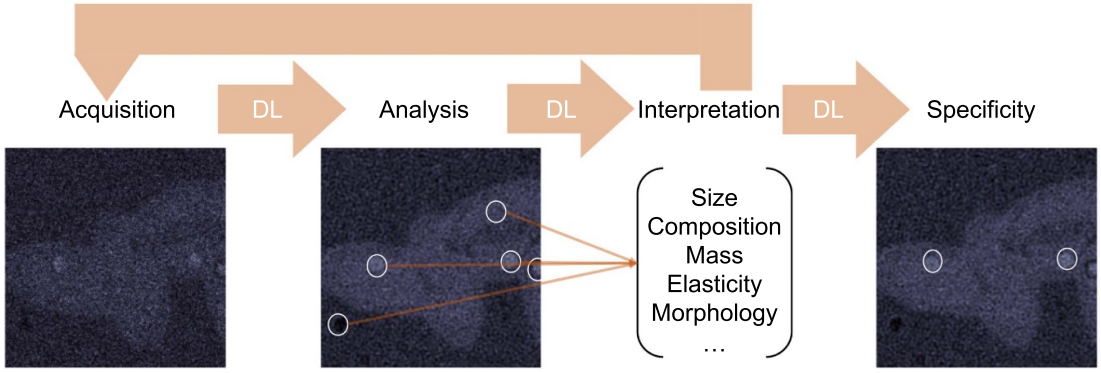

Figure 13. A tentative workflow for achieving quantitative microscopy with specificity. Starting from the acquisition, this must be sufficiently fast and display sufficient resolution to resolve the properties and dynamics of the objects of interest. Next, the data is analyzed, to enhance contrast and to resolve and detect objects in the data. This can be aided by deep learning (DL), as indicated by the arrow. Following object detection, a set of physical characteristics of the objects are quantified based on their optical responses, again potentially aided by deep learning. Advances in these steps in terms of noise resilience provides a feedback to the acquisition step, allowing faster acquisition (but also increasing noise in the data). Finally, specificity can be achieved by recognition of common characteristics of the objects of interest. Images are simulated using the software DeepTrack 2 [94].

quantitative imaging with high specificity. Extending this set of parameters without sacrificing spatiotemporal resolution is therefore a key challenge for these techniques (figure 13, 'Interpretation'). A particular challenge will be the quantification of objects larger than the Rayleigh limit, but smaller than the diffraction limit, as the relation between mass and light scattering is ambiguous in this regime and depends on the collection angle of the optical system and the internal mass distribution within the object [95].

For Raman and Brillouin imaging, the information content in each pixel consists of a frequency spectrum quantifying the local photon/phonon interaction. In this case the signal is in fact material specific, and can be used to quantify the abundance and distribution of different materials across a sample [96], resolving requirement (**1**) above. The remaining challenges in this field relate to the detection and quantification of individual objects in a complex sample (**2**) at high spatial (**3**) and temporal (**4**) resolution (figure 13, 'Analysis').

The key challenges are therefore:

1. extending the set of quantifiable parameters for interferometric imaging techniques, without sacrificing spatiotemporal resolution, and
2. improving spatiotemporal resolution and noise resilience of multispectral quantitative imaging techniques without sacrificing material specificity.

Advances in science and technology to meet challenges

The necessary technological and scientific steps in order to address these challenges are the following.

First, the amount of information in interferometric images can be enhanced through wavelength and/or modality multiplexing, which will ease the quantification of auxiliary physical parameters.

Second, the full information content of interferometric scattering patterns need to be utilized in the analysis. Specifically, the angular distribution of the optical field scattered from an object depends not only on its size, but also on its shape and internal mass distribution. This information is encoded in interferometric scattering patterns, but extracting it would generally require inversely solving Maxwell's equations. This is computationally expensive and extremely sensitive to measurement noise, in particular when approaching the Rayleigh limit. It has been demonstrated that DL-powered approaches enable much faster determination of object properties [91], while retaining accuracy even at low SNRs at sub-wavelength length scales [92]. Extending such quantification approaches to quantify also the morphology and internal mass distribution of objects covering a range of sizes from $<$100 nm to several micrometers will be a necessary step to achieve specific interferometric imaging.

For multispectral quantitative imaging, the key challenge is related to increasing spatiotemporal resolution. Recent works have demonstrated that stimulated Raman [96] and Brillouin microscopy [97] do have the potential for fast multispectral imaging, and work along those lines is expected to push the resolution of those techniques in the coming years. Integrating these imaging modalities with data-driven interpretation frameworks such as virtual staining (section 27) could further enhance specificity. DL powered analysis techniques can assist in this development by reconstructing physical parameters from

incomplete spectral information [96]. In this way, less data may be needed to reach the same level of specificity, thereby increasing temporal resolution further.

Combining such developments with advances in virtual staining, which has been demonstrated to recognize, classify and quantify structures in non-specific imaging modalities [98] is predicted to enable specific quantitative imaging through DL assisted analysis.

Concluding remarks

In conclusion, the big challenge in quantitative microscopy for the foreseeable future relates to combining quantitative measurements with having high specificity, thereby transforming such techniques into viable alternatives to fluorescence imaging. In this roadmap, I have discussed how DL powered data analysis may aid this development, highlighting some key steps that need to be realized in order to achieve this goal.

Acknowledgments

The author acknowledges funding from the Swedish Research Council, Grant Number 2019–05071.

9. Computational phase microscopy

Hao Wang[1], *Tongyu Li*[1] *and Lei Tian*[1,2]

[1] Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, United States of America
[2] Department of Biomedical Engineering, Boston University, Boston, MA 02215, United States of America

Status

QPI enables label-free imaging of transparent biological samples such as unstained cells and tissues. Many holographic-based phase imaging techniques have been developed to extract the phase information based on the principle of interferometry. However, these techniques generally suffer from complex instrumentation. In this perspective, we focus on computational phase microscopy techniques which recover phase information from a non-holographic setup using intensity-only measurements (alternative strategies are discussed in section 2 for learning with differentiable physical models). The main difficulty of reconstructing phase from intensity-only measurements is that it is generally an ill-posed inverse problem. Classical 'phase retrieval' algorithms are developed based on the Gerchberg–Saxton–Fienup algorithm, which uses an iterative reconstruction procedure and incorporates additional physical constraints to find the solution. In recent years, DL has shown tremendous success in solving such computational imaging problems. Here, we highlight several major advances on how DL pushes the imaging performance of computational phase microscopy, as summarized in figure 14.

In general, DL techniques have been developed to solve both the 2D phase recovery and 3D tomographic reconstruction problems, as summarized in figure 14. For 2D computational phase imaging, the first supervised learning-based reconstruction algorithm proposed in Rivenson *et al* [99] showed that the twin-image and self-interference related diffraction artifacts can be eliminated by a DNN. To push the performance at low-light conditions for 2D phase imaging, a DL technique has been developed to provide high-fidelity recovery when the photon numbers are low [100]. For 3D computational phase imaging, DL frameworks have been developed to solve the challenging 3D tomographic phase reconstruction from intensity-only measurements [104, 105]. In addition, frameworks have been developed to mitigate multiple-scattering effects and overcome the limitations in commonly used linear single-scattering approximation models [104, 105].

Although these supervised DL methods have made significant progress, they require large-scale training datasets containing paired measurements and ground-truth images. While early attempts focus on generating experimentally captured dataset [99, 100], in practice the accessibility to large-scale training data in experimental settings can be limited. To overcome this limitation, several innovative approaches have been developed. For example, computationally efficient and accurate simulations based on multiple-scattering models have been developed to generate synthetic training datasets for 3D problems [104, 105]. Instead of using the supervised DL framework, the unsupervised 'untrained neural network' framework based on Deep Image Prior [101] and Deep Decoder [102] have been adapted for 2D phase retrieval without needing any network pre-training nor training data. In this approach, the phase map is parameterized by an untrained network, whose parameters are optimized by matching the predicted measurement to the actual measurement using the known physical model via a standard iterative 'network training' procedure.

In addition to only performing the post-hoc 2D or 3D reconstruction tasks in computational phase imaging, DL methods have also been developed to co-optimize the physical design of measurement process along with the reconstruction. Pioneering work demonstrates an 'unrolled neural network' framework to optimize the coded-Illumination pattern and high-resolution 2D phase reconstruction using much reduced measurements than model-based method [19].

Current and future challenges

Despite achieving state-of-the-art performances in many tasks, multiple challenges remain to be solved both in terms of physical principles and computational frameworks, including:

1. **Multiple scattering effects and missing-cone problems**
   The multiple scattering effects become significant with the increase of refractive index contrast, structure complexity, and size of the biological samples. The measurement is often confounded by the 'missing-cone' problem, which does not provide access to a large amount of axial spatial frequency information. Together, these make the 3D reconstruction suffer from poor axial resolution

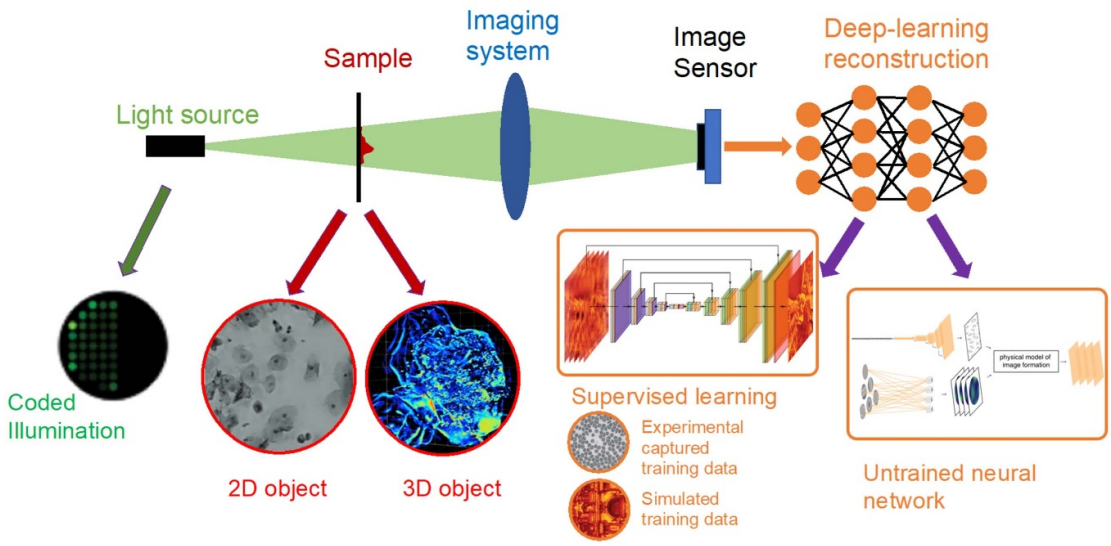

Figure 14. Overview of deep learning for computational phase microscopy. Deep learning has been used to recover both 2D [19, 99–103] and 3D [104–107] phase objects using both supervised learning [19, 99, 100, 103–106] and untrained neural networks [101, 102, 107]. For supervised learning, the training data is obtained either in experiments [99, 100, 103] or in simulations [19, 104–106]. Deep learning has also been applied to optimize the coded illumination patterns [19, 103]. The insets at the bottom are © [2019] IEEE. Reprinted, with permission, from [19]. Reproduced from [99]. CC BY 4.0. Reprinted with permission from [105] © The Optical Society. Reprinted with permission from [102] © The Optical Society.

and degraded reconstruction accuracy. Although DL frameworks have been developed to reconstruct highly scattering objects [104–106], their effectiveness is still limited. Similar challenges of data completeness and reconstruction under physical constraints also arise in quantitative microscopy (section 7) as well as other imaging modalities (sections 9–13).

2. **Reliable DL prediction**
   Another major limitation is that the trained DL model is often not robust to experimental variations. To overcome this issue, an uncertainty quantification framework based on the Bayesian learning framework has been proposed [103], which allows evaluating the confidence of the prediction result without knowing the ground truth. However, this type of Bayesian learning framework is still at its nascent stage and requires significant process to provide more reliable DL predictions.
3. **Large-scale computation**In DL-based reconstruction, the size of the NN generally increases as the size of the input images. Thus, the practically achievable space-bandwidth product (SBP) of the reconstructed image without excessive image stitching is limited by the memory of the computer. The associated computational cost becomes a bottleneck for emerging large-SBP imaging techniques, such as Fourier ptychographic microscopy for gigapixel 2D phase imaging [103] and diffraction tomography on large-scale 3D objects [105].

Advances in science and technology to meet challenges

Here, we outline a few promising directions to pursue to overcome the above challenges and further push the fundamental limit of DL for computational phase imaging.

1. **Prior knowledge incorporation**
   An overall strategy to overcome the multiple scattering and missing-cone problems as well as the reliability of DL prediction is to incorporate additional physical knowledge into the NN designs. In simulation-based training [104, 105], the multiple-scattering physics is incorporated into the simulator. In untrained NNs [101, 102], the physical model of the imaging system is incorporated during network training. In Bayesian learning [103], the priors are learned by a NN, which is then used to quantify the uncertainty at the test stage. A few other network designs implicitly take the physical insights into consideration. For example, reference [100] separately processes low and high spatial frequency information using a synthesis network. Kang *et al* [106] treats the sequential measurement as a dynamical system and solves the tomographic reconstruction problem by a RNN. We anticipate that hybrid strategies that incorporate multiple types of priors, e.g. multiple-scattering

physics, imaging model, and experimental data distribution, are promising directions to investigate in the future.

2. **Computationally efficient framework** Future development is needed to enable memory-efficient computational frameworks for large-SBP imaging applications. A promising solution is the neural fields (a.k.a. implicit neural representations) that parameterize the object by a coordinate-based deep network. The object is represented by the parameters of a small-scale NN, instead of the dense voxel/pixel grids. Promising results using this framework have been recently demonstrated for large-scale 3D phase recovery in Liu *et al* [107] and dynamic imaging in Cao *et al* [108]. We anticipate novel neural representations for efficient information embedding may be a promising direction to pursue for large-SBP reconstructions.

10. Multimodal microscopy image registration

Nataša Sladoje and Joakim Lindblad

Department of Information Technology, Uppsala University, Uppsala, Sweden

Status

The technological advancement enables highly informative imaging, allowing scientists to see what previously was hard to even imagine. The vast collection of imaging techniques used in the biological and preclinical areas indicates the breadth of possibilities [109]. Powerful and complex devices reveal a variety of properties of a specimen, such as morphological structure, chemical composition, dynamics, function; however, most often they reveal only one such property at a time. We are witnessing a growing popularity, particularly within life sciences, of *correlative multimodal imaging* (CMI) approaches, which combine complementary information from different imaging modalities to create a holistic, composite view of the sample, maximizing the extracted information about an object of interest [110]. The most well-established CMI technique is CLEM, combining spatial and functional information within the subcellular context.

To enable joint analysis and fusion of the heterogeneous information captured by different devices, the first requirement is to establish precise geometric correspondence between the acquired images, i.e. to find a spatial transformation which best aligns the data in the same coordinate system. This process, known as *image registration*, is traditionally performed as an iterative optimization process, aiming to find the transformation which maximizes a similarity measure between the images to align. Depending on the application scenario, different types of transformations (e.g. rigid, affine, deformable) are used, making the task more or less constrained. The optimization is typically highly non-convex, which makes the process complicated and time consuming, while still often only delivering suboptimal solutions. This in particular holds for multimodal registration, where the images that need to be aligned, aiming to capture complementary information, often look very different from each other.

To solve or circumvent the observed problems, researchers have turned to approaches based on DL, mirroring the broader trend toward hybrid model- and data-driven frameworks described in section 4 for PnP computational imaging. Three main directions have been followed: (i) learning a suitable similarity measure, (ii) speeding up the registration process by directly predicting the transformation, and (iii) reducing multimodal alignment to a monomodal task, by applying image-to-image (I2I) translation techniques. These approaches have been evaluated on medical data [111–113] and the most successful ones are steadily growing in popularity. However, their use in biomedical/microscopy image registration is still very limited. In particular, DL-based multimodal microscopy image registration has been attempted in only a few works [114, 115]. The extraordinary boost, observed for other image analysis tasks, such as image segmentation and classification, is still lacking—the DL-revolution is yet to come to multimodal microscopy image registration.

Current and future challenges

To reach, and benefit from, the full potential of DL in multimodal microscopy image registration, researchers need to respond to several challenges (figure 15):

Challenge 1—DL methods still do not meet the expectations in image registration

Considering that different combinations of modalities may require different measures of image similarity, the idea to learn such a measure from data comes naturally. Given sufficient annotated training data, proposed learned similarity measures perform well. However, they are most often used to replace conventional measures within a slow iterative registration procedure, which reduces the benefits.

Once trained, DL models are typically very fast and therefore appealing to use as regressors, to directly predict the transformation parameters to reach image alignment. This approach is taken by currently most popular methods which use DL for registration. While considerably reducing runtime, they still fall behind conventional iterative methods in terms of accuracy.

I2I translation approaches, typically based on generative adversarial networks (GANs), perform impressively in learning to mimic, and combine, content and styles captured by image data, e.g. enabling 'virtual staining' of label-free tissue. However, generated images, even if convincing in appearance, often do not preserve well the information important for image alignment [115]. Relying on aligned pairs, the coordinated representation learning approach [114] delivers superior performance (figure 16).

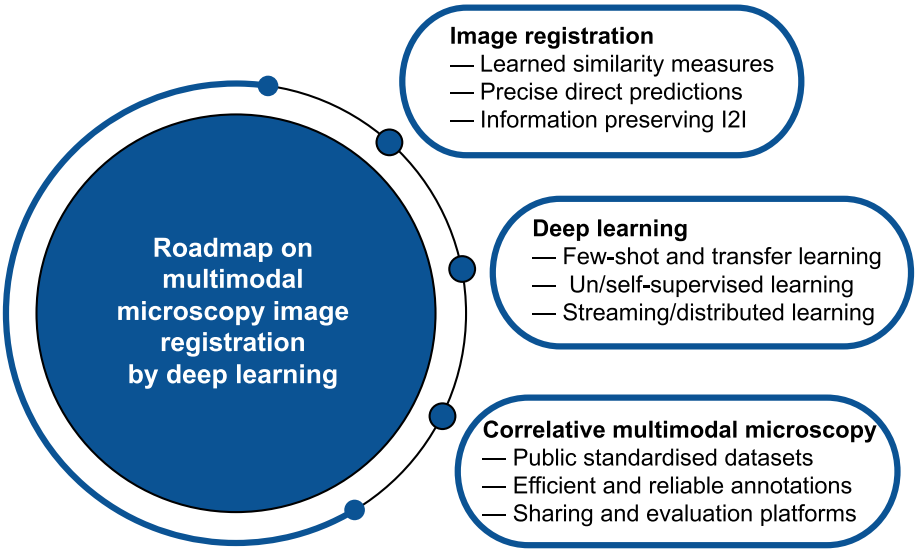

Figure 15. Needs and challenges. Overview of needs and challenges for multimodal microscopy image registration by deep learning.

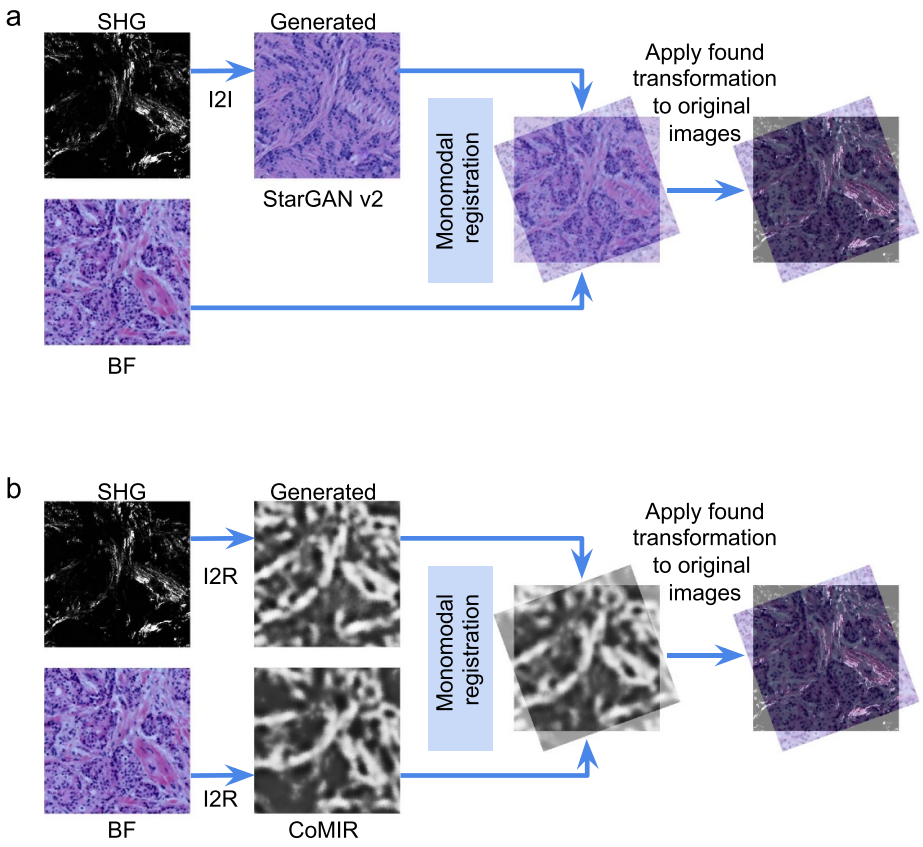

Figure 16. Multimodal image registration. Multimodal image registration can be addressed by converting the multimodal problem into a, presumably less challenging, monomodal one. Here, this is illustrated on the difficult task of aligning Second Harmonic Generation (SHG) and BrightField (BF) microscopy images. (a) One approach is to synthesize images acquired by one modality to appear as if acquired by the other; this type of image-to-image (I2I) translation is typically done by generative adversarial networks (here, using StarGAN v2). However, a comparative performance analysis [115] indicates that the synthesized images, even when convincing in appearance, seem to lack sufficiently reliable features to ensure successful subsequent monomodal alignment. (b) Another approach is to, instead of learning the appearances, learn representations (I2R) of the joint content of multimodal images, transforming them into a common 'virtual' modality. This coordinated representation learning approach, in combination with a suitable monomodal registration method, reaches excellent performance on alignment of SHG and BF images [114].

Challenge 2—DL methods require a lot of data, but not any data

Correlative multimodal microscopy typically results in few very large images. This is far from the millions of relatively small annotated images collected in publicly available datasets which are used for development of state-of-the art DL models in CV. Handling such massive, high-dimensional data volumes aligns closely with the computational bottlenecks discussed for large 3D image reconstruction in section 3. The complexity of the acquisition process, typically requiring significant manual labor, combined with exploratory aims of the performed research, implies that images acquired are often counted in single-digit numbers. Transfer learning, aimed at reusing trained models on new datasets, often delivers diminishing returns on the diverse and heterogeneous biomedical image data. At the same time, very high-resolution imaging provides terabyte-sized (or even larger) image volumes, necessitating advanced algorithmic solutions for their processing.

Challenge 3—annotated multimodal microscopy data are critically lacking

To ensure reliable performance, DL approaches generally require large amounts of annotated training data. Annotation of biomedical data requires not only extensive time, but also considerable expertise, making it costly to collect. DL models which require aligned image pairs for training find limited use, since the alignment is, for many modality combinations, simply too difficult to be performed manually. Few-shot and un/self-supervised methods may reduce the need for training data, but this often comes at a cost reduced performance.

Advances in science and technology to meet challenges

Availability of high quality training data is identified as a *sine qua non* condition for successful DL approaches. Notably, similar issues about the critical need for annotated datasets also arise in high-content screening and biomedical imaging applications (section 30), highlighting the broader ecosystem challenges. The challenges of DL-based multimodal microscopy registration all relate to the need for, and difficulty to provide in sufficient amounts, accurately aligned multimodal microscopy image pairs for diverse combinations of modalities. This indicates two directions forward:

i. develop methods which reach good performance while requiring only few, or no aligned image pairs;
ii. collect large annotated multimodal microscopy datasets and make them broadly available.

Advances in methodological development

Iterative maximization of mutual information (MI) is still the most popular multimodal registration method; it is generally applicable, performs reasonably well, and does not require any training data. A recently proposed FFT-based algorithm [116] advances MI-driven rigid multimodal registration, both in terms of speed and accuracy, outperforming DL-based competitors on two multimodal microscopy datasets. Novel DL-based registration methods need to deliver more.

Requirements for extensive training data of aligned multimodal image pairs need to be reduced, or removed. We need to turn to novel and innovative learning strategies which deliver highly performing, yet scalable solutions. General improvement of data-efficient DL strategies, such as few-shot learning, domain transfer, and self-supervised learning, will advance multimodal microscopy image registration as well.

Unsupervised methods, not requiring any aligned image pairs for training, are already available. Examples include I2I translation approaches based on unsupervised GANs [115]. However, their performance needs to be further improved.

Advances in data collection, open science, and standardized benchmarks

Annotated (ground truth) data is needed not only for DL model training, but also for evaluation of novel methods. It is therefore of critical importance to assemble and publish high quality datasets, enabling quality control and reproducibility. Automated approaches may reduce need for manual annotation, e.g. by generating landmarks through segmentation of identified common structures in multimodal data.

Integrated imaging, where different modalities are simultaneously acquired and therefore aligned, may be a rewarding path for generating large amounts of high-quality ground-truth registration data, for both training and performance evaluation. This option is, however, only available for a few combinations of modalities. Advances in integrated imaging will at the same time reduce the need for multimodal registration methods.

Considering the specificities and large size of multimodal microscopy datasets, availability of suitable data sharing platforms is a necessity. Standardization of acquisition processes will contribute to improved

quality of data and increased homogeneity (e.g. within a particular modality), leading to increasingly successful training of DL models and improved registration quality. Open platforms for standardized method evaluation will boost the quality of novel approaches. A recent contribution is an open evaluation framework for rigid multimodal registration methods, [115].

Concluding remarks

In the era of outstanding performance of DL methods, which continuously advance image data analysis, most popular methods used in practice for multimodal microscopy image registration still rely on semi-automatic approaches. A successful example is eC-CLEM [117]—relying on user-interaction, it has demonstrated applicability to registration tasks involving a wide range of modality combinations. However, the increasing diversity of multimodal microscopy image registration problems and the growing scale and dimensionality of acquired data urgently call for the power and flexibility of efficient data-driven approaches, to bring the quality of the methods and analysis results to the next level. By addressing the identified challenges of the field (overviewed in figure 15), DL-based techniques have the potential to deliver modality-agnostic, fast, and generally applicable image registration solutions, to advance CMI, and ultimately—life sciences.

Acknowledgments

The authors are inspired by the numerous activities and initiatives of the COST consortium 'Correlated Multimodal Imaging in Life Sciences' (COMULIS)—a collaborative European network aiming to foster interdisciplinary communication and collaboration on all aspects of Correlative Multimodal Imaging in Life Sciences. The authors are financially supported by VINNOVA (MedTech4Health project 2017–02447 and 2021–01420), Swedish Research Council (Grant 2022–03580) and Cancerfonden (projects 22 2353 and 22 2357).

11. Fluorescence lifetime imaging (FLI)

Jason T Smith[1,2], *Marien Ochoa*[1,3], *Margarida Barroso*[4] *and Xavier Intes*[1]

[1] Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, United States of America
[2] ECS Federal, Fairfax, VA 22031, United States of America
[3] University of Wisconsin–Madison, Department of Medical Physics, Madison, WI 53705, United States of America
[4] Department of Molecular and Cellular Physiology, Albany Medical College, Albany, NY 12208, United States of America

Status

FLI provides distinctive contrast mechanisms for the interrogation of biological samples. A key strength of FLI is that it can uniquely provide absolute measurements directly related to the fluorescent molecule(s) state and its/their interaction with the surrounding molecular micro-environment: such as temperature, pH, viscosity, polarity and mechanical forces. The principles and various technical implementations of FLI have been established over the last three decades, but until recently, have typically remained an expert field. With the advent of turn-key commercial FLI capable imaging platforms, FLI is currently being increasingly embraced by end-user communities (e.g. molecular biologists, drug development experts) with demonstrated increased utility in microscopic and macroscopic preclinical and clinical applications. Still, a main challenge in FLI resides in the estimation of the lifetime(s) or associated parameters. This is a complex computational task, in which accuracy can be highly dependent on the model selected, the set of parameters used, and the SNR of the acquired measurements (typically a photon starved application, hence, low SNR). Therefore, there are still large efforts focusing on providing robust, user-friendly, and accurate methodologies, including DL models, for FLI quantification and image formation. The use of ML approaches for FLI image formation was first using an ANN approach with promising results [118]. This work was followed by the first DL model, FLI-NET, designed to simultaneously produce 2D images of all the lifetime-based parameters associated with a double exponential model [119]. FLI-NET was validated both with microscopic and preclinical data sets and for two main instrumental detection technologies, time-correlated single photon counting (TCSPC) and gated ICCDs. Since then, an increasing number of reports have demonstrated the potential of DL models to accurately estimate lifetime parameters without any user input, over large FOVs, with extremely fast inference times (at, or close to real-time), and with better performances at low photon counts. There is also an increased interest in developing E2E DL models for pixel-wise classification based on spatio-temporal inputs. An example of an ANN based FLI image formation over a large FOV is provided in figure 17.

Current and future challenges

As the field of DL for FLI continues to mature, it is facing the same challenges currently encountered in the development and validation of DL models for biomedical imaging at large. These can be summarized as: the acquisition/availability of large representative data sets, the generalization of DL models, and instilling explainability and/or trustworthiness of the DL model output. In numerous fields, DL successes at large can be attributed to the confluence of increased computational power with accessibility to large data sets. However, for biomedical applications, large data sets that have been thoroughly curated are rarely available to the community. This is particularly true for FLI. To date, most of the work reported has implemented experimentally representative data simulation routines to generate large data sets for efficient DL model training and validation. This approach is cost effective and has been demonstrated to perform well for processing experimental data not used during training with high accuracy. This mitigates the requirements of large experimental data sets for training, but still requires expertise in developing the simulation environment that closely replicates the specificity of the application—such as instrumental characteristics (especially the instrument response function, IRF, which characterizes the temporal behavior of the system), lifetime-based parameters range and noise distributions.

Though, there are additional factors that can affect FLI quantification from experimental data and that can be challenging to represent during the training phase. These include laser jitter and instrumental drift, changes in sample to detector distance, pixel-dependent variation of the IRF, photobleaching of the fluorophore, saturation of the detector. Other factors include simulation model bias such as imaging a sample with fluorescence lifetime outside of the range of that used during model training and more complex signal signatures such as multi-exponential features beyond bi-exponential behaviors. All and each of these factors can negatively affect the model output independently. This also highlights the

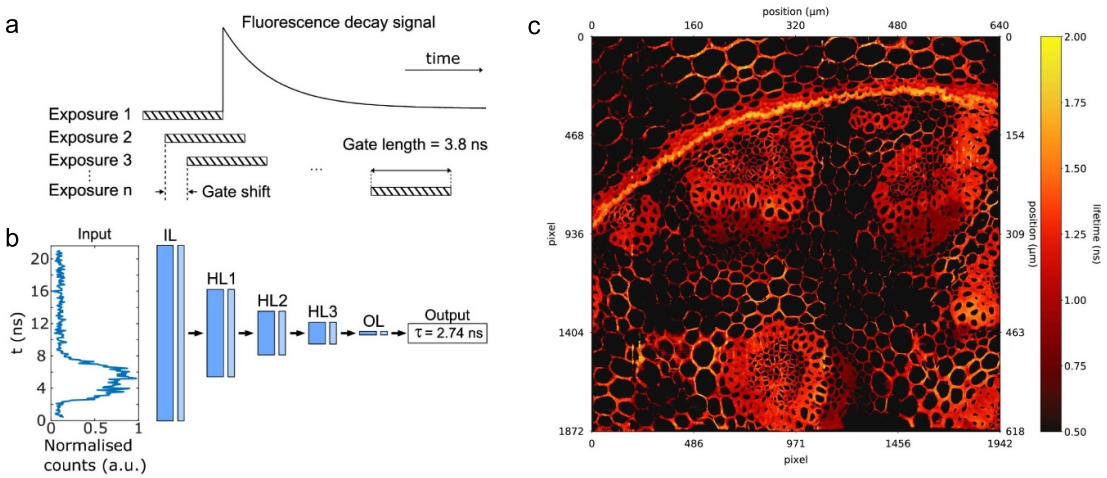

Figure 17. FLI quantification with ANN. Diagram of FLI quantification using an ANN trained with simulated data (a), (b). (c), Example of model used for lifetime quantification of a Convallaria sample of high spatial dimensionality (1875 × 1942 pixels). The total processing time of the ANN was approximately 36 s. Reproduced from [120]. CC BY 4.0.

difficulty to generalize the developed DL models beyond the imaging system and application at hand. FLI-Net, the first DL model applied to different FLI applications and technologies, needed to be trained specifically for each case. This comes with an added computational burden comparatively to classical iterative fitting approaches. Additionally, this leads to assessing the trustworthiness of the DL model output. In this regard, classical iterative fitting-based approaches can assess the fidelity of the quantification, as well as the quality of the FLI data itself, through residual error between the data and the approximated fits. This is an important means for quality assessment that is commonly employed during FLI analysis, as FLI data can be comprised of pixels with decays that are not suitable for quantification due to a range of factors (e.g. low SNR, motion artifacts, improper parameter settings). When using DL for either image formation or for classification, these poor-quality decays can be either removed or artificially enhanced via rudimentary pre-processing steps prior to network inference. However, this preprocessing step precludes the application of DL models towards real-time and/or can lead to large bias as it enforces expected features in the data that could not always be valid (for instance bi-exponential behavior while more complex biological distribution is present in the sample).

Advances in science and technology to meet challenges

The use of DL methodologies for FLIM is a nascent field that promises to greatly increase the widespread utility of lifetime-based contrast in biology, as well as its use in translational medicine. To date, a pioneering set of reports have laid the foundation for the development of efficient DL models that are dedicated to specific applications and technologies. One important aspect is that, in many cases, experimental validation of DL models has been primarily performed with relatively bright and long lifetime samples that are 'best case' scenarios. Still, it can be argued that the biggest challenge in the field of lifetime imaging at present is that of low photon count detection with high background noise levels. Indeed, proper FLI data acquisition requires many photons to provide high quality decays in all pixels. However, low counts often force the use of high exposure times or illumination powers that can lead to fluorophore photobleaching or cell/tissue damage. Low counts also lead to increased binning resulting in reduced resolution [121]. However, collection of suitably high counts is oftentimes unfeasible when dealing with sensitive samples or with applications requiring even modest (i.e. sub-second) framerates. FLI-Net was the first reported model to exhibit increased accuracy in low photon regime compared to traditional iterative fitting techniques. Since, numerous studies have focused on establishing improved DL models for this specific scenario across various applications [121–123]. In parallel, next-generation time-resolved detectors, such as single-photon avalanche diode (SPAD) arrays, are poised to enable faster and more photon-efficient FLI data acquisition with improved SNRs [124]. Furthermore, next generation microscopy systems that leverage computational imaging approaches like single-pixel detection will allow for improved collection efficiency especially when coupled with DL models [125]. Also, the implementation of detectors that leverage additional information content (e.g. hyperspectral detection arrangements) will allow for increased specificity, especially for conditions where spectral emissions from target fluorophores are hard to isolate due to spectral bleedthrough [126]. Additionally, there are still large

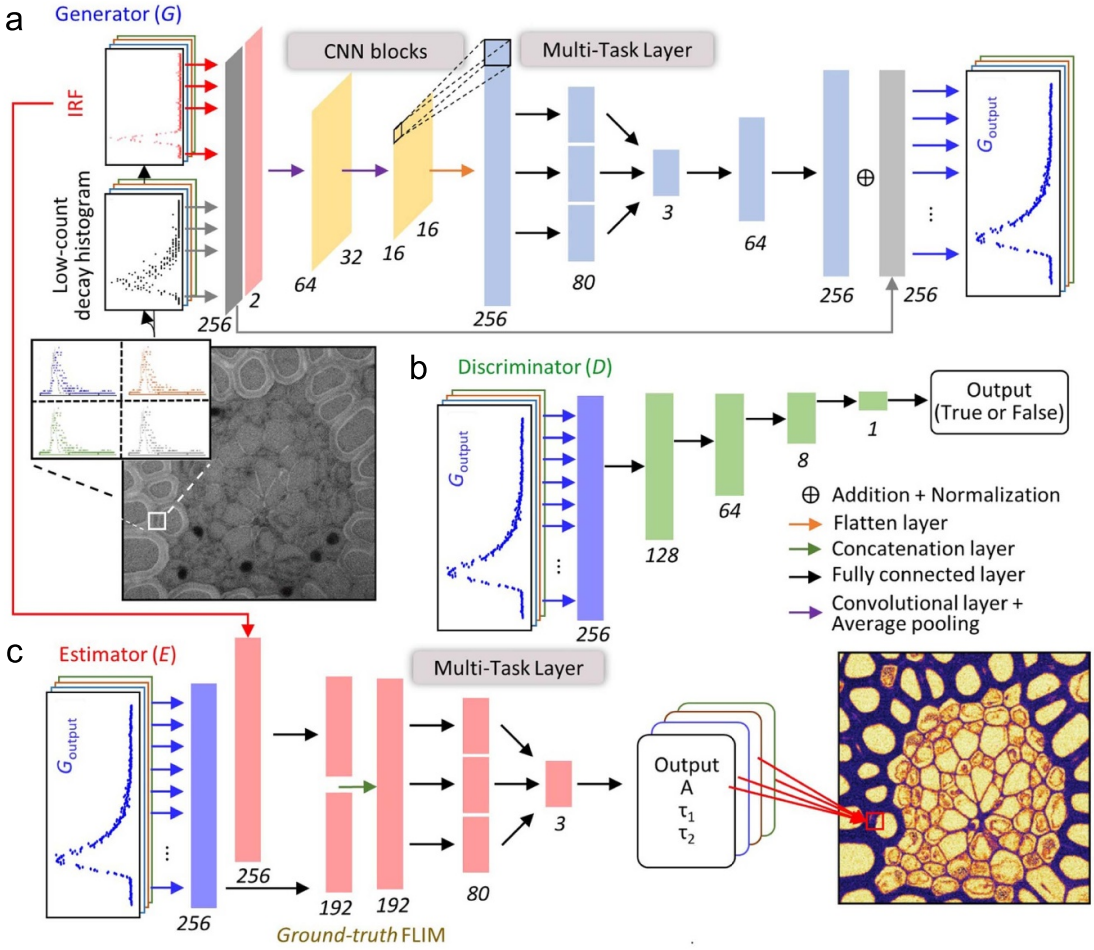

Figure 18. Diagram of flimGANE (fluorescence lifetime imaging based on generative adversarial network estimation). Low SNR fluorescence decay input is denoised, and potential laser jitter is accounted for, by the generator (a). The discriminator (b) enforces high-quality decays by an additional loss term. The estimator (c) then uses the denoised output of the generator to retrieve lifetime quantification. Reproduced from [128]. CC BY 4.0.

efforts focusing on developing more stable exogenous fluorophores with high quantum efficiencies and low cytotoxicity for improved FLI signal detection.

On the account of output trustworthiness, there has been an increased number of tools available for eXplainable AI (XAI) [127]. Though, if current methods can be leveraged efficiently in FLI classification tasks, there is still a lack of appropriate tools in DL image formation. To date, DL model outputs must be assessed by an expert and unexpected results, forensically investigated. From experience, these unexpected DL results are typically attributed to some variations in the experimental acquisition parameters. Such variations can be mitigated by employing generative models. For instance, flimGANE takes into account the noise distribution as well as laser jitter in FLI data collected by TCSPC (figure 18 [128]). In turn, this enables the development of DL models that are more generalizable. Still, they typically come with an additional computational complexity that does not make them competitive for fast inference, which is required in certain clinical scenarios. In this regard, simpler DL models have been proposed that process data for individual pixels with potential to be implemented directly on the acquisition hardware. Coupled with efficient training methodologies, they herald embedded hardware implementation, coupling with sensors and readout circuits to achieve fast on-chip training and inference [129, 130]. Especially, by integrating computationally efficient DL workflows directly onto FPGAs, real-time edge computing FLI can be achieved [131].

Last, to maximize reproducibility and accessibility across the FLI community, the development of user-friendly open-sourced software will have a large and significant impact (see also sections 32–35). Providing a workflow that allows for deconvolution of the end-users IRF (i.e. the detector-specific response) [132] along with subsequent data simulation and model training across wide parameter bounds could provide a cross-reference platform for benchmarking. Combined with the mandate from

institutional funding agencies to support the release of publicly available data, it will elicit even further developments in the field by opening it to the computer science community.

Concluding remarks

FLIM has firmly established itself as a pillar of cellular imaging from microscopic to macroscopic platforms. However, FLI utility is associated with parameter(s) quantification obtained via computational approaches, which has reduced its impact and wide-spread use due to the expertise requirements. The implementation of DL methodologies is expected to enable real-time FLI quantification free of user-related bias. Beyond simplifying and enhancing the data processing pipeline, DL is poised to greatly impact FLI imaging protocols and imaging platforms. Due to its enhanced robustness at dim signals, DL is expected to relax the need of long integration time, spatial binning, and increased illumination power. Altogether, DL-enhanced FLI will be better-positioned to impact applications ranging from fundamental biology to clinical practice.

12. Multi-modal nonlinear microscopy

Tong Qiu, Li-Yu Yu and Sixian You

Research Laboratory of Electronics, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge MA, United States of America

Status

The development of nonlinear optical theory and microscopy has enabled unprecedented opportunities to look at living cells, tissues, and animals with submicron resolution in real-time. Over the past two decades, a variety of biological phenomena have been investigated using images based on the fluorescence excited by multiphoton processes, harmonic generation from specially structured molecules, and chemical profiling based on coherent Raman scattering. Integration of these modalities has been shown to provide metabolic, chemical, and structural profiling of cells in the context of living tissues. Despite its strong promise, multimodal nonlinear microscopy has not yet reached its full potential. Improvements in hardware and software are needed for further translation to biomedicine. These challenges and opportunities are similar to those of traditional microscopy modalities explored in sections 7–10, 12, and 13. This perspective focuses on the important gaps that can be potentially addressed by DL from two angles: (1) how DL can help overcome the technical limits of multimodal nonlinear microscopy through DL-based reconstruction and augmentation, and (2) how DL can help increase the clinical and biological relevance of multimodal nonlinear microscopy through DL-based image and video understanding.

Current and future challenges

One challenge lies in the longstanding technical limitations of nonlinear microscopy, such as the penetration depth, imaging speed, signal strength, and resolution associated with *in vivo* imaging. Various hardware improvements have been proposed in recent decades, including adaptive optics, adaptive laser sources, and fiber-based endoscopy for deeper tissue imaging, multi-foci and wide-field temporal focusing for higher speed, and pulse shaping for stronger signals. However, optimization of one of these parameters based on hardware is usually associated with the degradation of performance in the other parameters. For example, pulse shaping boosts signal generation efficiency by compensating for dispersion at the expense of laser power and cost. Computational reconstruction, such as optimization and learning-based algorithms, provides a promising alternative to overcome such inherent tradeoffs by regularization based on principle-based or learned data priors.

The other challenge is the automated image and video understanding of multimodal nonlinear microscopy. The information captured by multimodal nonlinear microscopy is a pixel-coregistered multimodal quantitative measurement of fluorophores (two-photon, three-photon absorption fluorescence), molecular structures (second, third-harmonic generation), and chemical bonds (coherent Raman anti-Stokes scattering, stimulated Raman scattering). Despite its rich information, the translation of the multimodal information to biological and pathological analysis is not yet readily accessible to biologists and clinicians who are experts in immunohistopathology. Algorithms that can address the challenges of image and video understanding of multimodal nonlinear microscopy are in great demand. One direction is to directly transform multimodal nonlinear microscopy to hematoxylin and eosin (H&E)-like images, which facilitates the biomedical relevance of the new image dataset. Cahill *et al* have shown a color metric-based method that reliably translates nonlinear microscopy images to H&E images [133]. The other direction is the direct extraction of quantitative information for specific applications. For example, Walsh *et al* performed classification and single-cell analysis of quiescent and activated T cells using quantitative features of nonlinear autofluorescence microscopy, which revealed the correlation between the autofluorescence features and the metabolism of T cells [134]. These insights establish robust protocols and propose new mechanisms but rely on users' mastery of both nonlinear microscopy and the specific biomedical applications. DL methods promise to alleviate the burden of domain expertise and further streamline the process via data-driven and E2E learning.

Advances in science and technology to meet challenges

As mentioned in the previous section, computational reconstruction is uniquely positioned towards pushing the physical limits of nonlinear microscopy via proper data priors. For example, to faithfully reconstruct wide-field multiplexed measurements in deeper layers, Zheng *et al* [135] leveraged temporal focusing for randomly patterned wide-field illumination and physics-based inversion for computational reconstruction of the multiplexed signals (figure 19(a)). To enable video-rate multiphoton

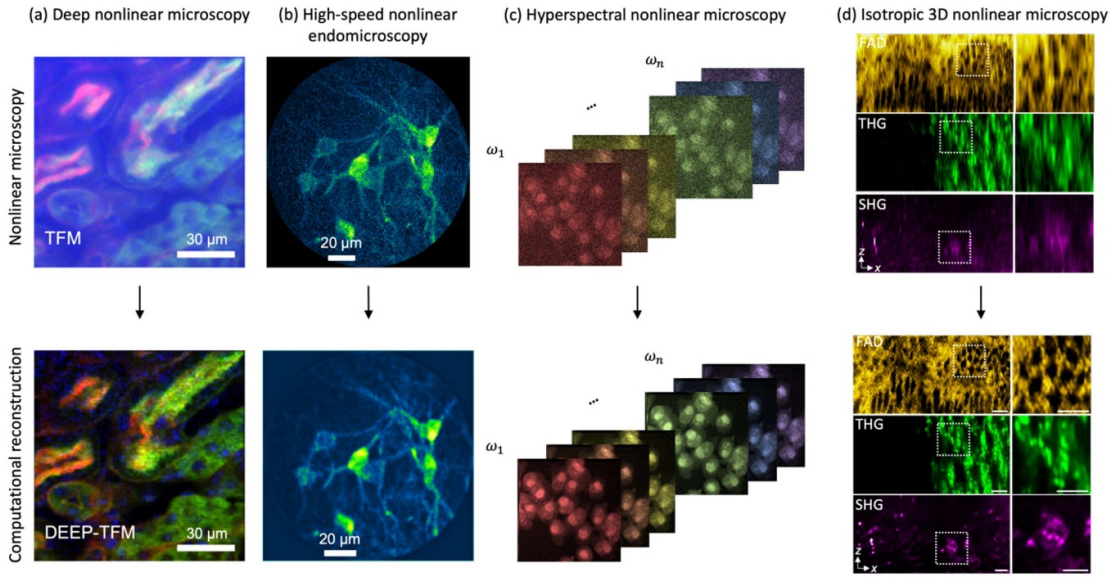

Figure 19. Examples of how computational reconstruction improves nonlinear microscopy images. (a) Deep wide-field two-photon microscopy of ex vivo muscle tissues with patterned illumination and computational inversion. Reproduced from [135]. CC BY 4.0. (b) Video-rate *in vivo* imaging of neurons in freely behaving mice using two-photon fiberscopy and two-stage learning transfer strategy. Reproduced from [136]. CC BY 4.0. (c) High-speed high-fidelity hyperspectral stimulated Raman spectroscopic imaging of biomolecules with deep learning-enhanced signals. Reproduced from [96]. CC BY 4.0. (d) Isotropic three-dimensional label-free nonlinear microscopic imaging via physics-informed deep learning axial deblurring [139].

endomicroscopy, Guan *et al* [136] devised a two-stage learning transfer strategy to generate augmented training datasets that would be otherwise challenging to obtain experimentally *in vivo*, which effectively recovered the loss of SNR and resolution associated with a high frame rate (figure 19(b)). Towards high-speed high-fidelity hyperspectral nonlinear microscopy, Lin *et al* [96] achieved ultrafast tuning based on a polygon scanner and enhanced the weak Raman signals via a spatial-spectral residual learning network (figure 19(c)). Weigert *et al* [8] and McAleer *et al* [137] showed promising results in high-quality image reconstruction with reduced light dosage or acquisition time via learning-based algorithms. Fan *et al* demonstrated an enhanced second harmonic imaging using a DL decipher for efficient and resilient phase retrieval [138]. Towards isotropic three-dimensional imaging, Han *et al* [139] leveraged a sampled-informed synthetic dataset to adapt a pre-trained deblurring network to axial resolution recovery in label-free nonlinear microscopy (figure 19(d)). Although these computational strategies have the potential to bypass the physical limitations of nonlinear microscopy, a few challenges remain open to more investigations, including how to obtain ground truth data in demanding applications, how to avoid hallucination in data-driven reconstruction, and how to avoid data-driven bias in learning-enabled biological discoveries and clinical diagnosis.

Towards better image and video understanding of multimodal nonlinear microscopy, DL is a promising tool due to its E2E data-driven principle, i.e. not relying on hand-crafted features, and its demonstrated capability of generalization to unseen data. Rapid advances in CV pave the way for numerous applications in optical microscopy, and many methods can be adapted to nonlinear microscopy analysis such as virtual staining [140–142], cell segmentation [143, 144], and cancer classification [145, 146] (see also sections 27 and 30). For example, to generate histology-like images, Sun *et al* [141] have demonstrated virtual H&E staining based on SLAM images, segmentation NN, and color translation metrics (figure 20(a)). To carry out single-cell analysis of the tumor microenvironment, You *et al* [143] performed multiclass pixel-wise segmentation using a supervised U-net (figure 20(b)). To enable real-time intraoperative assessment, Hollon *et al* [145] demonstrated real-time tumor detection based on stimulated Raman histology images using a CNN-based architecture (figure 20(c)). Despite these encouraging results, many real-world biomedical applications face the scarcity of adequate, curated training datasets, which poses a significant problem for supervised learning. To tackle this challenge, Shi *et al* [147] have demonstrated weakly supervised learning to extract conventional and unconventional cancer biomarkers from the optical signatures obtained by multimodal nonlinear microscopy. While these DL-based approaches open a wide array of applications, several challenges need to be addressed in the future. First, to alleviate the requirement of extensive training datasets, weakly supervised or self-supervised methods are in great demand. Generalizability is another issue. It remains a challenge to deploy well-developed

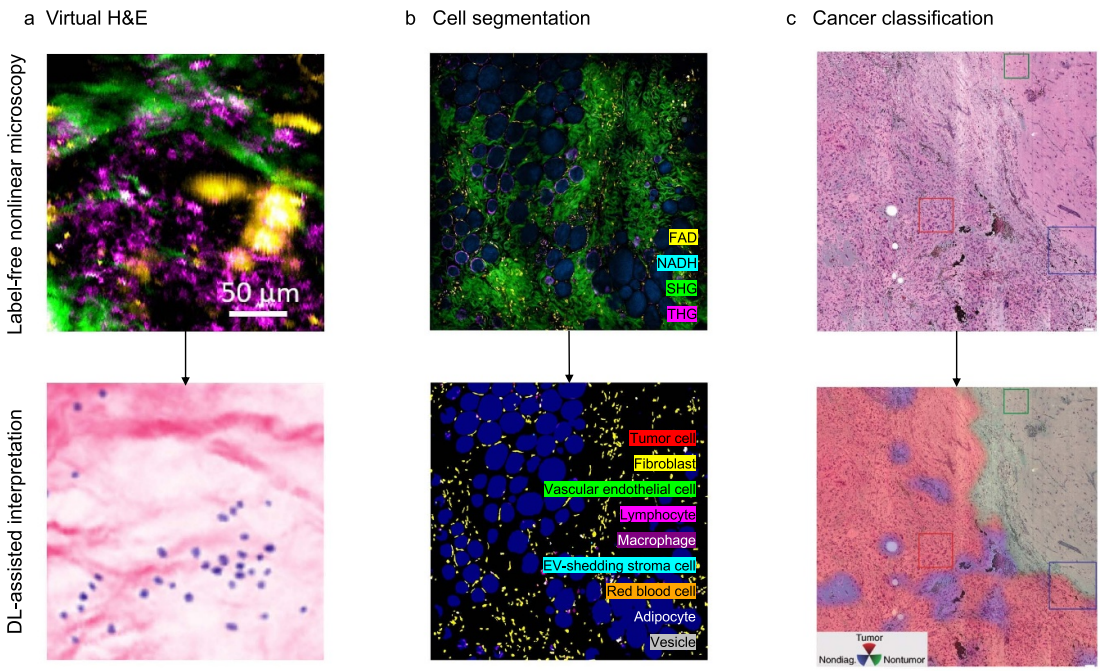

Figure 20. Deep learning-assisted biomedical imaging processing and analysis in multimodal nonlinear microscopy. (a) Virtual H&E staining of living rat mammary tissue [140]. (b) Segmentation of cells and extracellular for characterization of the tumor microenvironment [141]. (c) Diagnosis of brain tumor from stimulated Raman histology images [142]. (b) Reproduced from [143]. CC BY 4.0. (c) Reproduced from [145], with permission from Springer Nature.

methods trained on one dataset to other different biomedical applications. You *et al* [148] attempted to address the issue of quality discrepancy between lab-based data and intraoperative data through physics-based data augmentation. More efforts are needed to develop algorithms that can easily adapt to microscopy systems and user differences.

Concluding remarks

Multimodal nonlinear microscopy has become an indispensable tool for high-resolution imaging of living biological systems. Rapid advances in DL are and will keep being leveraged to push the limits of multimodal nonlinear microscopy. This perspective has focused on how DL helps push the technical limits of multimodal nonlinear microscopy as well as enhancing its biomedical translation and impact. Looking forward, we expect these efforts will accelerate the clinical translation and biological discovery based on nonlinear microscopy with caution and innovation. There will also be other exciting work showing how DL can further advance the design of nonlinear microscopy, such as learning-based imaging system design, as demonstrated in the prospering community of computational photography and microscopy.

Acknowledgments

T. Q., L. Y., and S. Y acknowledge the support of Jameel Clinic, Scialog RCSA Award, Amazon Research Award, and MIT.

13. Automated scanning probe microscopy (SPM)

Yongtao Liu[1], *Maxim A Ziatdinov*[1,2] *and Sergei V Kalinin*[3]

[1] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States of America
[2] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States of America
[3] Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37996, United States of America

Status

Thirty-five years after the development of AFM by Binnig and Rohrer [149, 150], SPM has become the mainstay technique in areas ranging from condensed matter physics and materials science to biology and medicine [151]. The highly robust nature of SPM has spawned multiple types of machines operating in the ambient environment, controlled atmosphere, liquid, and ultrahigh vacuum. Combined with the electrical, chemical, magnetic functionalization of the probes, the gamut of SPM imaging modes is truly broad. Complementing imaging, many SPM techniques allow a broad variety of spectroscopic measurements, ranging from the force–distance curves in conventional AFM, current–voltage curve in scanning tunneling microscopy (STM), and a broad variety of time- and voltage-spectroscopies in techniques such as piezoresponse force microscopy (PFM) [152]. The development of the SPM modes is seamlessly tied to the priorities of the R&D community, with the continuous growth wave of magnetic force microscopy in the late 1990s being driven by the magnetic hard drive industry, development of PFM stimulated by the ferroelectric non-volatile memories and tunneling barriers and recently discovery 2D ferroelectricity, and force–distance measurements providing insight into statistical physics of biomolecules. It can be argued that exponential growth of experimental effort in materials for energy storage and conversion will guide the SPM progress over the next decade. Here, techniques such as electrochemical strain microscopy [153] can become invaluable for probing electrochemical reactivity in nanoscale volumes of batteries and fuel cells, whereas light-assisted electrical SPMs are likely to emerge as techniques of choice for probing photovoltaic materials and devices. Similarly, STM and associated spectroscopies will grow as techniques of choice to explore quantum behavior of materials on atomic level, as well as enabling tool for single-atom manipulation and assembly of atomic scale devices.

Current and future challenges

Despite the tremendous progress in SPM instrumentation and continuously growing number of imaging modalities and experimental platforms worldwide, the basic principles of SPM remained unchanged from the early days of Binnig, Quate, and Rohrer [150]. The SPM is based on continuous raster scan of the probe, effectively sampling the probe-surface interactions over the uniform rectangular grid of points. The spectroscopic measurements are enabled either as point and click approach by operator, or via hyperspectral imaging modes in which the spectroscopic data is acquired over the uniform grid. This acquisition mode is convenient from the instrumental implementation, human perception, and mathematical analysis perspectives. For hyperspectral measurements, a number of physics- or data-driven approaches have been developed to convert the high dimensional spectroscopic data sets to the set of 2D representations [154]. However, for realistic materials the information of interest is often concentrated in a small amount of locations. For example, in biological systems the molecules deposited on the surface are often of a higher interest then the substrate between them. In ferroelectric materials, functional responses of structural defects such as grain boundaries or topological defects such as domain walls are often of higher interest then the responses in the uniform-domain regions. This consideration is particularly important for the spectroscopic measurements. Here, the grid measurements are often very time consuming and can be associated with the tip damage. Perhaps even more importantly, many spectroscopic measurements can affect the state of material due to reversible or irreversible processes, for example shift ferroelectric domain walls, induce local electrochemical reactions, or plastic deformation of material. Hence, it is of interest the development of the instrumental workflows with the varying density of imaging points, and particularly methods for active experiment in SPM. In these, the algorithm updates the locations for image or spectroscopic measurements based on the measurements results within the same experiment. Here, we disambiguate three types of automated experiment, namely (a) adapting sampling for the scalar or multimodal measurements (figure 21(a)), (b) forward spectroscopic experiment in which the *a priori* known objects of interest are discovered in real time and spectroscopic measurements are taken (figure 21(b)), and (c) the inverse experiments in which the spatial

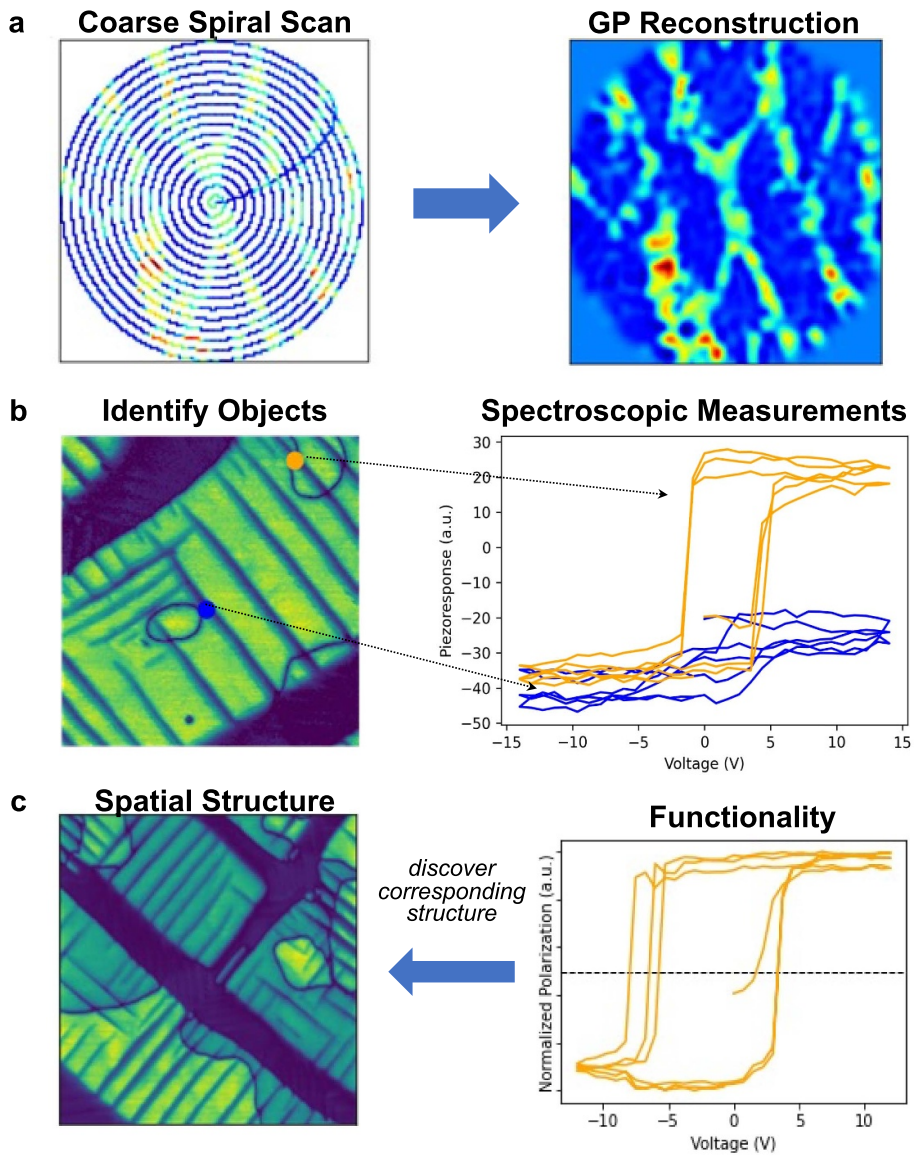

Figure 21. Three types of automated experiments in SPM. (a) Adapting sampling and reconstruction, where the full data set is reconstructed from sampling data [155]. John Wiley & Sons. © 2020 Wiley-VCH GmbH. (b) Forward experiment, in which the objects of interest are identified in image data, then spectroscopic measurements are taken at the identified objects. Reproduced from [156]. CC BY 4.0. (c) The inverse experiments, with known target functionality the corresponding structure or materials are discovered. Reproduced from [157], with permission from Springer Nature.

structures that correspond to functional behaviors of interest are discovered (figure 21(c)). We note that while (b) roughly corresponds to the operation of a human microscopist, the tasks (a) and (c) are purely amenable to human operation.

Advances in science and technology to meet challenges

Developing the automated experiment workflows requires the synergy of three components, namely engineering controls (i.e. the algorithm should be able to issue the control commands to the microscope), ML algorithms, and definition of the reward function. While the former two components are obvious, the third component (reward) is traditionally less recognized. However, it is clear that even for applications such as automated driving, the chosen pathway will be very different if reward is safety (favoring very slow driving) vs. time. For physical experiments, the reward is considerably more complex and has to be defined in the context of the specific physical experiment, prior hypotheses, etc. Here, for the reconstruction problems (a), the enabling algorithm can be variants of adaptive sparse sampling, e.g. based on Gaussian Processes. The reward function in this case is defined by the balance between the quality of reconstruction and minimization of samplings. However, while initially perceived to be promising, these algorithms often lead only to the insignificant (factor of 2–3) reduction of sampling points

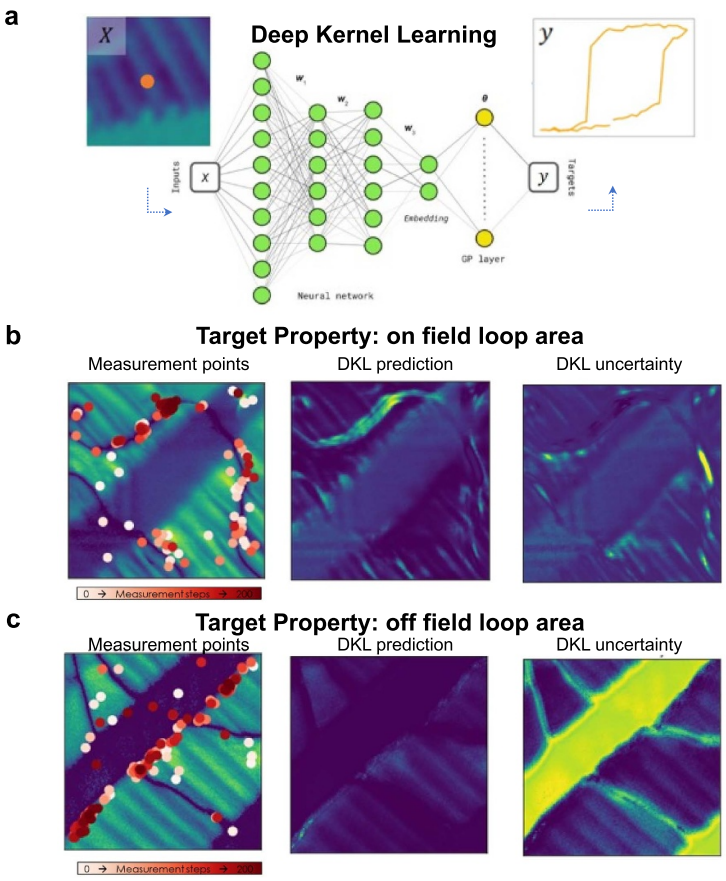

Figure 22. Deep kernel learning piezoresponse force microscopy (DKL-PFM). (a) Deep kernel learning structure: it contains a neural network to encode the properties of image data into a low-dimensional latent space and a Gaussian process layer operating over the latent space to analyze the relationship between image data and target property encoded in spectroscopic data. (b)–(c) DKL-PFM results on a $PbTiO_3$ thin film, which show that the DKL decided measurement points are around domain walls (indicating domain walls to be interesting) and predict that domain walls have large polarization dynamics. Reproduced from [157], with permission from Springer Nature.

at the cost of more complex scanning protocols. This behavior can be traced to the presence of the multiple length scales of the images that preclude effective construction of GP kernels. From a more general perspective, the classical uniform scanning of the SPM corresponds to the fully open prior, and hence is often optimal. The experiments (b) rely on the *a priori* defined objects of interest that can be recognized in real time using deep convolutional networks. Here, the emergence of the ensemble and iterative training methods allowed to address the inevitable out of distribution effects (i.e. capability of the trained network to recognize objects of interest if microscope parameters have changed). Recently, a deep residual learning framework with holistically-nested edge detection (ResHedNet) has been ensembled to minimize the out-of-distribution drift effects in real-time SPM measurement [156]. The ensembled ResHedNet is implemented in operating SPM, and converts the real-time PFM data stream to segmented ferroelastic domain wall images. Then, a pre-defined workflow uses the discovered domain walls as the coordinates for spectroscopic measurements. In doing so, the approach allows a thorough investigation of domain walls (virtually all locations at domain walls) in an automated manner, in contrast, traditional manual operation only allows us to investigate a limited amount of locations at domain walls. Using this approach, alternating high- and low- polarization dynamic ferroelastic domain walls in a $PbTiO_3$ thin film is observed. Finally, the emergence of the deep kernel learning (DKL) methods allows to implement the inverse spectroscopic workflow. Here, the operator defines the characteristics that make the spectrum of interesting, e.g. intensity of a specific feature, specific aspect of spectrum shape, or even maximal variability of spectra within the image. In other words, each collected spectrum can be associated with a single number defining how 'interesting' it is, in absolute sense for compared to previously acquired spectrum. The DKL algorithm learns what elements of the materials structure maximize this reward, and guides the exploration of materials surface accordingly. This DKL algorithm is recently implemented in SPM to investigate the relationship between ferroelectric domain structure and polarization dynamics [157]. Both the DKL exploration process and results are interesting. As show in figure 22,

the DKL exploration process demonstrates the domain walls to be interesting and the DKL results indicate the high polarization dynamic of 180° domain walls. Although these are expected by ferroelectric experts, DKL itself does not have any physical knowledge, all information is actively learned during the experiments.

Concluding remarks

ML methods have revolutionized many aspects of modern science including CV and image generation, medical and biological imaging, planning and prediction. The growth of the open code and data culture results in rapid propagation of ML algorithms between domains (see sections 32–35). Combined with the introduction of the Python control interfaces in modern microscopes, this poses the field of SPM for a transformative change. However, taking full advantage of this opportunity requires developing the connection between the domain areas and ML, including defining the domain-specific rewards that will steer the experiment, introducing the invariances and defining the biases that should be ignored, and adoption of the required ML tools required to build them. With these, automated experiments provide opportunities to perform experiments that are challenging in traditional methods because of experimental parameters are too complex and numerous (e.g. too many locations/conditions are required to investigate). In addition, automated experiments offer the opportunities for solving problems that are challenging to human beings, e.g. inverse problem. The active learning approach can learn the relationship between target functionality and material behavior quicker by several orders of magnitude than traditional method. As the development of ML and its connection to domain science, automated experiments are expected to accelerate the material and physics discovery.

Acknowledgments

This effort was primarily supported by the Center for 3D Ferroelectric Microelectronics (3DFeM), an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences under Award Number DE-SC0021118. The effort was partially supported (piezoresponse force microscopy) at Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS), a U.S. Department of Energy, Office of Science User Facility.

14. Image restoration for scanning microscopy

Arlo Sheridan[1] *and Uri Manor*[1,2,3]

[1] Salk Institute for Biological Studies, La Jolla, CA, United States of America
[2] Department of Cell & Developmental Biology, University of California, San Diego, La Jolla, CA, United States of America
[3] Halıcıoğlu Data Science Institute, University of California, San Diego, CA 92093, United States of America

Status

Point-scanning imaging systems are among the most commonly used microscopy imaging modalities due to their versatility and accessibility [8, 158]. While these tools have had a profound impact on advances in the life sciences, they come with a cost. Since it is difficult to simultaneously optimize imaging speed, resolution and sample preservation, one parameter is always compromised at the benefit of the others (aptly deemed the 'eternal triangle of compromise', figure 23) [158]. For example, imaging speed and sample damage can be enhanced by ensuring a fast pixel dwell time, but at the expense of a lower resolution and SNR. It is vitally important to correct for these shortcomings to facilitate future scientific breakthroughs. To address these challenges, DL-based methods have been developed to subsequently improve the negatively affected component(s) following image acquisition. Over the past several years, DL breakthroughs have advanced the field in both denoising and super-resolution of point-scanning microscopy modalities including fluorescence and EM. Similar advances addressing image degradation through DL have also been critical in computational phase microscopy (section 8). While initial approaches [8, 158, 159] centered around fully supervised learning via pairs of images of the same samples acquired at low versus high quality, promising new semi-supervised [160] and fully unsupervised methods [161–163] are beginning to demonstrate competitive results that allow investigators to avoid the technically difficult and costly process of acquiring paired data. Optimization of image acquisition is a costly process. DL-based restoration and super-resolution not only provide a solution to the 'eternal triangle of compromise', but substantially decrease the amount of imaging time (and subsequent costs) to conduct necessary experiments for downstream analysis. In order to further minimize these costs, future work should aim to improve fully unsupervised approaches and, ideally, integrate these into imaging hardware. Additionally, it is imperative to ensure that these methods are openly accessible to the broad community via intuitive, easy-to-use software that does not depend on a resident computer scientist. These advances will alleviate the burdens on researchers and allow them to focus on their scientific questions.

Current and future challenges

While DL approaches offer a practical solution for overcoming common pitfalls with scanning microscopy, there are several limitations that pose a challenge for current methods. Fully supervised methods still require the acquisition or generation of pairs of high- and low-quality images of the same samples to create training data [8, 158, 159]. Here we are using the term 'quality' to include both SNR and/or resolution. The high-quality data typically must exceed the desired quality of the output of the trained network. Acquiring image pairs of the same sample can be technically difficult, costly, and sometimes impossible; often, samples can only be imaged once [158]. Generating and validating semi-synthetic data (e.g. a 'crappifier' that generates low-quality data from high-quality data [158],) also requires significant effort, and validation still depends on real-world image pairs of the same sample. These data requirements are also hampered by the high cost of imaging and storing high quality data, which is proportional to image resolution [158]. Unsupervised approaches aim to solve this by enabling image restoration from noisy or low-resolution input images alone. While leading methods [161] generate impressive reconstructions, it is still challenging to match the accuracy of fully or semi-supervised results. For example, successfully denoising pixel-wise independent noise does not necessarily remove spatially correlated (structured) noise [161]. The extra effort to generate target data for fully supervised approaches is therefore sometimes unavoidable. Additionally, image restoration is an inherently ill-posed problem, i.e. multiple different high-quality images could be used to generate a single low-quality image [160, 163]. While this is less of a concern for popular tasks such as restoring photographs, restoring scientific imaging data can be challenging because of a greater need for accuracy. Any inconsistency in the fidelity of reconstructions can have harmful effects on downstream analyzes. For example, super-resolution can be used to quantify objects in microscopy data imaged with a lower resolution than is normally required to resolve these structures. For this approach to be useful, the model must be validated to show

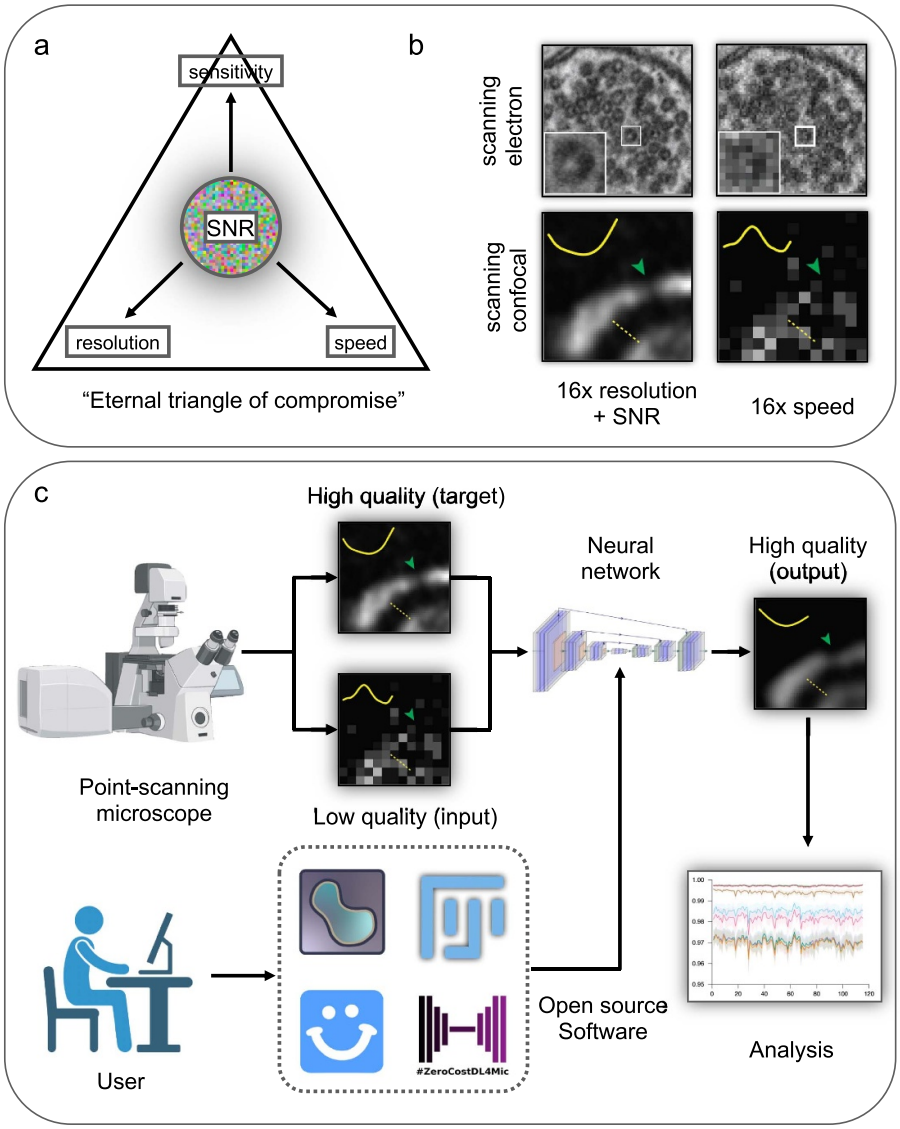

Figure 23. Image restoration for scanning microscopy. (a) The 'eternal triangle of compromise' illustrates the necessary compromises for optimizing each imaging component. (b) Examples illustrating the tradeoffs between imaging speed and resolution/SNR for scanning electron microscopy (SEM) and confocal microscopy data (top and bottom rows, respectively). (c) Overview of an example workflow for integrating deep learning with open-source software to accelerate scientific discovery. Microscope and user schematics created with BioRender.com.

it did not introduce errors resulting in inaccurate quantification of these objects [158]. Furthermore, for image restoration methods to realize their full potential they must be accessible to the community via easy-to-use graphical user interfaces (GUIs) [164, 165]. Most scientists do not have the expertise necessary to re-implement DL approaches, and running established code often poses a steep learning curve. However, it is not always trivial to implement robust solutions that satisfy all requirements. Ideally, software should be easily packaged, distributed, and hosted on a well-maintained platform. Critically, having turn-key methods for generating model outputs will greatly facilitate validation of model outputs by domain experts. Validation should never rely solely on pixel-based metrics such as PSNR or SSIM. Instead, validation workflows should integrate visual inspection by human experts, multiple pixel-based metrics, and, most importantly, comparing the final measurement of the experiment with ground truth, validated data. For example, the number and size of objects of interest (e.g. presynaptic vesicles) generated in the model output should match the ground truth data as closely as possible [158].

Advances in science and technology to meet challenges

Even though fully supervised methods require the acquisition of high-quality data, the standardization of imaging techniques and advances in data storage have been transformative over recent years. Many

microscopy imaging systems now allow for automated collection and alignment of images. It is also considerably cheaper nowadays to store large amounts of data (for example, storing a terabyte of data on AWS on average costs a relatively modest $20 per month[70]). As imaging datasets become more widely shared, it may be feasible to generate learned crappifiers that can be used to quickly train new supervised models. Nevertheless, in an ideal world, unsupervised approaches would be preferred. Despite the aforementioned challenges, denoising methods can now handle structured noise in addition to pixel-wise noise and recent results indicate that accuracy on par with fully supervised methods is achievable. Self-supervised and unsupervised methods for super-resolution are also rapidly advancing [166]. Furthermore, some methods now provide a distribution of restorations instead of a single output [163]. This is of great practical relevance for solving inverse problems in which multiple solutions may be proposed. Theoretically, it should allow a user to choose different solutions across regions based on additional context. For example, if one corner of an image is much noisier than the center, a single solution might provide reasonable results in the center and poor results in the corner. In these cases, having a range of solutions that a user can choose from is preferable. Recent work yielding significantly improved results using more sophisticated 'crappifiers' for generating semi-synthetic training data and more sophisticated loss functions highlights the huge potential for further improvements in the future [167].

The full impact of these advances can only be achieved if they are accessible to a wide range of users with various backgrounds. Efforts toward democratizing DL for microscopy workflows are critical to this process, as discussed in sections 32–35. In order to cater to the broad community, platforms with intuitive user interfaces and continued support must exist. Over the years, FIJI/ImageJ[71] (the *gold standard*) has allowed scientists to implement a wide range of image analysis techniques on their data. Next generation tools such as Napari[72], ImJoy[73], and ZerocostDL4mic [164] aim to extend this accessibility and enhance it through the integration of DL approaches. These tools and many others [165, 167] will allow users to analyze their data in ways that were previously unattainable without sufficient expertise.

Concluding remarks

Advances in biological imaging are moving at an exponentially increasing pace. As it becomes easier to acquire large amounts of images via point-scanning microscopy systems, there is an increasing need to accurately and efficiently analyze the resulting data. Since image acquisition poses challenges for the optimization of the 'eternal triangle of compromise', DL methods are exciting solutions for image restoration. Since these are difficult inverse problems to solve, ideal solutions need to be robust to data ambiguities and should limit the burden on researchers to collect and store high quality target data. Additionally, efforts should ensure the accessibility of methods to a broad user base across scientific domains. With the development of recent DL approaches and the growing communities dedicated to developing open-source software and datasets, it is clear that the field is moving in the right direction.

Acknowledgments

Arlo Sheridan and Uri Manor are supported by the Waitt Foundation, Core Grant application NCI CCSG (CA014195), NIH (R21 DC018237), NSF NeuroNex Award (2014862) and the Chan-Zuckerberg Initiative Imaging Scientist Award.

[70] https://aws.amazn.com/s3/pricing/.
[71] https://imagej.net/software/fiji/.
[72] https://napari.org.
[73] https://imjoy.io.

15. Single molecule localization

Elias Nehme[1,2], *Ofri Goldenberg*[2] *and Yoav Shechtman*[2]

[1] Department of Electrical and Computer Engineering, Technion—IIT, Haifa, Israel
[2] Department of Biomedical Engineering and Lorry I. Lokey Interdisciplinary Center for Life Sciences & Engineering, Technion—IIT, Haifa, Israel

Status

The spatial resolution of conventional microscopes is fundamentally bounded by the diffraction limit at approximately half the wavelength of the light, which, in the visible range, practically corresponds to 200–300 nm. To overcome this limitation, a variety of super-resolution microscopy techniques have been developed including structured illumination microscopy, stimulated emission depletion, as well as single molecule localization microscopy (SMLM) [168], which is the focus of this perspective. The main working principle of SMLM relies on a space–time trade-off: instead of capturing a single image of a fluorescent sample, a movie consisting of many frames (typically thousands) of temporally blinking fluorophores is acquired. In each frame, only a sparse, random set of fluorophores is activated, and their positions are determined computationally. There are multiple chemical and physical mechanisms to obtain blinking, and a myriad of associated acronyms, most notably PALM, STORM, and PAINT [168]; however, subsequent analysis is similar between the different variations. After data acquisition, the resulting localizations are combined numerically to render a single, computationally super-resolved image, typically, with an order of magnitude improvement in resolution (figure 24).

SMLM revolutionized biological research, enabling *nanoscale* imaging of biological structures and tracking of single-particles [169], thereby earning its pioneers the Nobel Prize in Chemistry in 2014. Compared to other high-resolution imaging modalities like EM, SMLM offers the high specificity and SNR of fluorescence microscopy, as well as the possibility to image living cells.

With the great advancement in resolution introduced by SMLM, came unique experimental and algorithmic challenges. The main issue with trading-off space over time, manifested in the per-frame emitter *sparsity* constraint, is the sacrifice of temporal resolution, which imposes strict limitations on the ability to image dynamic processes. This problem becomes more acute as we are interested in sensing more and more physical properties of the imaged sample, such as depth (3D), color information, and molecular orientation.

In recent years, DL has found tremendous success in handling some of these challenges [170]. In particular, DL-based E2E optimization for joint design of sensors and algorithms [171–173] has led to powerful task-driven experimental designs, significantly pushing the barriers of the spatiotemporal trade-off in localization microscopy (figure 25). This convergence of sensor design and learning-based reconstruction is conceptually related to trends in computational phase microscopy highlighted in section 8. Such designs are starting to find their applications in scientific research [174, 175]; however, work remains to be done in making these methods fully-mature and widely adoptable by the biological research community. Specifically, efforts should be invested in robustifying SMLM algorithms for a wide range of experimental conditions and providing accessible software packages to end users [164].

Current and future challenges

Here, we outline four main challenges that are key to address in order to improve the capabilities and accessibility of SMLM:

1. **Improving spatiotemporal resolution**. For imaging fast dynamical processes at high resolution using SMLM, its temporal resolution needs to be improved. The trick to enhance its spatio-temporal resolution is to trade off experimental time with information-rich image data containing complex patterns. The latter needs to be then analyzed by powerful algorithms to extract the underlying physical information. For example, relaxing the emitter sparsity constraint to increase temporal resolution in SMLM requires image processing algorithms that can handle nearby emitters with overlapping point spread functions (PSFs) [9, 177].
2. **Enhancing information extraction**. There is often a need to look beyond the 2D positions of emitters. For example, by tagging different targets (e.g. proteins) with distinct fluorescent emitters, we can capture correlative information that reports on inter-specie relations. One exemplary challenge, in this case, is to classify the color of each emitter based on the acquired PSFs, which are typically captured on highly photon sensitive, yet spectrally insensitive (grayscale) detectors. Another physical

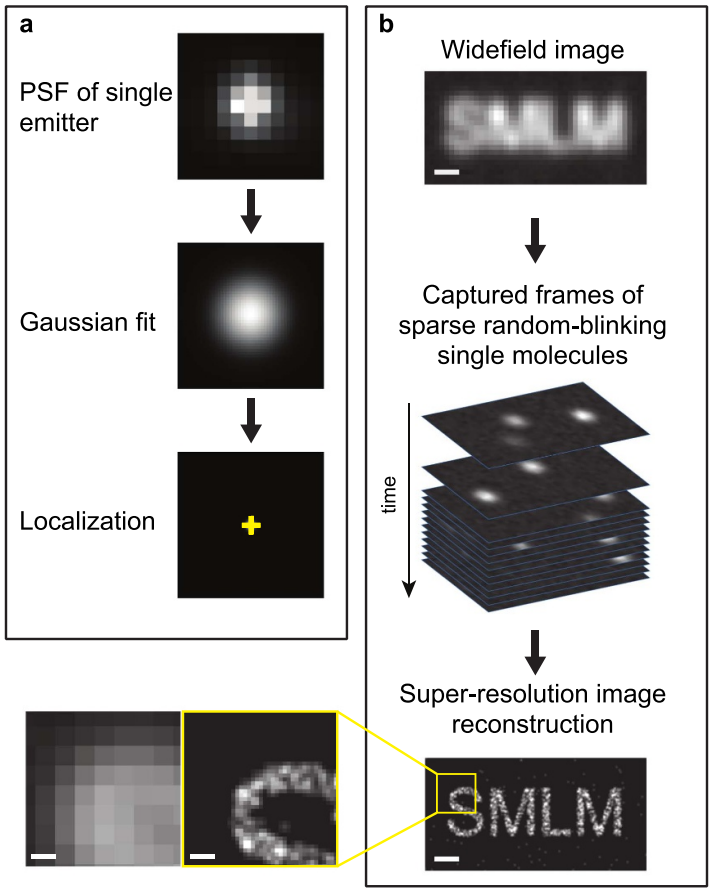

Figure 24. Single molecule localization scheme. (a) Single emitter position computed by fitting to a PSF model. (b) SMLM experimental pipeline. Scale bar = 0.4 μm (inset scale bar = 0.1 μm).

property of interest is emitter depth; since biological structures are three-dimensional, we can gain significant insight by looking into their 3D organization (3D SMLM).

3. **Optimal information acquisition**. Imaging systems in SMLM are typically optimized through physical intuition-based heuristics or through mathematical measures quantifying information content, e.g. Fisher Information. However, such methods are still limited in their ability to adapt to challenging experimental conditions and are difficult to generalize to systems acquiring multiple physical properties jointly. Hence, optimal designs of acquisition schemes are key to unlocking the full potential of SMLM. Similarly, in FLI (section 10), optimizing acquisition in conjunction with inference models has proven crucial for extracting weak lifetime signals under photon-starved conditions.
4. **Increasing algorithm reliability**. DL algorithms are extremely powerful and have led to a performance revolution in solving inverse problems. However, little is understood about their inner-workings and their failure modes. Specifically, for SMLM, there is need for sample-adaptive self-tuning algorithms, that can handle arbitrary experimental conditions without need for extensive calibration. More importantly, in a field such as biological research where algorithms are expected to drive scientific discovery, we need to be able to quantify uncertainty and bias in the reconstructions [177]. At the moment, DL algorithms have a hard time 'knowing when they do not know', and we anticipate extensive work to be done on this front to improve their reliability.

Advances in science and technology to meet challenges

DL has proven to be highly successful in handling some of the fundamental challenges in SRM [170]. DL is particularly suited for SMLM, because large, paired training sets, which are the main bottleneck in supervised learning, can be generated easily using accurate simulators based on physical models. The first application of DL to 2D SMLM was presented in [9]. The authors showed that the acquisition time could be shortened significantly down to a few hundred frames by proposing a DL algorithm that is able to reconstruct dense emitters with overlapping PSFs. A similar strategy was used later in Barth

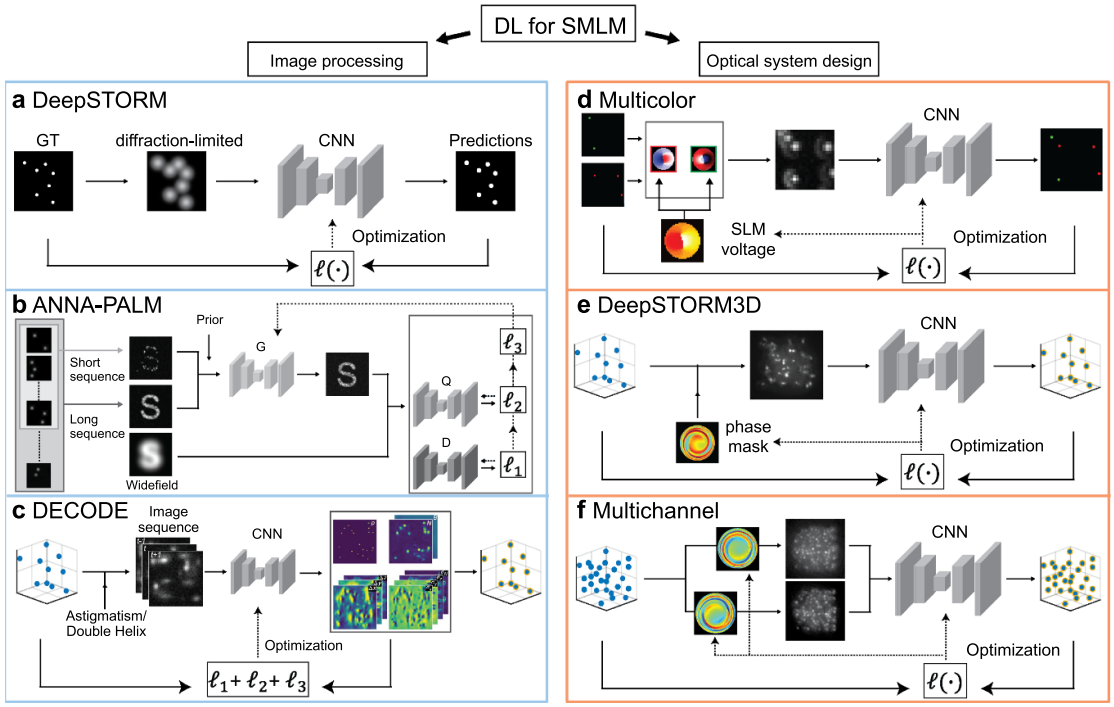

Figure 25. Deep learning applications in SMLM. (a) DeepSTORM [9], a neural network for fast and precise reconstruction of a super-resolved image from raw STORM frames. (b) ANNA-PALM [176], a neural network for interpolating a super resolution PALM reconstruction from a small number of frames. (c) DECODE [177], a neural network for dense molecule localizations in 3D that outputs maps of detection probability, subpixel coordinates and their uncertainty, brightness, and background. Reproduced from [177], with permission from Springer Nature. (d) Multicolor localization and classification [171], E2E learning of a neural network and chromatically-dependent engineered PSF for localization and classification of multicolor molecules from grayscale measurements. (e) DeepSTORM3D [172], E2E learning of a neural network and engineered PSF for dense 3D localization. (f) Multichannel [173], E2E learning of a bifurcated optical system with two engineered PSFs and a neural network for ultra-dense 3D localization. © (2021) IEEE. Reprinted, with permission, from [173]. Adapted from Speiser *et al* [177].

et al [174] for tracking chromatin dynamics in living cells. Concurrent work proposed content-aware strategies to increase the temporal resolution [176], and even decrease the number of necessary frames to a single image [8]. Similarly, DL was also applied for sensing additional physical properties from grayscale 2D measurements such as color classification [171, 178], depth estimation from high-density data with engineered PSFs [172, 177, 179], as well as 3D molecular orientation [180–182]. Additionally, DL networks have been used for background estimation to improve localization accuracy, as pre-processing for single molecule fitting algorithms [183] or directly in the localization pipeline [177].

A particularly promising application of DL in SMLM is the ability to tailor the acquisition scheme to the task at hand. Specifically, E2E optimization of the physical acquisition system jointly with the data processing algorithm, holds great promise [171–173, 184]. E2E optimization taps into the full potential of existing hardware as well as making use of new hardware in a task-driven manner. This synergy between acquisition and reconstruction enables us to design significantly more complex sensing paradigms that make full use of the photons, as well as powerful DL-based reconstruction algorithms. DL was originally invented to approximate functions which were hard to mathematically define, and optical instrumentation design is no different. Recently, E2E optimization was shown to be extendible to multiple sensor designs [173, 185], offering even greater temporal resolution at the expense of a more complex optical setup.

Additional progress within the last few years addressed temporal resolution limitations (super spatiotemporal resolution reconstruction [186]), field-dependent aberrations (large FOV in 3D SMLM [187, 188]), processing speed [189], performance and versatility improvement [190], and learning and correcting optical aberrations in near real-time [191].

Concluding remarks

SMLM is a powerful tool in bio-imaging, with new applications emerging quickly. The combination of SMLM with DL enables efficient analysis and system-design for obtaining information-rich imaging data. Due to the nature of E2E designs, this interdisciplinary field requires tight collaboration between experimentalists and data scientists to fully exploit existing sensors and compute power. Focus thus far has

been mainly in improving reconstruction algorithms [192]; however, we anticipate that more and more applications will emerge in which acquisition systems and algorithms are designed jointly. Furthermore, the development of new sensors and hardware will drive DL-based design of new optimal acquisition systems. In any case, there will be need for standardized metrics to assess performance, reproducibility, and reliability of DL-based SMLM algorithms. Finally, caution needs to be taken and uncertainty quantification should be addressed before existing tools can be used for scientific discoveries; this is especially important for methods that assume strong prior information on the object to be recovered.

Acknowledgments

This work was supported by H2020 European Research Council Horizon 2020 (802567) and by the Israel Science Foundation (Grant No. 450/18).

16. Nanofluidic scattering microscopy (NSM)

Henrik K Moberg[1], *Christoph Langhammer*[1] *and Barbora Špačková*[2]

[1] Department of Physics, Chalmers University of Technology, Gothenburg, Sweden
[2] FZU, Institute of Physics of the Czech Academy of Sciences, Prague, Czechia

Status

Recent years have witnessed the rapid development of new fields of science, such as nanobiology, single-molecule biochemistry, and biophysics. A myriad of information has been obtained which had not been anticipated based on conventional ensemble-averaged measurement and microscopy tools. At the same time, single-molecule imaging methods have long relied on fluorescent labeling [193]. This labeling involves the chemical attachment of light-emitting moieties to the biomolecules of interest, which in turn may alter their natural state and, thus, their biological activity and function. This drawback has been the driving force for the development of label-free optical microscopy methods. Among these techniques, iSCAT has been successfully employed in many applications of imaging and analysis of individual biomolecules [194]. These studies are however limited to biomolecules attached to a surface and do not reveal their natural diffusive motion in solution.

To this end, we have recently introduced NSM [195], whose unique underlying principle enabled to bypass those limitations: NSM is capable of imaging individual biomolecules in free motion without any label. This tool relies on the interference of light scattered from a nanofluidic channel and a nano-object inside it—such as a biomolecule (figure 26). To enable quick and accurate analysis of the recorded images, we employ a DL-based CV pipeline, consisting of three CNN architectures (figure 27). In Špačková *et al* [195], this pipeline is employed to identify and characterize a diverse array of biological nanoparticles ranging from extracellular vesicles down to DNA molecules and proteins of molecular weight down to the tens of kDa regime. Specifically, given an image containing the scattered light of several biomolecules inside a nanochannel over time (known as a 'kymograph'), the CV pipeline outputs the trajectory of each separate biomolecule along with its two key characteristics—its molecular weight (MW) and its hydrodynamic radius (R_S). Their determination is enabled by the fact that the MW is proportional to integrated optical contrast (iOC) of a molecule and that R_S correlates with the diffusivity (D), i.e. the characteristic of its movement. It has been demonstrated that conventional analysis algorithms can, in principle, handle these aspects of NSM data processing [195]; however, DL offers distinct advantages. Compared to traditional methods, DL enables significantly faster processing, making it well-suited for the efficient analysis of large-scale image datasets. Moreover, it is fully automated, eliminating the need for manual parameter tuning. This enhances robustness by allowing the model to generalize across diverse biological nanoparticle species and experimental conditions.

With further advances in this novel technology, concurrently on the experimental and DL side, we expect NSM to have a significant impact in the field of single-molecule science.

Current and future challenges

NSM has proven its unprecedented performance in visualizing and analyzing of individual biomolecules diffusing in a solution inside a nanofluidic channel. However, numerous challenges have still to be addressed to allow the widespread adoption of the technique in research and industry.

The most significant drawback is its low throughput, which makes the analysis of heterogeneous samples very time-consuming. This is the consequence of the current limitation of the experiment to measure one nanochannel at a time. Parallel analysis of hundreds of nanochannels filling the whole field of view of a microscope is possible, in principle, but it is hampered by the lower performance of cameras when operated in the full frame regime. Similar bottlenecks between acquisition hardware and computational scalability are encountered in FLI (section 10), especially when high frame rates and low photon counts are involved. Such an approach would also generate a vast amount of data, making conventional processing inefficient. DL provides a scalable solution for rapid, high-throughput data analysis, enabling real-time extraction of molecular properties across multiple channels and overcoming the computational bottlenecks associated with conventional analysis methods.

There is also a significant limitation related to the current CV pipeline, as the performance degrades with high-concentration samples. This limits NSM's applicability to more complex biological systems, where sample heterogeneity and molecular crowding are critical factors.

Finally, the reported performance of NSM allows to analyze biomolecules down to tens of kDa in molecular weight. There is however a whole realm of biomatter below this limit that is central in many

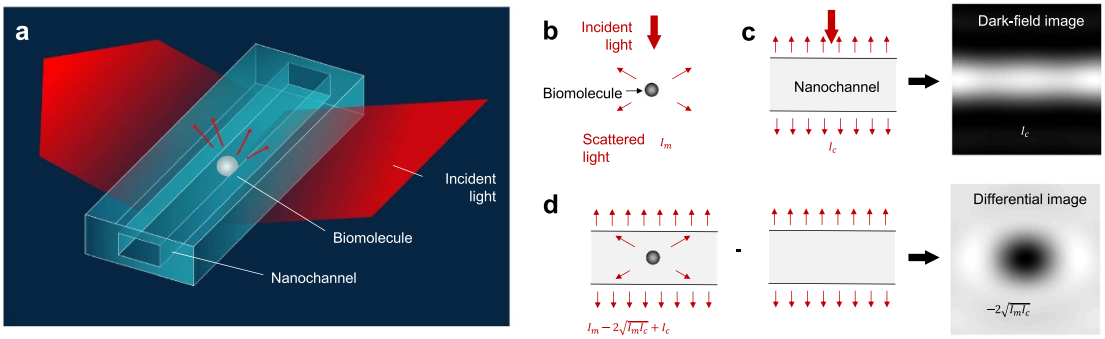

Figure 26. The NSM principle. (a) Artist rendition of the experimental configuration where visible light is irradiated onto a nanochannel with a biomolecule inside, and where the light scattered from the system is collected in a dark-field configuration. (b) Schematic of the light scattered by a single biomolecule. (c) Schematic of the light scattered by a nanochannel and the corresponding dark-field image. (d) Schematic of light scattered by a nanochannel with a single biomolecule inside, and the corresponding differential dark-field image obtained by subtracting an image of the empty nanochannel from the image of the nanochannel with the biomolecule inside. Reproduced from [195]. CC BY 4.0.

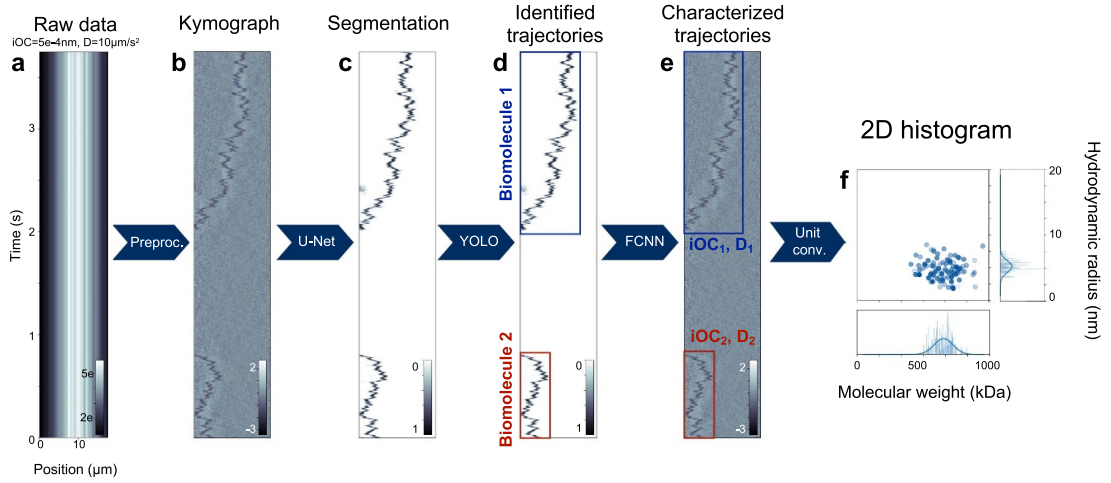

Figure 27. Computer vision (CV) analysis workflow. (a) Recorded data—time evolution of the scattered light of a nanochannel with a single biomolecule inside. (b) Kymograph—preprocessed data obtained by removal of the background estimated by local mean. (c) Segmented image where the particle positions are detected using a U-net neural network architecture. (d) Single biomolecule trajectory identification using the YOLOv3 algorithm. (e) Property calculation (*iOC* and *D*) using a custom fully connected neural network (FCNN). (f) Unit conversion from *iOC* to molecular weight (*MW*) and from *D* to hydrodynamic radius (R_S) plotted in a 2D histogram for illustrational purposes. Reproduced from [195]. CC BY 4.0.

applications of biomolecular research. Pushing the detection limit will require both improved instrumentation for higher sensitivity and advanced data analysis to extract weaker signals more effectively.

Besides the weighting and sizing of individual biomolecules, NSM has potential in several other applications. First, it can be applied to label-free and tether-free investigation of biomolecular interactions which is central for progress in many important areas (e.g. medical and pharmaceutical research). Second, integrating NSM with spectroscopic techniques could provide a powerful multimodal platform for characterizing molecular composition alongside structural and dynamic properties. The successful implementation of these future applications will require overcoming challenges in both experimental design and data analysis methods, particularly in developing DL approaches for accurate and efficient interpretation of complex datasets.

Advances in science and technology to meet challenges

Further development of NSM is a multidisciplinary effort, with advancements in DL techniques playing a central role alongside innovations in detector technology, micro- and nanofluidic technology, and nanofabrication.

Future DL innovations are well poised to address the challenges in analysis of high-concentration samples by advancing segmentation and tracking algorithms. Self-supervised learning techniques [196]

and transformer-based architectures [197], which excel at capturing long-range dependencies in kymographs, could significantly enhance molecular tracking in dense environments. These directions strongly resonate with trends seen in modern unsupervised image restoration and enhancement strategies, such as those applied to point-scanning microscopy in section 13. Geometric DL methods [198, 199] could offer improved particle tracking and characterization, mitigating issues caused by overlapping molecular signals, as described in sections 17 and 20.

Since the NSM AI models are trained entirely on simulated data, their performance is limited by available computational power. Thus, the detection limits of the NSM analysis will improve in tandem with the breakneck speeds at which newer generations of GPU and TPU operates. Most recent theoretical analysis predicts that the access to already-available higher computational power combined with further model optimization leads to an order of magnitude improvement in terms of detection limits. Thus, it can be expected that the analysis of biomolecules in the single-digit kDa regime will become available soon.

Expanding NSM to study biomolecular interactions and integrating it with spectroscopic techniques will require further advancements in DL. For biomolecular interaction studies, DL models must be capable of accurately distinguishing weak, transient binding events from random diffusion, a challenge for conventional tracking methods. Likewise, applying NSM to spectroscopic techniques will necessitate DL algorithms optimized for low-photon-count conditions, where traditional image processing struggles with increased noise and reduced contrast. Furthermore, as NSM will integrate multi-modal data sources, DL approaches must evolve to effectively fuse diverse datasets, extracting meaningful correlations while maintaining high accuracy across different signal types and scales.

Regarding the advances in detector technology, the CMOS cameras available on the market have quickly increasing frame rate, improved sensitivity, and larger dynamic range. The NSM method will thus reach higher throughput via parallel analysis of multiple channels and achieve more accurate measurements of even smaller molecules.

Nano- and microfluidic technology is a swiftly evolving field which keeps bringing new possibilities for further development and application of the NSM method. New tools (such as nanofluidic valves [200] and traps [201]) are introduced to provide means for efficient confinement and release of ultra-low volumes. This ability is crucial for development of future applications, such as investigation of biomolecular interactions inside a nanofluidic channel.

New materials available for nanofabrication will potentially open up new possibilities for increasing the sensitivity of the method [202]. In addition, advances in large-scale nanofabrication will pave the way for NSM becoming a reliable and cost-effective bioanalytical method.

Concluding remarks

In its current form, NSM already enables highly accurate label-free and tether-free characterization of individual biomolecules and biological nanoparticles in a wide range of biofluids. The expected advances of the instrumentation and DL tools will push the performance even further. In particular, high-throughput, resolution in single-digit kDa regime and at high concentration will find numerous bioanalytical applications requiring analysis of highly heterogeneous samples. Long-term monitoring of individual biomolecules diffusing in solution represents a yet unexplored opportunity for studies of conformational changes, aggregation processes and interactions between individual biomolecules. In addition, due to minimized dilution of the sample by the nanofluidic platform, NSM is highly efficient for transporting ultra-low volumes, such as intracellular content or secreted metabolites of a single cell, thereby paving the way to real-time label-free single-cell studies. Moreover, it can be expected that NSM will find applications outside the field of bioanalysis, such as characterization of inorganic nanoparticles, particle counting, or single particle analysis.

Acknowledgments

The authors acknowledge financial support from the Swedish Foundation for Strategic Research project FFL15-0087 (C.L.) and the Czech Science Foundation project 22–18203S (B.S.). Part of this research has been executed at the Chalmers Nanofabrication Laboratory MC2 and under the umbrella of the Chalmers Excellence Initiative Nano.

17. Particle tracking

Saga Helgadottir

SciLifeLab BioImage Informatics Unit, Dept. of Information Technology, Uppsala University
Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

Status

Tracking the motion of single particles has become a critical tool for probing the microscopic world. Particle tracking has come a long way from manually locating particles' positions over a century ago to using standard algorithmic approaches to the use of DL in microscopy. A pioneering example of particle tracking is when Jean Perrin proved the physical existence of atoms in 1910. Perrin projected the image of microscopic colloidal particles in a solution on a sheet of paper and manually tracked their positions, managing to quantify their Brownian motion despite a time resolution of just 30 s [203].

Modern particle tracking has largely been dominated by a technique generally referred to as 'digital video microscopy', introduced over 20 years ago [204], in which a video of microscopic particles is acquired and the particles' positions in each frame are determined using computer algorithms. Until a few years ago, the standard algorithms have often been based on the measures of the centroid of the particles in a black-and-white thresholded version of the image or the radial symmetry center of the particles [205], and can successfully achieve subpixel resolution when their underlying assumptions for ideal experimental conditions are satisfied—generally that the particle is spherically symmetric and imaged with a homogenous and constant illumination. Their performance decreases drastically with less-than-ideal experimental conditions and significant user intervention is required, which in turn is time-consuming and can introduce user bias.

In the last few years DL, a kind of ML built on ANNs, has started to be employed for digital video microscopy. In contrast to standard algorithms, DL algorithms autonomously learn to determine rules to perform specific tasks using a series of input data and corresponding desired outputs—which, in the case of particle tracking, would be images of particles and their coordinates in the image. Early success of DL for particle tracking has already shown that DL outperforms standard algorithms in accurately localizing particles in challenging experimental conditions and in many cases is able to eliminate user bias by simulating training data [72] (figures 28(a)–(d)). This parallels early progress seen in single-molecule localization microscopy (section 14), where DL-based methods significantly improved localization under challenging conditions. However, there is still a lot of room for improvement and DL for particle tracking has yet to reach its peak potential.

Current and future challenges

One major challenge is the availability of training data, both in quantity and quality [94]. Even though it is possible to train DL algorithms using experimentally-acquired data, this becomes increasingly difficult for the application of particle tracking. Considerable amount of data needs to be acquired for each experimental setting, for which it is also extremely difficult to determine the ground-truth particle positions with sufficient accuracy as they often need to be manually annotated. This process is time-consuming and limits the algorithms to human-level accuracy as well as introducing bias. This has been partly solved by training the algorithms using simulated particle images for which the ground-truth positions can be known exactly. It is especially useful to be able to physically simulate particle images replicating each user's experimental setup, from properties of the particle to optical properties of the instrument used to capture the data. However, simulating training data is often not feasible and this is in fact true for non-symmetric and biological objects acquired in most imaging modalities. Similarly, multimodal microscopy approaches (section 11) face challenges in generalizing learning-based models across varying biological structures and experimental settings. When it is not possible to accurately recreate the experimental setups numerically, the original experimental dataset can be artificially expanded using augmentation [206]. A special type of DL algorithm has even been developed that is able to generate additional synthetic training images as a form of data augmentation. Augmentation can aid in the training process but cannot replace high-quality training data.

A second major challenge is, as for standard algorithms, that the trained algorithms are usually tuned for a specific problem and are not easily generalizable, meaning that they most often cannot be used to analyze data containing different particle types or obtained with another experimental setup. This, in addition to the steep learning curve for developing custom DL solutions, makes DL underutilized in

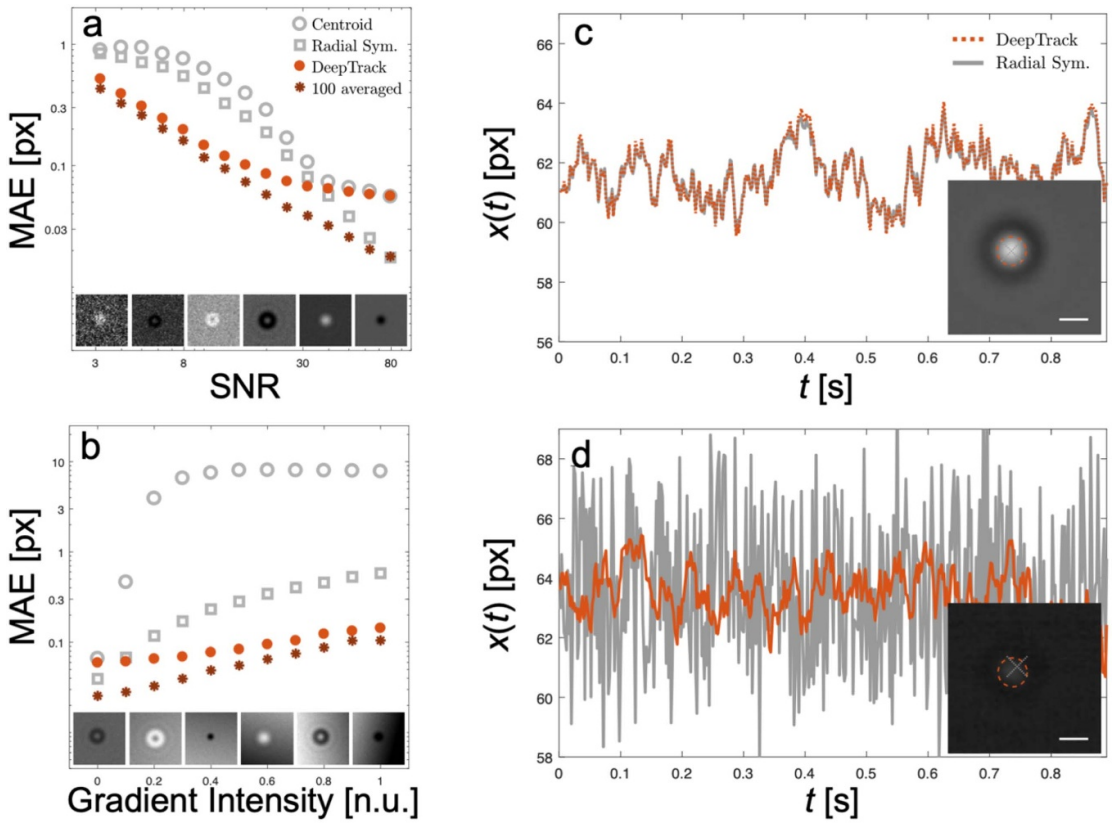

Figure 28. Particle tracking with neural networks. (a)–(d) Deep learning algorithms (orange) trained on simulated particle images outperform standard algorithms (gray) in accurately detecting particle positions in high-noise conditions for simulated image of symmetrical particles affected by a range of signal-to-noise ratios (a) and illumination gradients (b) (here, measured from the mean absolute error (MAE) in pixels), and experimental tracking of optically trapped particle (c), (d). Deep learning and standard algorithms lead to the same results under optimal illumination conditions (c) but the tracking from the standard algorithm becomes unstable in challenging conditions (d). Adapted from (a)–(d) deep learning algorithms (orange) trained on simulated particle images outperform standard algorithms (gray) in accurately detecting particle positions in high-noise conditions for simulated image of symmetrical particles affected by a range of signal-to-noise ratios (a) and illumination gradients (b), and experimental tracking of optically trapped particle (c)–(d). Deep learning and standard algorithms lead to the same results under optimal illumination conditions (c) but the tracking from the standard algorithm becomes unstable in challenging conditions (d). Reproduced from [72]. CC BY 4.0.

particle tracking and digital video microscopy as a whole. Pre-trained algorithms have been made available and are in practice easily applied to new data [207]. The results are however often acceptable only in the absence of a better alternative customized for the problem at hand.

Advances in science and technology to meet challenges

The most obvious way forward for DL for particle tracking is to focus on easier access to high-quality training data in order to custom train algorithms for each set of experiments or even to be able to train a generalized particle tracking algorithm. In this case, the continuous development of accurate ways to synthetically recreate more complex experimental setups imaging non-symmetrical particles with various imaging modalities is of upmost importance. The collaborative culture of the field provides a prime environment for a fast growth with open-source software packages emerging with the possibility of other researchers to contribute with additional functionalities related to their areas of expertise [206]. These software packages also attempt on being user friendly to allow for the development of custom DL solutions without the need for special coding skills, making them available for a broader audience.

A promising alternative is the new family of single-shot, self-supervised networks (section 17) that learn directly from one, completely unlabeled particle image. Such self-supervised approaches are conceptually similar to recent unsupervised restoration methods for scanning microscopy data discussed in section 13, where training without explicit labels has become a major breakthrough. In these methods—exemplified by the LodeSTAR architecture [208]—the network is trained to output two numbers that must transform exactly like the particle center under any in-plane rotation or translation that is applied

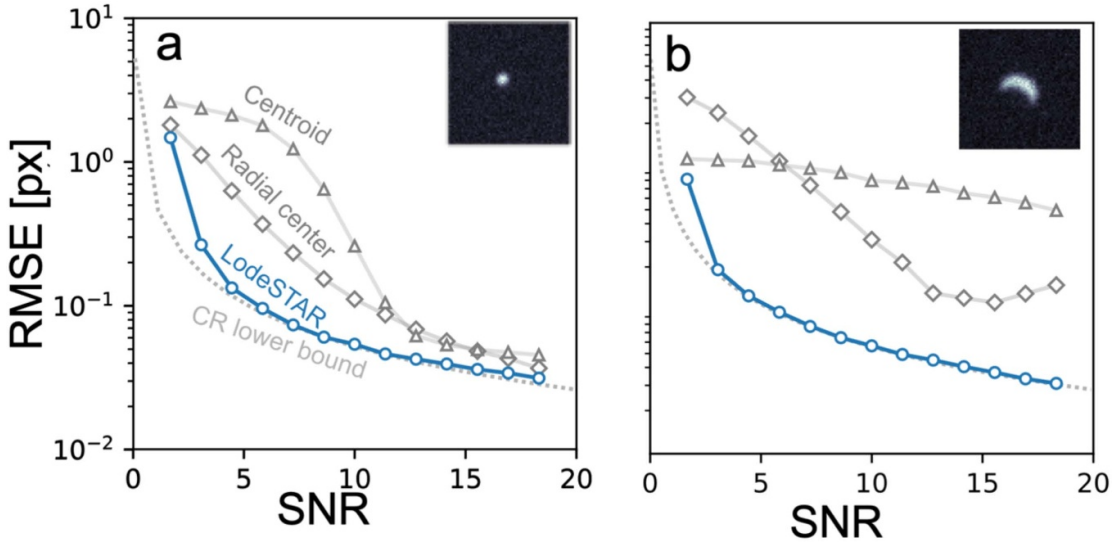

Figure 29. Self-supervised P´ particle tracking with neural networks. Novel deep learning algorithms can be trained using a single unlabeled experimental image, outperforming standard algorithms for symmetrical (e) and non-symmetrical (f) particles (here measured in root mean square error (RMSE) in pixels) affected by a range of signal-to-noise ratios (SNRs). Reproduced from [208]. CC BY 4.0.

to the input image. Because the only solution that satisfies this equivariance constraint is the true center itself, the network converges to sub-pixel accuracy without ever seeing ground-truth coordinates or simulated data (figure 29). This strategy therefore eliminates the time-consuming step of creating large, labeled training sets and is agnostic to particle shape, making it applicable to a wide range of morphologies and imaging modalities. The training runs in a few minutes on a standard laptop and the resulting model can be re-trained or fine-tuned easily, putting custom, high-precision tracking within reach of non-experts.

DL approaches have so far followed the conventional way of particle tracking, providing only a data-driven version of standard algorithms focusing on finding the coordinates of single particles in an image. For further analysis, the coordinates are then linked into single particle trajectories using other standard approaches. New approaches are able to take more advantages of the possibilities of DL to analyze the particles' dynamics to directly produce linked trajectories of particle coordinates as well as inferring local and global dynamic properties of the sample [46]. This has the potential to improve the analysis of samples where particles are lost, new particles introduced or change shape during the experiment, which is highly relevant in biological samples. When trained with manually annotated experimental data, the algorithm can even learn how to override imperfect annotation of the training dataset.

Concluding remarks

Despite its success in recent years, DL for particle tracking still has huge potential. Improved physical simulations of particle images will allow for more accurate tracking of single particles, bypassing the time-consuming and bias-inducing manual ground-truth annotations. Thanks to increasingly available inference speed, trained algorithms could also be implemented in the experimental setup for real-time particle tracking and decision making. Going one step further, new, fast and easily trained algorithms that only require a single unlabeled training image even give rise to the possibility to train and track particles on the spot during experiments.

Tracking particles is usually only the first necessary step in order to further analyze the dynamics of a system. DL can be used to go beyond only acquiring particle coordinates and simultaneously calculate other particle characteristics, as well as directly returning single particle trajectories and the underlying dynamics of the whole system.

Acknowledgments

The work for this project has received funding from the H2020 European Research Council (ERC) Starting Grant ComplexSwimmers (677511) and the SciLifeLab BioImage Informatics Unit.

18. Single-shot self-supervised object detection

Benjamin Midtvedt and Giovanni Volpe

Department of Physics, University of Gothenburg, Gothenburg, Sweden

Status

DL is a rapidly growing field in microscopy that aims to improve image analysis and automate the process of identifying and extracting information from images, enabling more efficient and accurate data analysis (see for example section 16 for particle tracking). Object detection is one such task where DL excels over traditional methods. However, most DL methods for object detection in microscopy require either large manually annotated datasets or highly realistic physical simulations of the experiment [94].

For many applications, this is a prohibitively expensive barrier to overcome. Recently, we introduced LodeSTAR, which addresses this challenge by enabling a NN to learn to perform near statistically optimal object localization, directly on experimental data, without any annotations [196]. This approach builds on the need for minimal training data and annotation, similar to the motivations behind advances in particle tracking discussed in section 16. This is achieved by devising a novel training scheme, where instead of providing target positions as ground truth, LodeSTAR is trained to produce two scalar values that are equivariant to Euclidean transformations applied to the experimental image. We show that the only solution to such a training scheme is for these two scalars to be the coordinates of the object's center, provided that the object has a well-defined optical center. Experimentally, we find that LodeSTAR approaches the Cramér-Rao bound of localization error for a range of geometries. This is particularly useful in applications where precise object localization is critical, such as in particle tracking, single-molecule localization, or super-resolution microscopy.

Next, we show that by carefully considering the design of the NN architecture, a localization model can be converted into a DL model without any additional training. In practice, this means that the localization model is imbued with translational equivariance through a center-of-mass pooling layer (where the mass is also learned by the model) of a convolutional backbone. By removing this pooling step, the model instead yields a spatially distributed coordinate field and a weight map, where clustered coordinate predictions that coincide with high weights are considered as detections. We find that LodeSTAR can be trained on a single or a small number of unannotated experimental images. In fact, LodeSTAR is shown to match the performance of state-of-the-art cell detection methods in the Cell Tracking Challenge [209] using 1000 times less training data (figures 30(a) and (b)). This makes it an ideal solution for researchers working in fields where annotated datasets are not available, or where the annotation process is time-consuming and costly.

Beyond detection, LodeSTAR can also perform certain regression tasks if there is a corresponding relationship to exploit. In many interferometric modalities, such as holographic microscopes, the focal plane can be numerically changed after-the-fact. This creates a relationship between the z-position of the object and a numerical repropagation transformation. Similarly, in many quantitative modalities, the strength of the signal is directly proportional to some physical property of the object. LodeSTAR has used these relations to position sub-wavelength polystyrene beads in 3D (figure 30(c)), and to measure the optical mass of intracellular aggregates (figure 30(d)).

Current and future challenges

While LodeSTAR achieves high performance with minimal training data, it should be noted that it currently only provides the position of objects, and not their size or morphological information. This may be a limitation for certain applications that require more detailed information about the objects being detected. Nonetheless, in applications such as cell counting and particle tracking, this is not a limiting factor. Indeed, low-data regimes are common across microscopy modalities, as also highlighted for single-molecule localization microscopy (section 14), where balancing data efficiency and precision is critical.

Another limitation of LodeSTAR is that it cannot efficiently utilize large amounts of data to gain more specificity. While it outperforms other methods in the low-shot regime (1–10 object instances), it may fall behind when unlimited data is available. This is a general problem with self-supervised learning, which is often less directed due to the lack of supervision.

Another limitation of LodeSTAR is that it is optimized for small to medium-sized objects, typically up to 60–100 pixels. While it has been shown to be effective for detecting objects of this size, it may not perform as well on larger objects. This is because the method relies on exploiting the inherent roto-translational symmetries of the task of object detection, which may not be as prevalent in larger objects.

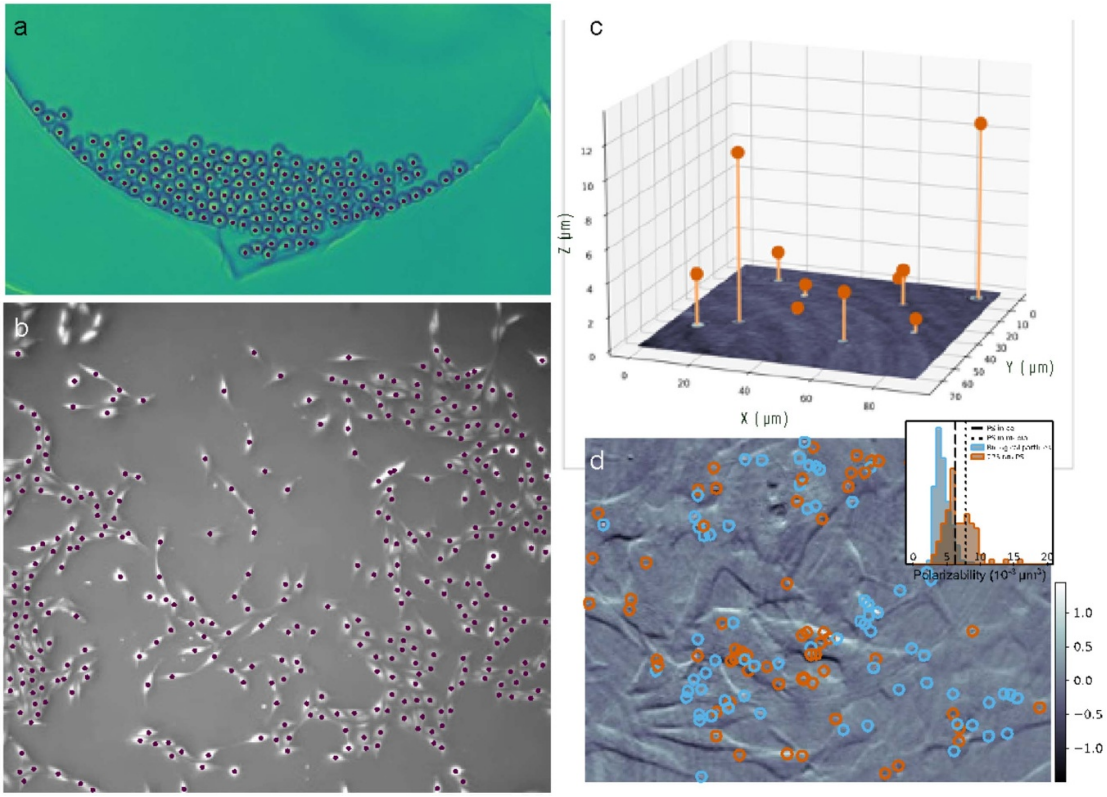

Figure 30. High quality detection with LodeSTAR. LodeSTAR achieves state of the art detection performance on a range of experimental configurations using only a single image of a single object instance to train. (a) LodeSTAR detects brightfield images of densely packed mouse stem cells. (b) LodeSTAR detects phase contrast images of pancreatic stem cells. (c) LodeSTAR detects and positions sub-wavelength polystyrene beads in 3D-space from off-axis holographic images. (d) LodeSTAR detects and measures the polarizability of intracellular objects (225 nm polystyrene beads in orange, biological aggregates in blue). Reproduced from [196]. CC BY 4.0.

This limitation may be less important for applications such as single-molecule localization, where the objects of interest are relatively small. However, it could be a limitation for other applications such as detecting large cells or structures in tissue imaging. In these cases, the issue can be resolved by sampling the image at a lower resolution. This improves detection at the cost of positional accuracy. However, for large structures, the notion of a single position is less well-defined and often less relevant to the applications.

Advances in science and technology to meet challenges

An important avenue for future research is to expand the set of transformations and symmetries that LodeSTAR utilizes in its learning process. I expect that exploring modality-specific transformations may be particularly fruitful. A prime example is imaging modalities capable of reconstructing the complex-valued light field, where a rich set of symmetries and transformations in Fourier space can be exploited. Another avenue would be to train LodeSTAR to produce morphological information. For example, by incorporating scale transformations, LodeSTAR could be taught to predict the relative sizes of objects. However, this alone would not be sufficient to provide absolute sizes without additional transformations.

Another approach could be to use a more interpretable network architecture, such as vision transformers (ViTs) [210]. These architectures have been shown to naturally produce segmentations of objects when solving consistency tasks. However, it is important to note that replacing the current convolutional backbone with a ViT would likely lead to a loss of translational equivariance, which may negatively impact LodeSTAR's positioning performance.

Currently, LodeSTAR is trained on exclusively positive samples, which are images containing the object to be detected. One potential area of research is to expand the training procedure to include negative samples, which are images with no instances of the object. This could help improve specificity in samples with multiple visually similar classes, where only a subset of the classes is of interest.

Lastly, LodeSTAR is currently designed to train on images with exactly one instance of the object per view, where detection capabilities are an emergent behavior resulting from the design of the model. Instead, it could be extended to allow multiple instances of the object per view during training, enabling

it to learn from larger datasets where the number of objects in view may vary. However, this would also decrease the specificity of the training, as all objects in view would be detected regardless of interest. Finding ways to mitigate this would be an important area of research.

Concluding remarks

LodeSTAR is a DL method for object detection in microscopy that is particularly well-suited for low-data regimes. LodeSTAR exploits the inherent roto-translational symmetries of object detection to achieve high performance with minimal training data. Its unique self-supervised training process allows it to learn from a single or a small number of unannotated experimental images, making it a cost-effective solution for researchers working in fields where annotated datasets are not available. Furthermore, LodeSTAR's ability to find the sub-pixel position of objects with high accuracy is a key feature that sets it apart from traditional methods. This level of precision is particularly useful in applications where precise object localization is critical, such as in particle tracking, single-molecule localization, or super-resolution microscopy. This, combined with LodeSTAR's ability to utilize additional relationships and symmetries to measure additional physical quantities, makes it a versatile and powerful method in the field of microscopy.

Acknowledgments

The authors would like to acknowledge funding from the H2020 European Research Council (ERC) Starting Grant ComplexSwimmers (Grant No. 677511), the Horizon Europe ERC Consolidator Grant MAPEI (Grant No. 101001267), and the Knut and Alice Wallenberg Foundation (Grant No. 2019.0079).

19. Force fields calibration

Aykut Argun[1], *Tobias Thalheim*[2], *Frank Cichos*[2] *and Giovanni Volpe*[1]

[1] Department of Physics, University of Gothenburg, Origovägen 6B, SE-41296 Gothenburg, Sweden
[2] Peter Debye Institute for Soft Matter Physics, Molecular Nanophotonics Group, Leipzig University, Linnéstraße 5, 04103 Leipzig, Germany

Status

Accurate calibration of microscopic force fields is fundamental for a wide range of experiments including OTs, DNA stretching, and non-equilibrium physics [211]. Traditionally, these force fields can be characterized by analyzing trajectories of Brownian particles that are subjected to them. This information then can be used to infer the underlying dynamics, forecast a future state of the particles, or to calibrate the experimental setup. This calibration sometimes needs to be done even in real-time [212].

One of the most commonly used microscopic force fields is the harmonic force field. This can be generated by an OT, where the force has the form $F_{\mathrm{h}} = -kx$ with k the stiffness of the harmonic trap and x the distance from equilibrium. For this case, there are already many standard calibration methods, such as the variance method, the autocorrelation method, and power spectrum methods [211]. The variance method directly determines the stiffness from the variance of the particle position in the trap ($k = k_{\mathrm{B}}T/x^2$, where $k_{\mathrm{B}}T$ represents the thermal energy). The autocorrelation method determines k by fitting the exponential decorrelation curve of the particle position in the trap. The power spectrum method is a further powerful method especially at high frequencies, which fits the power spectrum of the Brownian particle to a Lorentzian and infers the stiffness. All these standard algorithms work well if the measurements contain sufficient data points and are error free [211]. However, the need to develop calibration methods robust to noisy, incomplete data closely parallels the efforts in particle tracking (sections 16 and 17), where DL has similarly improved robustness under experimental imperfections. More efficient standard algorithms include linear maximum likelihood methods, like the FORMA algorithm [213], that apply a linear regression, which makes a first order approximation to the force around the equilibrium point. This permits a more accurate and faster calibration of force fields compared to standard algorithms [213] as well as the ability to calibrate non-conservative force fields, such as a rotational force field.

Current and future challenges

Due to the need for averaging, the calibration methods require more data for higher accuracy because the microscopic trajectories are stochastic. If there is enough statistical information in the data, averaging of displacements provides accurate force measurements [214]. However, in biophysics, this is frequently not achievable. Furthermore, the number of non-standard force field calibration procedures is extremely limited. Examples of such non-standard force fields range from non-harmonic potentials to rotational as well as dynamic non-equilibrium force fields. For complicated potential landscapes, the amount of data necessary for a correct measurement grows exponentially as the number of parameters to be calculated from a trajectory increases. Although the maximum likelihood algorithms like FORMA can perform better in these potential landscapes, they can solely do so provided that the data is acquired at a high frequency and the force field is not time-varying. Nevertheless, there are experimental systems that need real time calibration, e.g. because they undergo changes or a degradation of the setup over time. Examples of such systems are forces generated by electrodes inside liquid environments, by temperature gradients, chemical forces, and bacterial forces, with some of them being the results of feedback loops [212, 215] involving additional delays in the control process, which further complicates the calibration process.

Data acquisition can also be challenging in the fields of soft matter physics, biophysics, and single-molecule physics. More advanced approaches are enabling faster and more accurate localization of particles. Recent advances have allowed high-frequency trajectory measurements, where particle trajectories are also complicated by hydrodynamic memory [216]. In addition, the advancements in data-driven image analysis methods facilitated the possibility of enhanced image analysis [94].

Advances in science and technology to meet challenges

There is currently a growing need for more advanced calibration methods for experimental purposes. At the same time, data-driven analysis methods, such as RNNs, have proven to be a very powerful tool for extracting information from time series data. This has led to the development of DeepCalib [73], a free software package that uses DNNs to calibrate force fields from trajectories. Specifically, DeepCalib

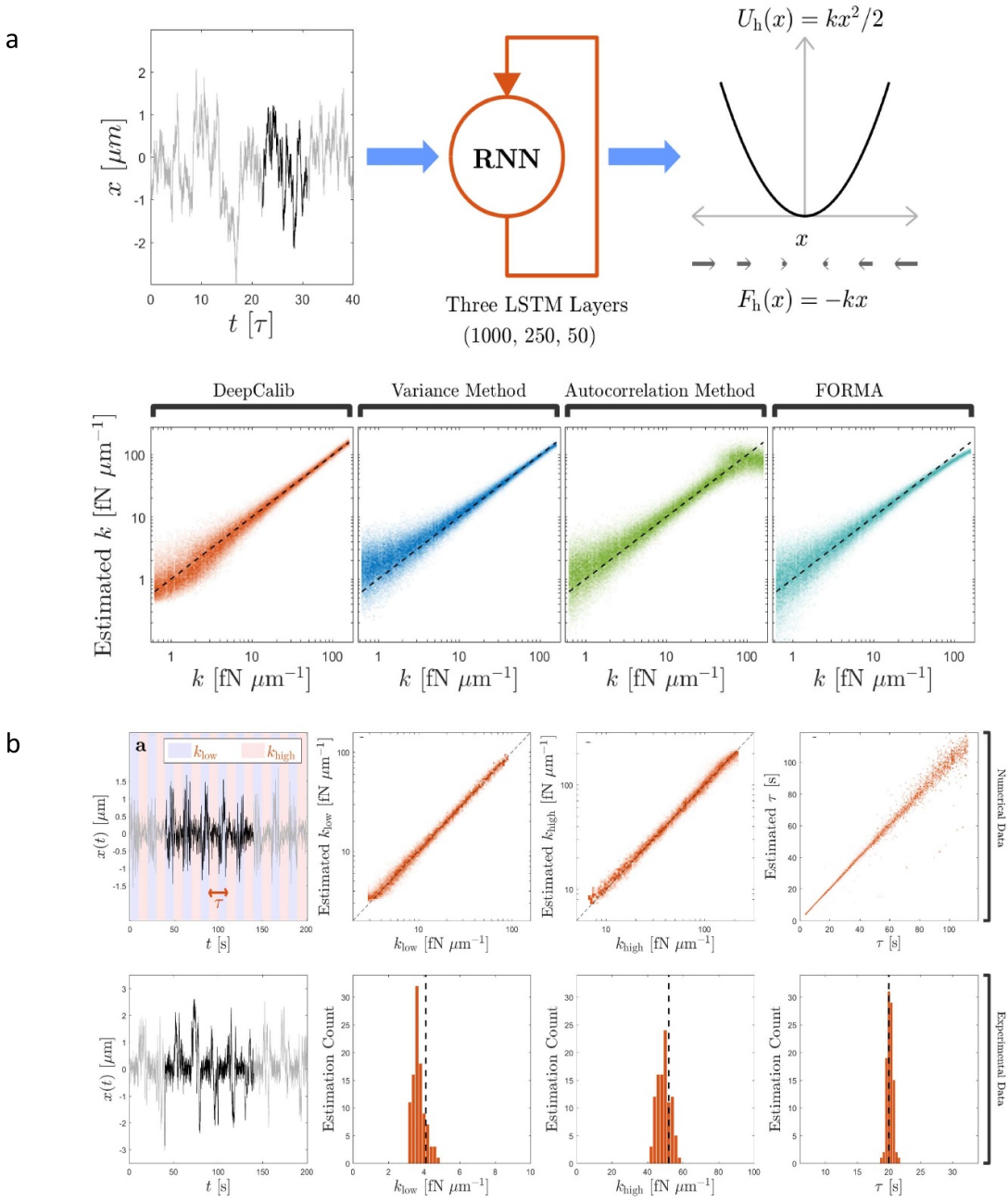

Figure 31. Calibration of force fields with deep learning. (a) The trajectory of a Brownian particle in a harmonic trap is given as an input to the recurrent neural network (RNN) to predict the stiffness parameter of the trap (k). The bottom panel shows comparisons of the predictions from various methods. As it can be seen, DeepCalib provides the most accurate results through the entire range. (b) Calibration of a dynamical non-equilibrium force field, where the stiffness of a harmonic trap is oscillating between k_{low} and k_{high} at a period τ. There are no existing calibration methods for such a force field that is time-varying. Evidently, DeepCalib successfully predicts all parameters both for simulated (top panel) and experimental (bottom panel) trajectories. All experimental data were generated employing the versatile feedback mode of a thermophoretic trap. Reproduced from [73]. CC BY 4.0.

uses LSTM (long short-term memory) layers to extract information from trajectories and uses simulated data to train the NN. Particularly for shorter trajectories or smaller force values, when the calibration is challenging, DeepCalib is proven to work better than other algorithms for harmonic potentials as shown in figure 31(a). In addition, DeepCalib can accurately calibrate non-conservative rotational force fields better than the existing methods. In this case, the force field has two parameters: the central force stiffness k and the rotational parameter Ω. Although being accurate for lower values of k, FORMA struggles at higher forces because of the data points becoming uncorrelated. DeepCalib is able to estimate the parameters of the force fields better than FORMA. The difference of DeepCalib turns

out to be even more prominent as we look at further non-standard cases, such as a double-well potential. Here, the force field is parametrized by the equilibrium distance L and the energy barrier height E_B. DeepCalib provides more accurate results when predicting the parameters of this double-well potential. Finally, and most importantly, DeepCalib can calibrate any force fields, such as time varying force fields that yield a dynamical non-equilibrium system. In this case, a harmonic potential with switching stiffnesses from k_{low} to k_{high} at a period τ is considered. It is shown in figure 31(b) that DeepCalib can very accurately calibrate all the parameters of a short trajectory for such a challenging case, both for numerical and experimental data. It is also shown that DeepCalib is more robust against measurement noise and diffusion gradients [73]. Current experimental developments also help to reduce the reality gap by testing DeepCalib against complex force fields generated in the feedback mode of the thermophoretic trap [215]. This setup generates dynamic temperature fields on circular nanostructures in response to the current position of a Brownian particle in liquids. The response may thereby follow a user designed protocol that permits almost arbitrary dynamic free energy landscapes [215] to test DeepCalib.

Concluding remarks

With the introduction of DeepCalib it becomes evident that a data-driven, NN approach for the calibration of microscopic force fields outperforms the standard methods in challenging conditions and limited available data. More importantly, DeepCalib can be applied to non-conservative or time varying force fields for which no standard calibration methods exist. With these advantages and the great robustness to measurement noise, inhomogeneous environments, and the reality gap between experiments and theory, DeepCalib is a very powerful and flexible tool for analyzing trajectories to extract the force fields. It allows for just a minor change in the code to immediately adapt to a new force field, while standard techniques totally change if a different force field is considered. Therefore, DeepCalib is ideal to calibrate complex and non-standard force fields from short trajectories and it is readily available as a free Python software package [217]. It also proves that the RNNs are a very powerful tool to analyze Brownian trajectories. Similar techniques have also been successfully applied to characterize, for example, anomalous diffusion [218].

Acknowledgments

A. A. and G. V. acknowledge support from H2020 European Research Council (ERC) starting Grant ComplexSwimmers (Grant No. 677511). T. T. and F. C. acknowledge support from the SFB TRR102 'Polymers under multiple constraints: restricted and controlled molecular order and mobility' (DFG Project Number 189853844).

20. Diffusion characterization

Stefano Bo[1,2] *and Lars Hubatsch*[1,3]

[1] Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Straße 38, DE-01187 Dresden, Germany
[2] Department of Physics, King's College London, London WC2R 2LS, United Kingdom
[3] Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstraße 108, 01307 Dresden, Germany

Status

Live imaging has revolutionized our understanding of biological processes. Resolution and signal-to-noise ratio have steadily improved over the past decades. As a consequence, the trajectories of single particles can be recorded, enabling detailed descriptions of microscopic dynamics while avoiding some artefacts of bulk labeling. Single particle tracking (SPT) unveils features of the surrounding environment and of processes that would be lost by ensemble averaging, thus opening an observational window that provides a deeper understanding of many biological processes.

To probe the molecular properties in heterogeneous environments and to enable quantitative modeling of biological processes, it is important to tell whether particles undergo normal or anomalous diffusion (defined by the power-law scaling of the mean-squared displacement: MSD $\sim t^a$) and to measure their diffusion coefficients or anomalous scaling exponent a. Motion at these scales is inherently stochastic and can involve processes characterized by significant variability in time and space. This makes the study of diffusion from particle trajectories difficult. In figure 32, we outline a typical pipeline to meet the challenges associated with studying diffusion leveraging DL techniques.

A first challenge consists in identifying the particles, localizing them and reconstructing their trajectories. Various methods have been developed for these purposes in the past decade [219], and DL techniques are gaining considerable attention [170].

The second main challenge is the extraction of information from the reconstructed trajectories. This is complicated by the fact that diffusion is a phenomenon related to particle fluctuations; in other words, the signal is in the noise. Even for simple diffusion, the direct application of Einstein's 1905 result MSD $\sim$ Dt can result in severe biases, for example, depending on the localization precision, track length, number of available tracks, or fluctuations due to movement of the surrounding media. To tackle these issues, methods accounting for experimental noise and optimal fitting algorithms for different use cases were developed [220]. Anomalous diffusion dynamics, which can be modelled, for example, by *continuous time random walks* (CTRWs), *fractional Brownian motion* (FBM) and *Lévy walks*, are even more challenging to analyze due to long-term correlations in the dynamics and possible ergodicity breaking [221]. Most of the methods to analyze such dynamics rely on asymptotic behavior, requiring many long trajectories, which are typically difficult to obtain experimentally. This has motivated the development of DL techniques to measure the diffusion coefficient, the anomalous exponent and to identify the model underlying the anomalous behavior [221–224]. Such applications of DL to infer underlying physical dynamics from noisy, short time-series data builds also on advances in particle tracking such as those covered in sections 16 and 17. These techniques typically involve artificial CNNs or RNNs named LSTM. They have remarkably improved the analysis of short individual trajectories, also including cases where individual trajectories switch between different diffusive dynamics.

Current and future challenges

A promising feature of DL techniques is that they have the potential to generalize, i.e. correctly analyze scenarios that differ from the training data. For applications to single-particle trajectories this suggests that algorithms trained on synthetic data, which are cheaply generated by simulation, can be applied to experiments. The techniques developed in references [221–224] are networks trained on simulated data, using supervised methods, i.e. the parameters used to simulate the input data, such as diffusion coefficient, anomalous exponent or the type of model, are known and used to label the input. The networks are then asked to return a prediction that is as close as possible to the label. These methods have proven their versatility, correctly predicting experimental trajectories that they had not seen during training, and the ability to deal with measurement noise. These promising examples, however, do not generally guarantee that networks trained on simulated data can be carried over to experiments. To assess the issue of generalization, more work is required to understand what leads to a network's decision. This would contribute to making the predictions more interpretable and transparent. Interpretability can significantly extend the scientific applicability of DL models to diffusion, linking network behavior to the

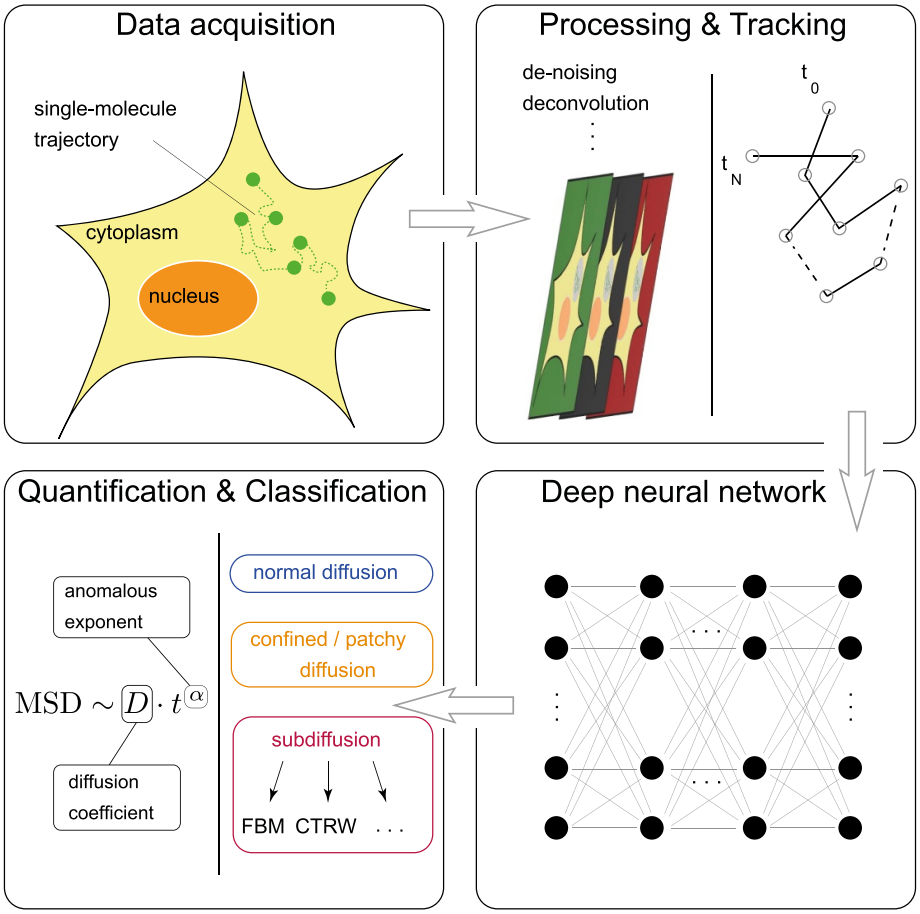

Figure 32. Typical pipeline to characterize diffusion from experimental single-particle trajectories using a deep learning method. The pipeline starts with data acquisition, followed by data processing, which may include tracking methods and de-noising steps. The processed data is then fed to a pre-trained deep neural network that typically returns a classification of the acquired data or a quantification of its physical properties.

underlying physical principles and structures. Much like developing a physical theory provides a mechanistic understanding of a physical phenomenon and goes beyond measurement and inference, improving the interpretability of DL contributes to deeper physical understanding. Furthermore, understanding how DL methods estimate diffusion from particle trajectories might contribute to the development of new algorithmic and statistical techniques, which are inspired by the solution found by the DL methods. Indeed, the progress in interpretability might pave the way towards learning from the machines. However, due to their high dimensionality and many parameters, it is notoriously difficult to unveil the inner workings of DL models.

An additional challenge is related to supervised methods. These methods, by their very construction, can only predict the labels they were taught during training. For instance, a method trained to only infer the diffusion coefficient will not be able to detect anomalous diffusion or its scaling exponent. Similarly, for the classification of the dynamics, typical networks are bound to return an answer within the classes they have been designed for. Recent work has begun addressing this issue [225]. A related challenge concerns the use of DL to capture the transient nature of anomalous diffusion, where different scaling regimes are observed at different scales.

Advances in science and technology to meet challenges

The field of DL is evolving rapidly, providing insights that can be applied to various sub-domains. For instance, several tools for the interpretability of DL techniques are currently being developed (see, e.g. [226]). The most promising results in this direction concern DNNs, such as CNN, employed for the classification of images. Similarly, self-supervised methods developed for object detection in microscopy (section 17) suggest promising strategies for exploring structure in unlabeled or weakly labeled diffusion data. These novel tools identify the most significant parts, or features of the images that led to their respective classification. For SPT, one could apply these algorithms (which should also be extended to LSTM) to determine which parts of a trajectory most heavily influence its classification.

In parallel, autoencoder and transformer architectures have been developed [227]. Transformers are currently among the best techniques to process natural language, for example, for translation purposes. They feature attention layers in their architectures, whose role is to identify how much different parts of an input are related to each other (e.g. learning that the subject in a sentence is strongly related to the verb). Notably, autoeconders can be used to generate diffusive trajectories [228].

The combination of the development of tools to interpret existing DL architectures and of more interpretable architectures offers the possibility of identifying the parts of a diffusive trajectory that are the most informative ones to characterize it. One can then perform statistical analysis on these results to extract information about what features or statistical properties of the trajectories are the most relevant to characterize their diffusive nature. This sheds light on how the DL methods make their predictions, contributing to their interpretability.

Additional advances in deep unsupervised learning can provide methods that are more insightful when applied to trajectories belonging to none of the models they have been trained on (see [229] for a first step in this direction using anomaly detection). Indeed, these models are not trained using specific labels but focus on reconstructing the input trajectory, thereby learning its salient features. Unsupervised methods might also successfully address the challenges posed by trajectories featuring different diffusive scalings at different scales.

The advances discussed so far concern general improvements for the interpretability of DL methods. Additional progress is likely to come from the development of physics-informed methods for the study of diffusion [230]. Such methods are designed to take into account the physical knowledge about the diffusive dynamics under consideration, for example considering key physical and statistical properties, such as the self-similarity of certain anomalous diffusion dynamics.

Concluding remarks

DL methods have proven to be invaluable tools for the study of diffusive dynamics, displaying remarkable robustness and showing promising potential to generalize results to unseen data and experiments. They excellently complement traditional algorithmic and statistical techniques, providing better performance. This makes it possible to characterize short individual trajectories, thereby expanding the range of systems for which diffusion can be quantitatively investigated, including many biological systems, for which track lengths are typically short. At the current stage, this performance improvement comes at the cost of a lower interpretability of the results compared to traditional algorithmic and statistical techniques. However, recent and ongoing progress in DL techniques with the development of novel architectures and methods to interpret their prediction together with physics-informed approaches suggest that more transparent DL techniques might be within close reach. These techniques have the potential to reliably unveil the physical properties of diffusive particles and the environments in which they move. These advances provide an important building block for future quantitative models of complex biological processes.

Acknowledgments

SB gratefully acknowledges inspiring discussions with Aykut Argun, Giovanni Volpe, Ralf Eichhorn, Alessia Gentili, Giorgio Volpe and the participants and organisers of the AnDi workshop hosted at ICFO in December 2021.

21. Microscopic motion characterization with MAGIK

Jesus Pineda[1], *Giovanni Volpe*[1] *and Carlo Manzo*[2]

[1] Department of Physics, University of Gothenburg, Gothenburg, Sweden
[2] Facultat de Ciències, Tecnologia i Enginyeries, Universitat de Vic—Universitat Central de Catalunya (UVic-UCC), Vic, Spain

Status

The characterization of dynamic processes in living systems provides essential information for advancing our understanding of life processes in health and diseases, as well as for developing new technologies and treatments [231]. In the past two decades, optical microscopy has undergone significant developments, enabling us to study the motion of cells, organelles, and individual molecules with unprecedented detail at various scales in space and time. However, analyzing the dynamic processes that occur in complex and crowded environments remains a challenge [209, 219, 221]. Similar challenges regarding variability and complex environmental interactions are encountered in the particle tracking workflows discussed in section 16. The diversity of biological motion is a significant contributor to the complexity of dynamic analysis. Cells, for example, exhibit a high degree of variability in their motion depending on factors such as cell type and environmental cues. This variability poses analytical challenges for current methods, which are typically designed for specific experiments or motion models. Thus, the analysis of experiments often requires manual adjustments of parameters to adapt to different dynamics, limiting their utilization and applicability.

Recently, we have proposed MAGIK [199], a DL framework for the analysis of biological system dynamics from time-lapse microscopy. MAGIK models the movement and interactions of particles through a directed graph where nodes represent detections and edges connect nodes that are spatiotemporally close (figures 33(a) and (b)). The framework utilizes an attention-based GNN to process the graph and modulate the strength of associations between its elements through two mechanisms. The first mechanism is a learnable local receptive field [232] that captures the complexity of local particle interactions. The second is a gated self-attention mechanism [233], enabling MAGIK to derive insights into the dynamics of each particle from regions within the graph that are not directly connected but provide valuable information about the overall dynamics.

MAGIK is a versatile tool capable of performing various tasks, from linking coordinates into trajectories to determining local and global dynamics (figure 33(c)). As shown in figure 34(a), MAGIK offers reliable performance in tracking HeLa cells, despite the challenges posed by the heterogeneity in cell shape and dynamics. MAGIK accurately identifies cell divisions and estimates trajectories in edge regions where cells are partially observed and move out of the field of view. Furthermore, MAGIK delivers outstanding results on several datasets from the 6th Cell Tracking Challenge across a range of microscopy techniques and cell types [199].

A unique feature of MAGIK is its ability to characterize dynamic aspects without the need for detection linking. In this way, MAGIK can provide information from high-density experiments, e.g. by resolving a spatially modulated diffusion landscape solely from particle localizations (figures 34(b) and (c)). Notably, most spatial features are accurately estimated from a 100-frame-long video. By skipping the linking step, MAGIK inherently reduces the propagation of linking errors to the quantification of relevant parameters and can thus unveils global dynamics, as shown for two relevant examples: the classification of the mode of motion of diffusing particles (figures 34(d) and (e)), and the quantification of anomalous diffusion [221] from ensembles of Brownian particles (figure 34(f)).

Current and future challenges

The rapid progress in DL has led to the development of various techniques for characterizing motion in biological systems [209, 219, 221]. However, several technical and interpretative challenges hinder the widespread use of these algorithms.

The scarcity of high-quality annotated microscopy datasets hampers the effective training and validation of DL models. Available datasets are often small and unrepresentative, making it difficult to create large and dependable models. In addition, the diversity of microscopy data and limited access pose further challenges to their utilization and integration.

DL models typically learn complex representations of the input data in an abstract, high-dimensional space, making interpretation of these abstractions difficult even for experts. A comparable issue of interpretability and trustworthiness also arises in force characterization (section 18) and diffusion characterization (section 19) from particle trajectories. This lack of interpretability is a barrier to the widespread

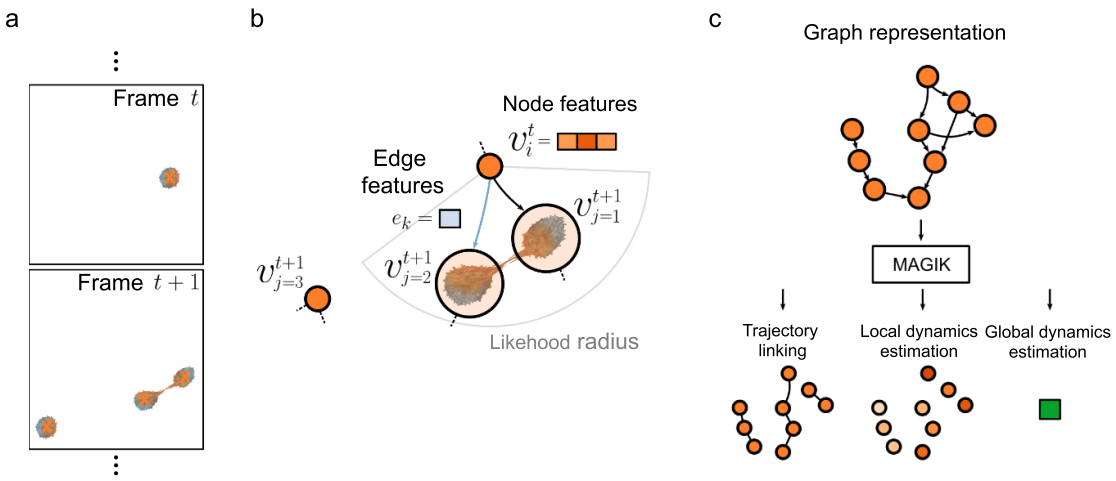

Figure 33. Estimation of spatiotemporal features using MAGIK. (a) Sequence of images illustrating the evolution of a group of cells over two consecutive frames. (b) The movement and interactions of the objects are modelled geometrically using a directed graph. In this graph, nodes (v) represent detections and edges connect spatiotemporally close objects. Each node contains the object's centroid and other important descriptors, such as morphological and intensity attributes. Edges (e), in turn, encode Euclidean distance between the centroids of the connecting objects. In this example, the node of interest (labeled with the subindex i and located in frame t) is connected to neighboring nodes in frame $t+1$ (labeled with the subindex j) within a distance-based likelihood radius. (c) The graph is processed through an adaptive attention-based GNN that estimates the associations among the objects and provides insights into the intrinsic dynamics of the systems at different scales. Reproduced from [199]. CC BY 4.0.

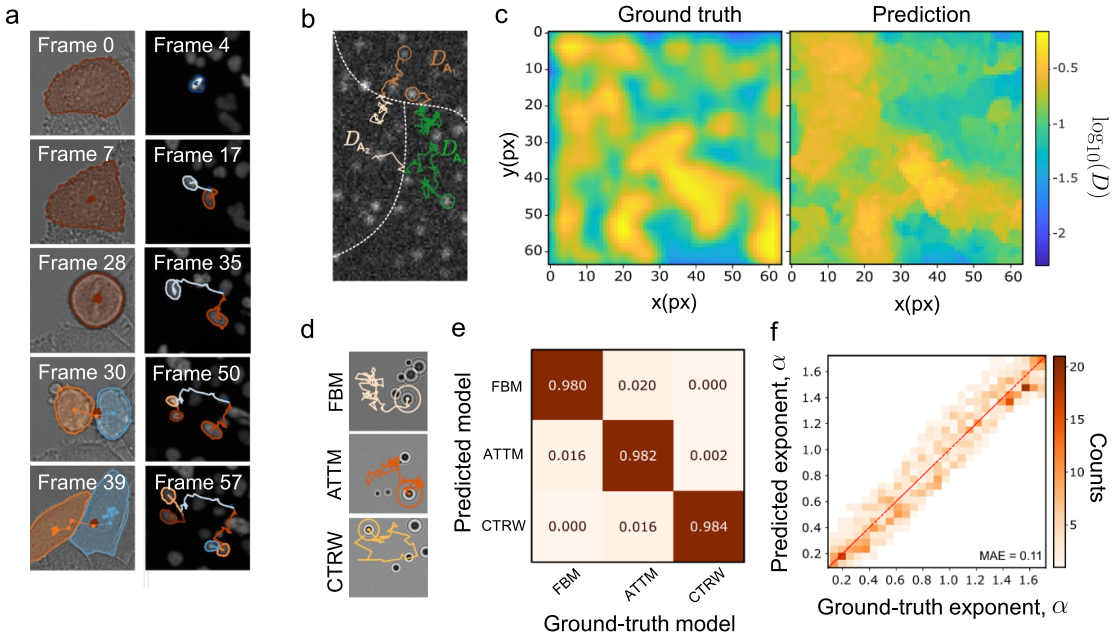

Figure 34. MAGIK analyzes a broad range of time-lapse biological experiments at different scales. (a) MAGIK effectively links trajectories across various experimental conditions. MAGIK demonstrates a 99.2% accuracy rate in tracking the movements of HeLa cells from the DIC-C2DH-HELA dataset, despite their changing shape and repeated division. Likewise, MAGIK achieves 98.4% accuracy in tracking HeLa cells from the Fluo-N2DL-HeLa dataset, despite the dense sample and frequent mitosis and collisions. (b)–(c) MAGIK estimates a spatially modulated diffusion landscape. (b) Fluorescence microscopy simulation of single-object tracking experiment, where molecules undergo Brownian motion with diffusivity D randomly varying in space. (c) Ground-truth spatial diffusivity pattern and prediction obtained by MAGIK using a 100-frame-long video. (d)–(e) MAGIK unveils global dynamics. (d) Simulated holographic videos, where objects follow fractional Brownian motion (FBM), annealed transient time motion (ATTM), and continuous-time random walk (CTRW). (e) Confusion matrix showing the network classification performance of the underlying diffusion model presented in 1.496 validation scenarios. (f) MAGIK predicts the anomalous diffusion exponent governing the motion of ensembles of objects performing FBM in 1.097 holographic videos. Reproduced from [199]. CC BY 4.0.

adoption of DL in medical and biological fields, particularly in contexts outside of research where interpretability is crucial [234]. Moreover, the proliferation of multiple methods for the same task without a clear evaluation of their performance can confuse non-experts and limit their usage.

The scalability of DL graph models is a major technical challenge. Processing large graph representations with numerous interactions is computationally demanding and requires significant memory and computing resources. This makes it challenging to directly use GNNs for analyzing dense and lengthy dynamic experiments.

Advances in science and technology to meet challenges

The MAGIK framework is continuously improved to address the limitations in the current DL approaches for motion characterization in biological systems. The focus of MAGIK's development is to create a framework that is both general and easily convergent. It has shown the capability to train using just one labeled video by utilizing tools that maximize information extraction from limited data, ensuring proper representation and stability during NN training [199]. MAGIK also uses transfer learning for migration experiments, enabling the trained network to be applied to other cell data without any reduction in performance, as shown in the MAGIK GitHub repository [235]. Further advancements in this regard are desirable through the implementation of self-supervised algorithms.

MAGIK is equipped with attention mechanisms that provide users with interpretability into the specific aspects of the data structure that the framework focuses on when making predictions. This approach resonates with broader efforts in computational microscopy, such as integrating attention and interpretable learning strategies in FLI (section 10). This offers a reliable and efficient method for analyzing dynamics and provides opportunities for discovering new features in the movement of living systems.

MAGIK is included in the DL package for microscopy, DeepTrack 2.1 [94], and is, therefore, undergoing continuous development and optimization with a focus on scalability and deployment improvements. Future efforts will focus on developing self-destillation-guided graph subsampling techniques [236] and resource-efficient GNN architectures [237].

Concluding remarks

DL frameworks for the analysis of biological system, such as MAGIK, can successfully handle the complexities of dynamic analysis in complex and crowded environments utilizing an attention-based GNN. MAGIK can perform various tasks, including tracking cells, determining local and global dynamics, and characterizing dynamic aspects without the need for detection linking. Despite these progresses, technical and interpretive challenges that hinder widespread use of these tools. With the continuous development of the MAGIK framework within the DeepTrack 2.1 package, we aim to address these limitations and offer a generalizable tool able to further provide interpretability into the data structure.

22. Quantification of subcellular dynamics from SPT

Carlo Manzo

Facultat de Ciències, Tecnologia i Enginyeries, Universitat de Vic—Universitat Central de Catalunya (UVic-UCC), Vic, Spain
Bioinformatics and Bioimaging, Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), Vic, Spain

Status

Time-lapse microscopy images of biological processes are widely used to observe the dynamics and behavior of live cells and unicellular microorganisms, with applications ranging from fundamental aspects of molecular and cell biology to medical practice. The development of single-molecule imaging and super-resolution microscopy has further extended the capability to resolve the dynamics of biological processes, reaching the subcellular and molecular scales [238]. The current technology thus enables the visualization of the motion of organelles, proteins, and lipids in their native environment. The observations provided by these experiments are valuable to decipher the interactions between cellular components and to disclose their role in fundamental processes such as signaling and function regulation. They are also helpful for biomedical applications related to pathogen infection and drug design. Nowadays, microscopes capable to perform live-cell single-molecule imaging are accessible in many research laboratories and, therefore, experiments are routinely performed. However, mining quantitative information from these experiments still poses several challenges.

Typically, the analysis pipeline is divided into two steps: SPT and trajectory analysis (figure 35). These standard pipelines share important conceptual similarities with those described for particle tracking (section 16) and diffusion characterization (section 19), both of which face similar challenges of noisy trajectories and complex dynamic behaviors. In the first step, the information contained in the image stream is converted into trajectories, i.e. time series of features associated with imaged particles (such as position, intensity, and size). Considerable efforts have been dedicated to developing automated algorithms for this task [219]. Relying on the advances of the research community in CV and multi-object tracking, the tracking-by-detection paradigm has gained increasing prevalence for SPT. Thus, images are first processed to detect features (detection), then features obtained at different times are connected using assignment algorithms to obtain trajectories (linking) (figure 35).

Once the trajectories are obtained, they are analyzed using statistical methods to extract information about the underlying dynamics of the particle, using estimators such as the MSD (figure 35). These analyzes aim at providing details about the type of transport being observed (Brownian, directed, confined, or anomalous), interactions with other particles and/or with the surrounding medium. Trajectories are also used to estimate biophysical parameters (e.g. the diffusion coefficient) or to determine whether the motion is compatible with a given theoretical model.

Life scientists dispose of a variety of algorithms to precisely track individual particles in living biological systems as well as many methods to interrogate trajectories. Recently, approaches based on DL have also been proposed, claiming remarkable improvements. The objective assessment of the performance of these methods is thus required to help end-users to pick the suitable tool.

Current and future challenges

Live-cell single-molecule imaging experiments typically record the motion of a subpopulation of individual particles (molecules, viruses, organelles) taking place in heterogeneous environments with the objective of detailing the molecular mechanism of transport and interactions with the environment. Technical and instrumental drawbacks impose limitations on the experimental conditions (e.g. the density of imaged particles and the temporal resolution) and affect quantitative parameters (e.g. the localization precision and the trajectory length) that eventually impact the precise characterization of the system [219]. A current challenge entails deploying approaches that can improve the performance of the methods that carry out the individual steps of the traditional analysis pipeline. In the last years, ML and single-molecule localization microscopy have produced a surge of methods for single-molecule detection and localization, mainly based on CNNs [177, 239]. More recently, DL approaches have also been proposed for the trajectory linking task [199]. These methods aim to provide an automated, unbiased, and reliable analysis of the image stream. The improvement of their performances enables experiments to be performed at faster image acquisition rates and higher labeling densities, increasing the temporal resolution and the spatial sampling. Similar advances aimed at enabling dense, high-speed data acquisition are also central in the development of motion analysis frameworks like MAGIK (section 20).

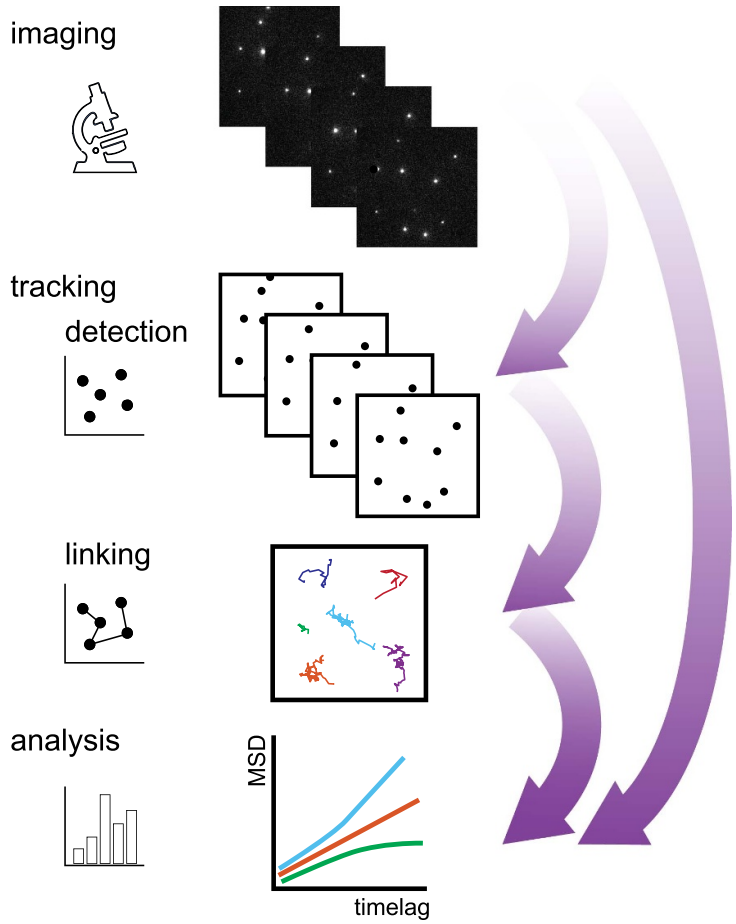

Figure 35. Steps of the analysis of live-cell single-molecule imaging experiments. The imaging experiments record the motion of individual particles. In the conventional analysis pipeline, images corresponding to individual frames are first processed to detect features (detection). Then, features are connected to obtain trajectories (linking). Finally, trajectories are analyzed to obtain quantitative information, e.g. by the calculation of the mean-squared displacement (MSD). Deep learning approaches are providing substantial improvements with respect to classical statistics and might further enable the direct extraction of dynamic features from unprocessed videos.

Because of the variety of phenomena taking place inside living cells, numerous approaches focusing on different aspects of particles' motion have been proposed to analyze trajectories. A challenging aspect is the characterization of individual trajectories, in particular when experimental conditions limit their length or localization precision. Very recently, pioneering works using ML have shown substantial improvements with respect to classical statistics and have demonstrated the ability of several architectures (random forest, convolutional, and RNNs) to provide the precise estimation of parameters (such as the diffusion coefficient for Brownian motion and the anomalous diffusion exponent) as well as trajectory classification, either with respect to the diffusion mode (described as immobile/confined/Brownian/directed or sub-/Brownian/super-diffusive) or to the underlying physical model [222, 223, 240, 241]. These results have led to the organization of the first anomalous diffusion (AnDi) challenge [221], a competition to objectively assess these methods, which fostered the development of novel approaches with outstanding performance [221]. The challenge also featured a task on trajectory segmentation for the detection of changes in dynamic behavior associated, e.g. with interactions with the environment.

Due to its implications for the characterization of biological systems, the development of trajectory segmentation methods for the detection of transient and short-lived events has recently gained further momentum, leading to the organization of the second AnDi challenge, which aims to evaluate and benchmark computational tools that can detect dynamic state transitions within individual trajectories, uncovering short-lived interactions that may indicate binding events or transient confinement [242].

Advances in science and technology to meet challenges

DL methods developed for both SPT and trajectory analysis outperform classical statistics counterparts in a wide range of conditions and promise to relax experimental constraints, providing more information at a faster speed from live-cell single-molecule imaging. However, several of the proposed

methods remain stuck at the stage of proof-of-principle and do not reach a widespread application in actual experiments. Various reasons might be contributing to this process. Most of the methods introduced so far involve supervised learning, but the lack of annotated data for this kind of experiment forces the training and validation over simulated datasets. Despite the realism of the simulations, the transfer learning to actual data might generate concerns from end-users, in addition to the black-box model concern. As done in other fields, progresses in this sense might be obtained by creating community efforts aimed at (i) producing public datasets to evaluate novel methods; (ii) periodically benchmarking existing methods using objective metrics to determine the state-of-the-art. It must be also considered that to work optimally, DL architectures must be trained on simulations reproducing the specific experimental conditions. It is thus recommendable to implement user-friendly interfaces to help non-experts to train and fine-tune the model.

The DL approaches implemented for SPT have incrementally improved on existing methodologies but have so far been bound to follow the standard analysis pipeline, providing data-driven versions of conventional approaches. A leap forward might be taken by developing 'tracking-free' methods capable of directly extracting dynamic features from unprocessed videos. Such approaches might be based on geometric DL or physics-informed ML architectures that include informative priors, i.e. physical constraints and inductive biases, on top of the observational data. This mirrors broader efforts across microscopy, such as physics-informed learning in computational phase microscopy (section 8) and particle tracking (section 17). In fact, besides producing faster training and more accurate predictions, these architectures will also increase the interpretability of the model. The direct use of raw data would also prevent the propagation of errors generated at the different steps of the pipeline that finally impact the extraction of dynamic information.

The development of techniques not requiring labeled datasets might further accelerate the application of DL to real data. Both unsupervised and self-supervised learning methods are advancing at a very rapid pace. In combination with innovative network architectures such as transformers, self-supervised techniques have demonstrated the ability to learn representations from unlabeled data, achieving outstanding results for image-based analysis. As such, they represent a promising approach for the development of the next generation of tools for SPT and analysis.

Concluding remarks

The DL revolution is yielding exciting perspectives for the quantitative analysis of live-cell single-molecule imaging. Yet, while recently developed tools are demonstrating their gain in performance, they still have important challenges to overcome to reach a widespread use of data-driven methods over classical tools. In this sense, it is advisable to promote community-driven actions to benchmark and validate methods [219, 221, 242]. Stepping beyond the tracking-by-detection paradigm might lead to a boost in performance by digging information directly from raw data. The integration of physical constraints and inductive biases into DL models can improve performance and interpretability. It is foreseeable that novel techniques not requiring labeled datasets will further boost this field and enable the study of molecular dynamics in living cells beyond current capabilities.

Acknowledgments

CM acknowledges support through Grant RYC-2015-17896 funded by MCIN/AEI/10.13039/501100011033 and 'ESF Investing in your future', Grants BFU2017-85693-R and PID2021-125386NB-I00 funded by MCIN/AEI/10.13039/501100011033/and FEDER 'ERDF A way of making Europe', and Grant AGAUR 2017SGR940 funded by the Generalitat de Catalunya.

23. Plankton life trajectories

Harshith Bachimanchi[1] *and Erik Selander*[2]

[1] Department of Physics, University of Gothenburg, Sweden
[2] Department of Biology, Lund University, Lund, Sweden

Status

Life likely began in the oceans 3.8 billion years ago [243]. Early cyanobacteria spread across the oceans and oxygenated the atmosphere [244]. Since then, the phytoplankton—the microscopic equivalents of terrestrial plants—have played a pivotal role in the Earth's ecosystem. Phytoplankton generate approximately half of the oxygen produced on Earth and fix 50 Peta-grams (Pg) of carbon every year, around five times the total emissions from fossil fuels [245]. Yet, because of their smaller size, our understanding of the lower aquatic food-web in many aspects is limited.

In order to effectively interrogate these organisms at a closer level, we depend on microscopy-based methods. Some of the application scenarios in this direction are plankton detection and counting, plankton segmentation and species classification, and long-term tracking of plankton cells. However, manual identification of single cells is a labor-intensive process considering the extraordinary diversity among plankton taxa that require trained taxonomists [246]. This also limits the possibilities to monitor and follow plankton communities with adequate resolution in time and space. Over the past decade, a multitude of CV-based imaging methods have been developed for object detection from microscopy images. These techniques range from traditional approaches that follow segmentation and binary thresholding methods, to more advanced techniques based on DL [247]. These trends parallel the shift from handcrafted to DL-based segmentation and classification methods observed in biomedical microscopy, particularly in SPT and cellular imaging (see, for example, sections 14 and 21).

Particularly, DL-based methods for plankton analysis have seen considerable success in recent years. Emerging as an alternative approach to established methods, DL offers objective schemes for investigating microorganisms in different environments. For example, a family of CNNs such as Faster R-CNN, U-Net, and YOLO are widely used for object detection, classification and segmentation problems [247]. The convolutional layers in these networks help to identify the high-level and low-level morphological features of cells in the image in contrast to traditional approaches. A family of generative models known as GANs are being used for generating new plankton imagery data from the existing data, to better evaluate the existing models and to further improve the accuracy in measurements [248]. Efforts are also being directed towards new problems such as quantitative tracking of planktons, and biomass estimation through DL [249].

Current and future challenges

The rapid development of AI inspires applications in plankton analysis, such as automatic identification and high volume and throughput *in-situ* monitoring efforts. In parallel there is a more demand-driven development to fine-tune and improve existing methods and identify techniques that work with the limited amount of labeled data, which is typically the bottleneck with DL based methods. Below we discuss the most pressing challenges.

Manual annotation of data: To train a DL network, one needs labeled data. Which means if an image contains an organism, we need experts identifying the organism to species level based on taxonomic knowledge from the literature and experience. This is a laborious task and often prone to errors. The problems with manual annotation of data can be tackled with unsupervised and self-supervised DL algorithms. Unsupervised classification of organisms, for example, can be performed by investigation the latent space distribution of variational autoencoders, followed by a subsequent evaluation of the predicted clustering by taxonomic experts.

Expanding the scope towards applications: DL applied research in plankton ecology is mostly restricted to detection, counting, and segmentation problems. This needs to be leveraged for more practical application scenarios to better understand the plankton dynamics. For instance, there is need to develop AI based tracking algorithms, that can track individual plankters over extended time periods. Mechanistic understanding of individual interactions such as predation, and resource exploitation are underexplored

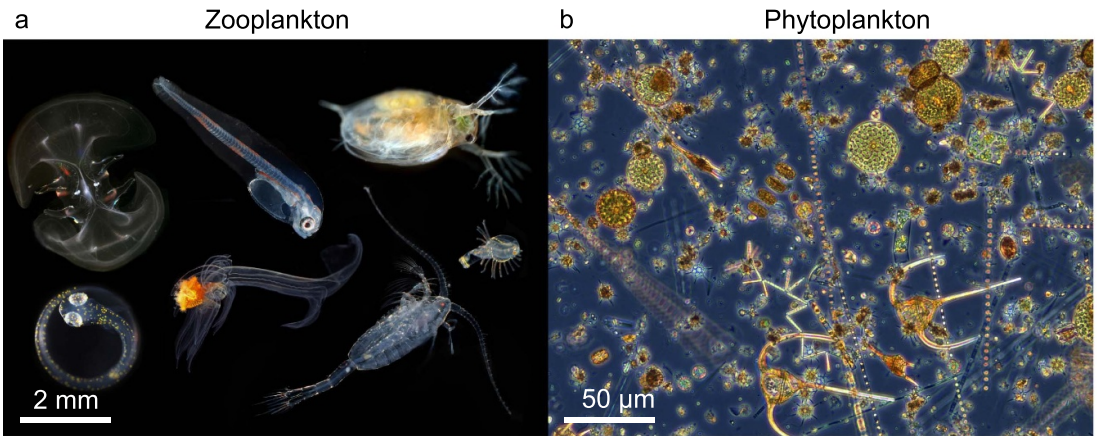

Figure 36. Plankton. Example images of zooplankton (left) and phytoplankton (right, photo credit Malin Mohlin) organisms. Plankton, as a broad term includes morphologically diverse organisms with complex geometries and deviating life stages that challenge AI driven classification and segmentation of images.

areas where boundaries are being pushed by DL algorithms. This ambition to move beyond static classification toward dynamic behavioral analysis is shared by efforts like MAGIK (section 20), which also seeks to uncover spatiotemporal patterns in biological motion.

Practical constraints: In field experiments and ship-based monitoring efforts, plankton communities are sampled at too low temporal and spatial resolution. Often, plankton communities in ocean are dynamic, and short-lived features such as blooms can easily form and disappear between two discrete sampling events. Apart from infrastructural difficulties, there is also a bottleneck in image analysis speeds. With the development of marine profiling instruments with AI based real time classification without manual annotations, the lower marine food web can be sampled at the appropriate spatial and temporal resolution. Furthermore, successful classification is a challenge given the overlapping morphologies in plankton cells (figure 36). While some can be accurately classified, there are other where overlapping morphology will prevent accurate classification even with large training sets.

Usability: Though DL methods are becoming increasingly accessible, most often they are not packaged in a ready-to-use interface. Efforts in this direction would benefit technicians and users with limited programming knowledge.

Advances in science and technology to meet challenges

Below we discuss the recent developments in DL that can be used to tackle some of the challenges in plankton ecology and likely open the doors for new applications.

Labeled data: Obtaining the labeled ground truth data is an undisputed problem in many DL applications. Lately, unsupervised and self-supervised DL algorithms have shown some promising results in this direction. Cycle-GANs, which belong to the family of GANs are widely used in style-transferring the images from one domain to different domain with *unpaired* images and ground truth data. From an image segmentation viewpoint of planktons (figure 37), this indicates that Cycle-GANs can be used for segmentation tasks when there is a limited amount of manually annotated data. Since the plankton images and the corresponding segmentation masks needed not be paired in order to train a Cycle-GAN, synthetically generated masks with comparable morphologies can be used as the ground truth data for real plankton images.

Additionally, microscopy imagery data of planktons can itself be synthetically generate either by GANs [248], or by simulating plankton-like objects. By employing the state-of-the-art computational optics and replicating the optical properties of the experimental devices that are used to record the data, a representative set of microscopy images can be generated on a large scale [94]. The advantage of synthetically generated data is that the ground truth is known beforehand and can be easily controlled. This has the potential to overcome the challenges that arise with manual annotation of data, specifically for the segmentation tasks where careful labeling of cell borders is crucial. As an example, the segmentations of plankton species shown in figure 37 are obtained by a U-Net model trained on simulated plankton

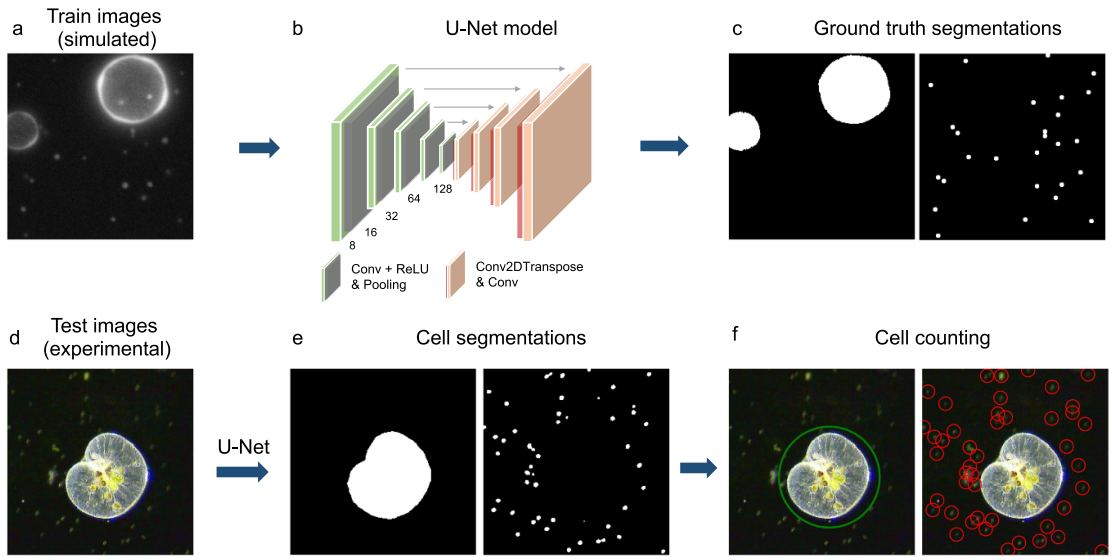

Figure 37. Plankton segmentation and counting using U-Net. (a) Simulated microscopic image containing plankton species, *Noctiluca scintillans* (Larger cell) and *Dunliella tertiolecta* (Smaller cells), are processed through (b) a U-Net model pre-trained on simulated data to obtain binary masks corresponding to *N. scintillans* cells (left) and *D. tertiolecta* cells (right). (d) Experimental test image presented to the trained network. (e) Network outputs: predicted masks of the *N. scintillans* cells (left) and of the *D. tertiolecta* population (right). (f) Post-processing for quantification: identification of *N. scintillans* cell (green outline on the left) and centroids of individual *D. tertiolecta* cells marked with red circles for automated counting (right).

data. This methodology also offers cell classification based on morphological properties, apart from the segmentation.

Recently, self-supervised DL algorithms have also shown some promising results in object segmentation and detection tasks. These advances are conceptually aligned with self-supervised methods explored for microscopy object detection (section 17) and for diffusion characterization (section 19), where minimal manual annotation is a key advantage. Unlike supervised algorithms where learning is based on labeled examples, self-supervised algorithms are provided with labels which are transformed versions of input images themselves. Particularly, ViTs, which employ an attention based mechanism to emphasize different regions of input image, have outperformed CNNs in many CV tasks [210]. Belonging to the family of ViTs, distillation with no labels, a self-supervised ViT [250], was able to successfully segment objects from images without any labeled data with a higher accuracy. Considering the diversity in plankton morphologies, self-supervised ViTs can used for segmentation and classification of complex taxa.

Concluding remarks

AI driven microscopy is a rapid developing field of research. The automatic identification, segmentation, and tracking of individual plankton organisms provides mechanistic insights beyond the current state and will revolutionize our understanding of the lower aquatic food web and allow observations of high numbers of organisms at the relevant spatial and temporal scales. Moreover, recent combinations of well-established techniques such as digital holography with DL algorithms will facilitate individual resolution of plankton organisms and interactions in a way that will catalyze plankton ecology in coming years.

24. Micro-physiological systems (MPS)

Antoni Homs-Corbera

Cherry Biotech, Paris, France
OWL Lifesciences, Paris, France

Status

MPS lying at the interface of microfluidics and biology are a promising tool for better understanding human biology, physiology and physiopathology. Otherwise termed organ on chip devices, these systems exploiting physical and chemical phenomena at the microscale, are able to replicate biological cellular and biomaterial ensembles as well as their conditions in the human body [251]. They provide functional units, emulating organs, that can be interconnected together [252], as is done by the blood and lymphatic streams in the naturally occurring systems, to provide simplified versions of the human body leading to a better understanding of complex multifactorial behaviors [253].

MPS are hybrid devices in which cells and biomaterials status during the experiments play a fundamental role on the accuracy and the relevance of results. Guaranteeing time-dependent spatial conditions and the monitoring of experimental events require gathering, managing and processing complex data in a non-invasive way while preserving the validity of the biological *in vitro* model (phenotype) at least in functional terms. The capability to provide large amounts of time-dependent data is inherent of the MPS which are potentially ready to integrate multiple sensors [254, 255] and to be imaged with sophisticated microscopy techniques [256]. This provides the means of controlling and analyzing biological complexity of several interconnected units, working synergistically, to gain insights into their systemic response.

The usage of ML techniques [254, 257] to control and to evaluate specific behaviors of MPS when disturbed by stimuli, such as drug compounds, could provide a reliable tool to predict the behavior of new potential therapies, to stir therapies discovery, to enable personalized medicine, or to spot processes that remain hidden to the human eye due to the complexity and volume of data that should be simultaneously evaluated. This vision shares strong conceptual links with the autonomous microscopy and modeling approaches seen in sections 10 and 18, which also aim to extract meaningful dynamic patterns under experimental perturbations. This could open a paradigm shift in the pharma and biotech industries with important socioeconomic positive impacts and the substitution, at least partially, of the current *in vivo* animal models [253, 258, 259]. The current parallel advancement of ML, MPS, sensing and imaging technologies provides a fertile field to achieve autonomous MPS platforms providing the necessary high-throughput and robustness to be used as true human physiology and physiopathology emulators. Despite current limitations, the combination of these technologies is already providing encouraging scientific results [257]. An example of how such a system could work is presented in figure 38.

Current and future challenges

Human physiology has a strong dependence on local physicochemical parameters. For long-term studies on MPS accurate control is critical to provide reliable and meaningful data. This control, when aimed massively, particularly for potential applications related to therapeutic screenings, cannot be done by human operators. Automated control systems able to take decisions in real time to fine-tune and maintain MPS parameters, such as dissolved oxygen contain, pH, temperature, fluidic profiles or induced mechanical forces, will be critical (figure 38). Furthermore, they could assess how the system is responding to specific stimulus and also provide an indicator of validity of the MPS model after each experiment based on monitored media components and metabolites as well as other events occurred.

The combination of sensing and imaging technologies with ML approaches provides a promising tool to achieve such a degree of control and evaluation in these complex systems [255–257]. However, sensing and imaging techniques need to be chosen wisely not to alter the MPS biology itself. Chemical and physical conditions that introduce bias (e.g. phototoxicity, genetic alterations, molecule absorption) have to be avoided to keep the models viable and realistic. Some of these undesirable effects can be also eliminated by ML techniques using advances coming from microscopic image analysis such as virtual staining [98] that could avoid the use of chemical staining and eliminate phototoxicity in some applications. These non-invasive approaches have also been explored in cell imaging pipelines (see, for example, section 27, where virtual staining enhances label-free analysis and preserves biological viability).

Nevertheless, acquiring and labeling massive input data to train future autonomous decision-making systems is a tedious and user-intensive task that is subject to certain bias. Also, continuous monitoring of the development and the function of these biological models implies real-time comparison and integration of multiple data sources from heterogeneous conditions which is a challenge by itself. The ML

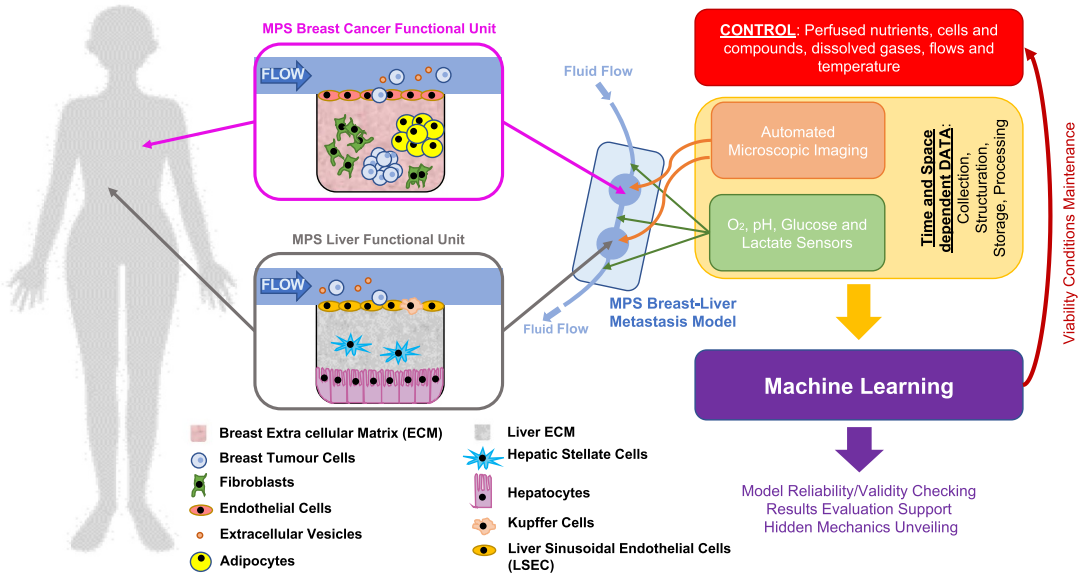

Figure 38. Example of a MPS system incorporating machine learning. This MPS system is focused onto unveiling the mechanisms of breast cancer metastasis into the liver potentially usable to serve as a model to find new therapies or to personalize the treatment of this pathology.

evaluation system should be able to classify time-dependent events expressing in different dimensional scopes. This is another considerable challenge linked to evaluate simultaneously cellular level characteristics, such as phenotypes, migration (e.g. immune and tumoral) or distant communication (e.g. extra cellular vesicles), as well as overall larger multicellular functional units and systemic responses.

The model proposed in figure 38 illustrates the degree of complexity that can be attained in an MPS. Recently, cell-derived vesicles secreted by breast cancer cells have shown to travel and activate liver sinusoidal endothelial cells (LSECs) in a liver MPS promoting the destruction of vessel barriers and unveiling metastatic mechanics [260]. At imaging level, higher resolution microscopy could be also required on top of conventional one to follow the morphology and density of fenestrations, nanostructured apertures present in the liver LSEC.

Multisource automated study of multidimensional time-dependent data is still in its youth for MPS. So far, ML for automation of single-sensor measurements, imaging systems or identification of key candidates in a drug screening procedure have been implemented, but a long road has to be paved to achieve reliable, usable and feasible multiplexed and complex MPS.

Advances in science and technology to meet challenges

ML techniques have already demonstrated their value and feasibility in simple MPS to generate virtual staining, to study cells phenotypes or to track migration events [257] both in supervised and non-supervised approaches. However, the usage of ML techniques to control the multiple parameters of MPS microenvironments to guarantee their viability and phenotype has so far not been demonstrated. Furthermore, limited data sources have been used to evaluate the MPS, such as a unique time-lapsed imaging approach or a limited number of combined analytical sensors, reducing the true potential capabilities of the approach and dismissing important physiological and physio pathological data.

MPS systems improvements in terms of usability, repeatability, high-throughput and robustness have to go hand in hand with imaging, sensing and ML advances. The provision of completely automated MPS systems with these characteristics for continuous, standardized, condition-controlled and reliable data provision for researchers should be the main aim of current developments. This will allow test and optimization of different ML strategies and mechanisms leading to exploit the full potential of these technologies. It will also lead to facilitate high-throughput analysis which is critical to generate the envisaged healthcare disrupting system that are aimed to be. Eventually, other data obtained by sampling the MPS fluids and tissues, such as genetic information, or by medical evaluation of human sources could be added to improve the results when therapeutic evaluation outcomes are targeted.

From the pure ML algorithms point of view, data augmentation, semi-supervised learning and transfer learning can be explored as means to reduce the amount of data needed for the training. Similarly,

automatic data annotation can be used to reduce the cost of manual labeling. Furthermore, specific requirements for DL algorithms applied to the MPS field involve the development of novel customized architecture and layout designs. Different approaches to automate design and gain efficiency have to be explored and catastrophic interference has to be prevented in any explored iterative upgrading. Also, the large amount of real-time data gathered by these systems coupled to sensors and imaging devices implies an accurate study of how to improve inference speeds by compressing the model volume while ensuring accuracy. Equivalently, the hardware utilized for training the DL algorithm and evaluating MPS systems needs to be chosen wisely and performance of the overall system optimized. Finally, interpretable DL technologies, mitigating the 'black box' effect, should be developed to facilitate the generation of meaningful physical explanations from a biological point of view.

Concluding remarks

Combination of MPS and ML techniques is an emerging field holding a great potential to impact human biology understanding, medical therapies development and personalized medicine. Ideally, the application of ML techniques to those multiparametric complex systems should in the future provide currently unattainable biological insights by using unsupervised learning approaches. On the other hand, by comparing results obtained with well-known and characterized therapeutic compounds libraries to the ones resulting with novel compounds and to the real *in vivo* outcomes these algorithms could also achieve good levels of prediction in drug discovery and therapy personalization. Eventually, medical and biological data coming from the human subjects providing tissue or cells in the studies, and obtained by other means, could be also included in the ML models providing extra information in therapeutic discovery and personalized medicine applications.

In the years to come, efforts into automatizing and standardizing the MPS data sources, providing experimental robustness, high-throughput and scalability, should improve the ML outcomes in this domain. The need of condition-controlled, massive and reliable data is paramount. Besides this, the specificities of MPS devices, their data heterogeneity and its multiples sources, have to be considering when designing and exploring strategies for any machine or DL approaches.

Acknowledgments

The author would like to thank the European Commission for continuous research and innovation financial support to Cherry Biotech SAS in the subjects treated in this article through different initiatives and programmes (MSCA-ITN-2018-812780, HORIZON-EIC-2021-PATHFINDEROPEN-01-01-101046928, HORIZON-EIC-2021-PATHFINDEROPEN-01-01-101046620, HORIZON-EIC-2022-TRANSITIONOPEN-01-101099775).

25. Self-learning thermofluidics

Martin Fränzl, Tobias Thalheim and Frank Cichos

Peter Debye Institute for Soft Matter Physics, Leipzig University

Status

In microfluidics, liquids and suspended objects are manipulated in tiny channels to trigger chemical reactions, assemble new materials, analyze single cells at high throughput, or to imitate organs on a chip [261]. Its innovation potential for fields like biology, chemistry, drug design, and medicine is driven by the superior mass and heat transport at these small length scales combined with an easy experimental accessibility and control. Most fluid manipulation approaches are pressure-based and require external pressure differentials applied at the inlet/outlet of complex channel systems. Such methods allow for complex flow designs, but need to move the fluid throughout the channel to affect the local composition, though other local approaches based on electro-kinetic (electrophoretic) actuation exist. In recent years new techniques for the manipulation of nano-objects and flows in a fluidic environment have been developed that are remotely controlled by light and go beyond the force generation of OTs. These approaches employ local temperature gradients to drive the migration of species suspended in liquids by thermophoresis [262], thermo-osmosis [263], or thermo-viscous [264] or other secondary effects [263] and may be summarized as thermofluidics (figure 39). In this way, schemes for manipulating colloids [265] and single molecules [215] have been developed to even provide new access to the study of protein aggregation [266]. Temperature gradients at liquid–solid boundaries in simple fluidic slit pores allow the generation of local flow patterns to guide, manipulate, and separate objects suspended in liquids without any external pressure [263]. Thermo-viscous effects allow dynamic flow generation by dynamic local heating of the liquid. Due to the small dimensions of microfluidic systems, heat-transfer is fast paving the way for feedback-controlled techniques in fluid manipulation [267]. This real-time, closed-loop strategy strongly aligns with similar efforts in automated microscopy and adaptive calibration described in sections 13, 18 and 23. A combination of all these effects would allow to stir, mix, separate, compress, and even heat/cool in microfluidic systems based on the simple application of light. Yet, large scale integration and application in microfluidics is still missing.

Current and future challenges

The increasing complexity in the phase composition of liquids, in the objects suspended, in multimodal sensing techniques used together with the specific goal of microfluidic applications raises the need for new ML-based analysis and control schemes that harvest the advantages of microfluidics [261] (figure 40). One of the current challenges that, for example, is often met in various applications is the appearance of heterogeneities in samples, which hampers the data analysis of molecular species. Such heterogeneous samples are, for example, highly relevant in the study of protein aggregation in assays to understand the origins of neurodegenerative diseases [266].

The experimental investigation of heterogeneous ensembles leads to ensemble averages that are dominated by the most abundant species, which, however, might not be the most important ones for the specific disease. The new local approaches of thermofluidics can be size- or even species-selective in their action allowing the spatial dispersion of the different species. Yet, their interaction with the temperature fields is often unknown. This mirrors challenges faced in dynamic imaging settings such as MPS (section 23) and FLI (section 10), where the measurement and control of local environments must be managed without disrupting system behavior. ML approaches, for example, for real-time visual classification and localization of species in a sample [268] combined with reinforcement learning [269] seem well-suited to meet this challenge of a self-learning variant of thermofluidics. Together with local spectroscopic information in real-time to improve the local homogeneity while increasing the spatial heterogeneity will readily lead to new reconfigurable self-learning tools to tackle chemical, medical, or physical questions with high flexibility.

Advances in science and technology to meet challenges

The main advances to meet these challenges constitute on one side the large-scale integration and testing of thermofluidic approaches into conventional microfluidic systems. This includes simple absorptive layers to induce laser-controlled local temperature increments to yield strong boundary flows that can be combined with additional fields such as electric fields to deliver even stronger effects of thermo-electrohydrodynamics. These approaches shall be studied with pressure-driven flows as they will provide many new variants, such as temperature-driven flow field fractionation. On the other side, the key

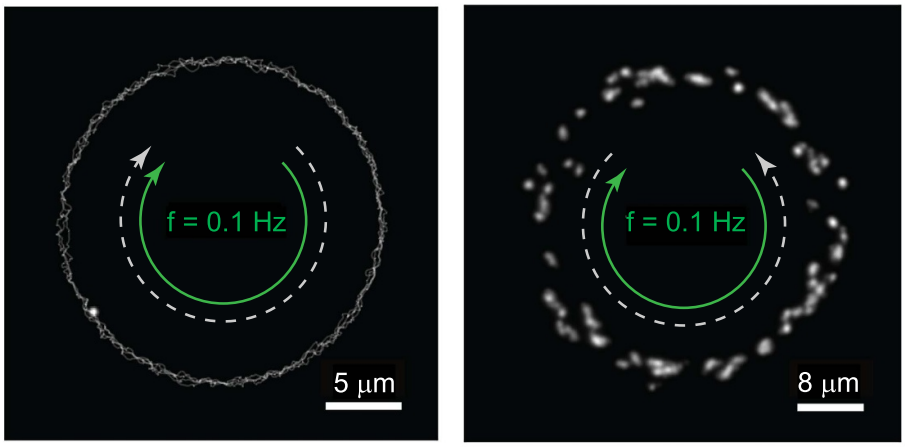

Figure 39. Colloidal motion through thermos-osmotic flows. Colloidal motion induced by controlled thermo-osmotic flows (left) and a combination of thermo-viscous and thermo-osmotic flows (right). Reproduced from [263]. CC BY 4.0.

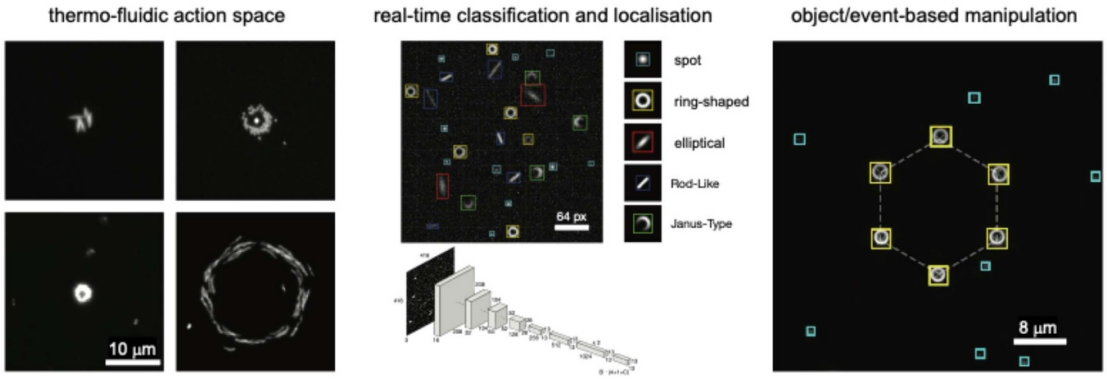

Figure 40. Machine learning-based thermo-fluidic control. Machine learning-based thermo-fluidic control of species using an action space that is determined by thermal-gradient-induced drifts of objects and liquids, a classification and localization of objects in space as well as an object and event-based manipulation that is adaptive due to a learning process. Reproduced from [263]. CC BY 4.0. Reproduced from [268]. CC BY 4.0.

approach that is suggested here is, however, the use of the freely configurable dynamic temperature fields with newly developed ML approaches (recurrent networks, CNNs, reinforcement learning) that pick up local signals with chemical resolution at a high sensitivity and speed. The ML techniques make use of the local character of the temperature perturbations to yield a goal-driven microfluidics. This further requires highly sensitive detection techniques, that deliver at best chemical resolution in real-time to allow the adaptive improvement suggested by ML such as deep reinforcement learning. Such highly sensitive experimental techniques may, for example, involve new variants of photothermal infrared microscopy, which recently revolutionized infrared microscopy but still have to be adopted for microfluidic applications.

Concluding remarks

Local thermofluidic effects have the potential to drive new approaches of ML-driven microfluidics as they provide a rich action set that can be applied to fluids of almost any type to manipulate suspended objects, but also to induce local flows that separate or mix different constituents. While the individual interactions of species with temperature fields might not be known in detail, ML of actions to yield specific outcomes directly during the experiment, will provide new approaches to manipulate species for applications in physics, chemistry, biology, and medicine.

Acknowledgments

We acknowledge support from the SFB TRR102 'Polymers under multiple constraints: restricted and controlled molecular order and mobility' (DFG Project Number 189853844).

26. Digital pathology

Carolina Wählby

Department IT and SciLifeLab, Uppsala University, Uppsala, Sweden

Status

Microscopy-aided visual assessment of tissue samples, such as biopsies and cytological smears, has been essential for diagnosing cancer and other disease for more than 100 years. Grading disease severity is critical to the clinical management of patients. It is however a difficult task, and the variability between pathologists poses clinical challenges, leading to both under- and overtreatment, which impacts patient morbidity, mortality, and healthcare costs. The first approaches to machine-aided diagnostics in pathology were made already in the 1930s, and more recently, efficient slide scanners and automated sample handling has boosted digital pathology and the development of DL-based image analysis systems for assisting pathologists in sample evaluation [270]. Such systems have the potential to increase both accuracy and cost efficiency in cancer screening, and are a prospective solution to the problem of high inter-pathologist variability [271]. Some of the most impressive results have been reached by global challenges such as the PANDA challenge [272], including training- and test-data from multiple hospitals and countries, leading to DL-based solutions based on ensembles of diverse models, featuring, for example, different data preprocessing approaches and different NN architectures. Similar large-scale efforts to develop generalized models across diverse biological samples are also highlighted in sections 21 and 27, in the context of single-particle dynamics and label-free virtual staining.

At the same time, automated systems based on DL are often very sensitive to sample-to-sample variation and artifacts stemming from processes during sample collection, sample handling, staining, and scanning. While a human is very efficient in adapting to such variability, very subtle variations, sometimes not even possible to notice by the human eye, can have catastrophic effects on automated detection and grading. Large multi-site efforts such as the PANDA challenge can overcome these issues simply by ensuring that the massive amounts of data included during training cover as much as possible of the sample variation that can be expected during model deployment in the clinic. However, such large-scale efforts are costly, and may be difficult to organize for rare disease where sample availability is limited. Another bottleneck is the reliability of the training data, depending heavily on the inter-pathologist variability in visual annotation of samples. With robust approaches for image normalization, augmentation, and novel learning-regimen that ignore non-essential variability, DL has the potential to become a broadly applicable and reliable tool in the clinic.

Current and future challenges

The two largest challenges in deployment of DL for digital pathology is the availability of reliable training data, and sample-to-sample variability.

Another word for training data is 'ground truth', an expression that comes from remote sensing, where data is collected from imaging devices attached to satellites or aircrafts, and automated analysis results, such as mapping of roads or classification of tree species, are compared to 'the truth' collected from observations made on the ground. Such observations of 'the truth' are not straight forward in digital pathology. Typically, the 'ground truth' is manual visual annotations by pathologist. Many times, people ask the question 'What precision does a learning-based decision system have to reach to be useful?'. This number must always be answered in relation to the 'truth' to which it is compared, and since pathologists often disagree in their visual assessments, both training and evaluation has to be done with care. One approach is to compare automated result from a learning-based decision system to multiple manual annotations, and in the same way also compare each of the manual annotators to one another [273]. One can also use other metrics as a means of evaluating method's performance. Patient survival is such a metric. This is however a very noisy metric, especially since a mm-sized tissue sample may not at all be representative of the cause of death of a patient. This challenge of defining meaningful 'ground truth' echoes difficulties in subcellular dynamics analysis (section 21), where ambiguity and variability in manual labels similarly limit model reliability.

Sample-to-sample variability and limited generalization performance is a fundamental problem when using DL applied to digital pathology, and lack of generalization may even introduce bias. For example, digitized tissue samples may be collected from a number of different hospitals. If one of these hospitals is specialized in, e.g. a severe type of breast cancer, there is a risk that the system learns to associate irrelevant hospital-specific image effects with sever breast cancer, rather than learning the actual features of the tissue morphology. This is typically approached by some form of *unification of data* from

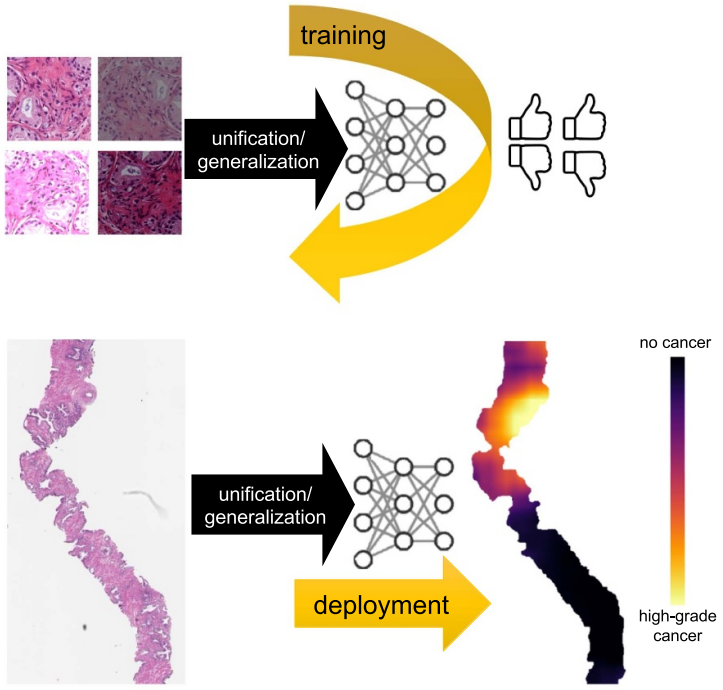

Figure 41. The need of deep learning in digital pathology. In digital pathology, sample-to-sample variation, as well as ambiguities in annotations, call for methods that can handle unification of samples and generalization of deep learning models, at training, inference, or both.

different sites to each other, for example by normalization [274], or by *generalization of the model* during training by creating artificial data so that samples from different sites span the same parameter space, see figure 41. The simplest approach to stain normalization is to separate the RGB-image into its stain components and scale each channel to a fixed intensity interval. More advanced methods, such as sparse non-negative matrix factorization have been successfully used to normalize individual stains [275].

Advances in science and technology to meet challenges

When training a DL model, the input data influences the model's ability to learn relevant features of the data and generalize to new data. A standard technique is to use color and texture augmentation of the training data, artificially generating more variations for the network to learn. However, it is typically difficult to produce a dataset without some bias toward any specific feature. DL models used in digital pathology have a tendency to overfit to the stain appearance of the training data. If a model is trained on data from one lab only, it will usually not be able to generalize to data from other labs.

Recent advances in GANs can reduce the effects of sample variation by being trained to mimic what an observed image would have looked if it was captured in a different batch or at a different site [276]. This type of NN-based sample unification can transfer images from one 'mode of variation' to another while preserving the phenotype of the tissue morphology. In this way, the training data can be extensively expanded to represent a feature space spanning both the non-important variation due to sample handling, and the disease-related variation that is to be learned. It is however important to note that false structures that could influence grading may be added.

Another promising approach is to use so-called domain-adversarial NNs, which are designed to prevent the model from being biased towards features that in reality are irrelevant, such as the origin of an image. Ultimately, such a system would adapt in a similar fashion as a human, resulting in no need for normalization or augmentation, as indicated by promising results on prostate cancer grading on datasets from different hospitals [277].

New molecular methods have the potential to bring DL in digital pathology beyond mimicking ambiguous manual annotations. Recently, new transcriptome-wide analysis technologies have enabled the study of RNA molecules directly in tissue samples, thus maintaining spatial resolution and complementing histological information with molecular information important for the understanding of many biological processes and potentially relevant for the clinical management of cancer patients [278]. Parallel application of standard clinical staining techniques and novel molecular methods introduces a novel type

independent sample annotation that can be used for model training, with the potential to discover of previously unseen but medically relevant tissue patterns.

Concluding remarks

Limited generalization across diverse multinational cohorts is one of the central barriers to implementation of DL in clinical practice. Strict protocols and quality control, all the way from sample collection to staining and scanning has the potential to reduce variability. At the same time, DL approaches that can adapt to shifted domains much like a human, learning to discriminate between image features relevant or irrelevant for decision making, are starting to make their way into digital pathology. Another recent approach to data-efficient models that can generalize and transfer to a wide range of diagnostically challenging digital pathology tasks is the use of foundation models, such as UNI [279], pretrained on more than 100 million images from over 100 000 diagnostic H&E-stained whole slide images.

As spatially resolved novel multiplexed and target-specific molecular methods can be directly correlated with prognosis and strategies for treatment, they may function as useful tools by themselves. They may however also function as a means to molecularly 'annotate' parallel slices of tissue samples exposed to standard clinical stains, thus providing input for DL systems that have the potential to go beyond mimicking what a human observer could do. This principle of using spatially resolved, multiplexed data to annotate microscopy inputs aligns with recent work in MPS (section 23), where multimodal fusion is becoming crucial for biological insight.

Despite many challenges still remaining before DL will be widely used in clinical practice, the attitudes are generally positive: a survey with 487 pathologists practicing in 54 countries [280] showed that nearly 75% reported interest or excitement in AI as a diagnostic tool in pathology.

Acknowledgments

This research was funded by the European Research Council via *ERC Consolidator Grant CoG 682810 to C.W. and the SciLifeLab Bioimage Informatics Facility.*

27. Virtual staining of histological tissue sections

Kevin de Haan, Yair Rivenson and Aydogan Ozcan

Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, United States of America

Status

The histochemical analysis of tissue samples is used as the gold standard for diagnosing diseases. It is a 100-year-old practice based on microscopic analysis of 5- to 10-micron-thin tissue sections stained to provide contrast optimized for the human visual system. While the most commonly used stain is H&E, a wide variety of stains are used to give different contrast to different tissue constituents. This staining process can be time-consuming and expensive and can create a significant amount of chemical waste, some of which is toxic. The staining process is also destructive to the specimen, so typically only a single stain can be performed on each tissue section. Two different DL-based methods have been recently developed to computationally generate an accurate artificial/virtual stain: (1) virtually staining label-free tissue sections, and (2) transforming one stain into another.

Virtual staining of an unlabeled tissue section can be used to eliminate the need for chemical staining altogether. It involves using a DNN to computationally transform images of label-free tissue into various stains (figure 42(i)). This technique has been developed using different imaging modalities as the input to the NN, with many different stains being developed. For example, autofluorescence microscopy [140], QPI [281], and reflectance confocal microscopy [282] have all been used to image the label-free tissue and achieve virtual staining. Furthermore, this label-free strategy aligns closely with the goals described in digital pathology (section 25), where reducing chemical variation and standardizing stain appearance is essential for clinical applicability.

Stain transformations—the second class of methods that can be used as an alternative to histochemical staining—use DNNs to perform transformations from one chemically labeled stain to another [283, 284] (figure 42(ii)). While these methods rely on images of chemically stained tissue as input to the NNs, they can be used to avoid the need for as many stains to be performed, without changing existing clinical pathology workflows. Furthermore, H&E, which many of the transformations are based on, can be performed consistently and is already used in nearly every clinical case. Therefore, the virtual stain transformation technique can be targeted to replace more difficult and costly special stains with their virtually stained counterparts.

Both the label-free virtual staining and stain transformation techniques have been proven to create highly accurate computationally generated stains, which are equivalent in quality to histochemical stains, and allow for accurate diagnoses to be performed [140, 283, 285]. In addition to avoiding the need for chemical labeling (reducing costs and eliminating chemical waste), these techniques have several advantages over standard histochemical staining. For example, virtual stains are standardized, as the same transformation is performed by the NN every time. This results in stain-to-stain variations being minimized. Another major benefit is that multiple stains can be performed on a single tissue cross-section, allowing pathologists to view the same area (and therefore the exact same cells) with multiple stains rather than relying upon staining of serial tissue sections. Furthermore, when performing virtual staining of an unlabeled tissue, the tissue is also preserved for future use if more advanced (e.g. molecular) analysis is needed.

Current and future challenges

One of the challenges posed during the creation of virtual staining models is the generation of the image data required to train the NNs. These data often need to go through an extensive pre-processing workflow, which can be time-consuming, particularly if manual steps such as data cleaning are necessary. Furthermore, a significant amount of data from various sources/labs is needed to allow the models to generalize to new patients or different sample processing procedures. This is a particular pain point for stain transformations, as there can be significant variations between the stains performed at different labs or even within a single lab. For these models to be useful, they must be able to generalize to any given stain (that is correctly performed). The challenges of generalization across heterogeneous data sources are also echoed in virtual microscopy workflows for plankton imaging (section 22) and multi-site diagnostic grading systems (section 25).

I2I translation techniques such as virtual staining are often performed by DNNs trained using supervised learning, taking advantage of loss functions that directly teach the network to perform a mapping between pixels. These structural loss functions ensure that an accurate transformation is learned.

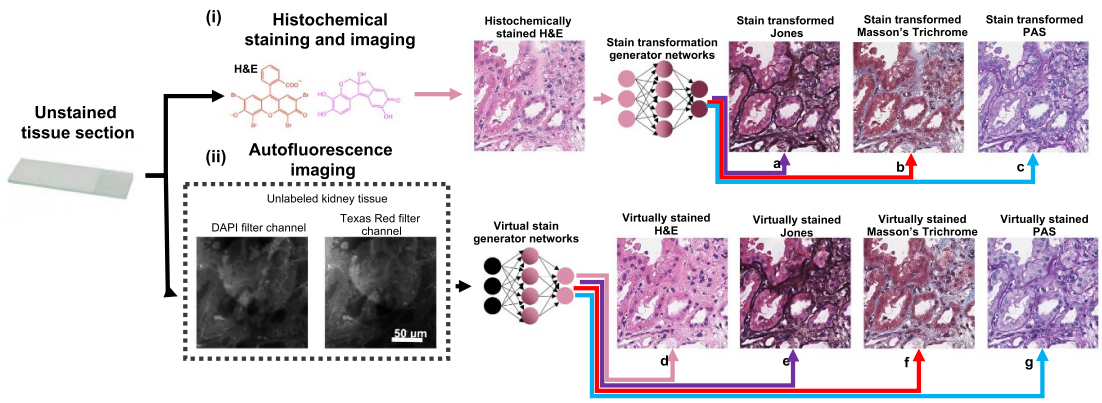

Figure 42. Example virtual staining workflows. Example workflows that can be used to computationally generate stained tissue images by (i) digitally transforming histochemically stained H&E tissue images into various other stains (here, (a) Jones silver stain, (b) Masson's trichrome, and (c) PAS), or (ii) virtually staining the tissue, transforming autofluorescence images of the unlabeled tissue into the various stains (here, (d) H&E, (e) Jones silver stain, (f) Masson's trichrome, and (g) PAS).

However, supervised learning techniques rely upon images of the same tissue section captured before and after the staining being matched at the pixel level. This image registration task can be difficult, particularly when the tissue is damaged or structurally altered during the staining process. Furthermore, for the stain transformation technique, the input to the NN is a stained tissue image. This further complicates the image registration process, as destaining and restaining of tissue is very difficult. While there has been some success with using unsupervised learning (e.g. with distribution matching losses) to perform transformations between stains [284], they may be limited as networks trained only using distribution matching losses are prone to potential hallucinations [286].

Advances in science and technology to meet challenges

These limitations have left room for technological advancements to improve upon previously developed virtual staining techniques. For example, techniques such as data augmentation through stain style transfer are effective at allowing stain transformation networks to generalize across a large sample distribution [283]. By developing and incorporating data augmentation techniques for other modalities such as fluorescence, virtual staining techniques may become more effective and generalizable. This push toward multimodal data fusion and augmentation resonates with approaches taken in MPS (section 23), where multimodal integration helps mitigate variability and expand applicability.

Significant improvements to the quality of virtual staining and ease of dataset development can be achieved by advancing the multimodal image registration techniques used to match tissue images before and after labeling. Synthetic data can be used to avoid the image matching process, and have effectively been used to train stain transformation networks in a supervised manner [283]. More advanced loss functions can also be developed to make the networks less reliant on perfect image registration. Finally, by modifying the chemical labeling process (i.e. the ground truth generation step), the stains may be optimized to reduce the physical damage to the tissue, making any image registration pipeline more accurate and easier to perform.

There is also room to expand upon the unique opportunities afforded by computationally generated stains rather than using histochemical staining. For example, one transformative aspect of virtual staining is the ability to multiplex stains by performing multiple stains all from a single scan (as shown in figure 42, where four different stains are generated from autofluorescence images of a single tissue section). When a single network is trained to perform multiple stains, the stains can also be blended to digitally generate new artificial stains, giving different levels of contrast to different tissue constituents [287]. Such multiplexing of virtual transformations is reminiscent of cross-parameter mappings explored in the MAGIK framework (section 20), where multiple dynamic models are predicted from shared data streams. Multiplexed staining also enables micro-structured staining, where different tissue areas in a single image of a tissue sample appear as different stains [287]. In contrast, histochemical staining is a destructive process, so each tissue section can only be stained a single time and by a single type of stain. Researchers may exploit these and other unique aspects of computational stains enabled by DL to improve the speed and accuracy of diagnoses, with further research being able to determine the best opportunities for virtual staining to improve existing clinical workflows.

There is also significant room for this technology to advance by taking advantage of the symbiosis between virtual staining and image analysis algorithms. Some preliminary studies have begun to show this potential, for example, by demonstrating that the information extracted by a stain transformation network can be used to improve image segmentation algorithms [288]. This same technology can be adapted to improve other downstream image analyzes, such as automated diagnostic algorithms, by incorporating virtual staining into ML-based disease detection workflows.

Concluding remarks

The use of virtual staining of label-free tissue, and stain transformations between histochemical stains is a rapidly developing field with significant potential to improve the field of pathology. These technologies can improve the speed of diagnoses while reducing costs and chemical waste. While this is still an emerging field of research, with continued advancements and regulatory approvals, there are many opportunities for it to assist both human and computer-based diagnoses. Furthermore, as the field of pathology continues to move away from manual inspection of glass slides and toward digitization, there will be more and more opportunities for alternative imaging modalities and digital visualization technologies to make their impact and improve upon standard pathology workflows.

Acknowledgments

The authors acknowledge the support of NSF Biophotonics Program.

28. Cell phenotype determination

Zofia Korczak and Caroline Beck Adiels

Department of Physics, University of Gothenburg, Sweden

Status

Identifying cell types and following their dynamic behavior is crucial when delineating biological mechanisms. Whether *in vitro* or *in vivo*, cell–environment interactions result in stimuli-dependent cell responses. These can constitute, e.g., stem cell differentiation during embryogenesis, cellular defense responses due to altered extracellular environment, or malfunctioning behaviors due to pathological circumstances. Typically, intracellular gene and protein expression profiles are obtained only after cell isolation, whereby the spatial cell–environment interaction is lost. Instead, phenotypic characteristics can rely on visual traits or cell morphology captured by imaging. The classical methodology uses a probe to target the cell, protein, or even particle of interest, which is detected and visualized by microscopy. The probe can constitute a dye that binds to a specific protein or subcellular organelle, a fusion protein expressed in genetically modified cells, or another type of particle or substance that provides information about the cell's phenotype or behavior. Any structural or behavioral cell differences are then analyzed via image processing, clustered, and commonly confirmed with additional biochemical analyses. The temporal and spatial resolution can be modulated depending on the imaging setup and the probe used. Some probes can be applied to and are compatible with living cells with minimal impact during a limited period. Others are harmful and will affect the cells and possibly the response, therefore are better used on terminated cell cultures or 'dead' tissue. Regardless, the specificity and sensitivity of the methodology rely on many different parameters, some hard to control. Further, analyzing the microscopy image outcome of sometimes subtle changes or variations is far from straightforward that may be misleading, especially if using rigid, more traditional image analysis tools, not to mention the risk of being biased.

With the advancement of imaging techniques and AI approaches, a new era of biological screening has emerged. DL mimics humans' learning process and includes statistics and predictive modeling. This growing trend toward label-free classification and analysis strongly complements the virtual staining techniques discussed in section 26, which similarly aim to reduce invasiveness while maximizing phenotypic insight. The technology has been successfully applied in a wide range of biological contexts, including clinical diagnostics (reviewed in [289]), and the quantification of dynamic information from single cells and their subcellular components (reviewed in [290]). It is also central in developing probe-free cell identification and tracking analysis tools. The raw input data can stem from various microscopy modalities: bright-field [291], fluorescence [98], structured illumination [292], scattering [92], or phase contrast [293] (see figure 43), to mention a few. The abovementioned models are trained using supervised learning, which is suitable when the input and the corresponding output are known. The network develops a strategy to transform an input image into the correct output from labeled training data. Another possibility is to train models using unsupervised learning. Unsupervised learning aims to find hidden patterns from the unlabeled dataset. It carries significant uncertainty about the features the model considers crucial but is supposed to be less biased than supervised learning models and is foreseen to be the next generation of software models for complex image analysis.

Current and future challenges

Despite the usefulness and many available applications of DL in cellular phenotyping, there are still challenges to overcome to improve the efficacy of the technology in complex biological settings. One aspect regards attaining dynamic ground truth data; the other regards choosing or developing optimal imaging modalities for distinct purposes.

For example, when targeting specific proteins, cell cultures must be fixed prior to the addition of immunofluorescent antibodies. This approach was used to generate ground truth data, which—when combined with a DL-based method—enabled successful prediction of progenitor cells' future differentiation direction paths from bright-field microscopy images [294]. However, this approach makes it impossible to monitor the same sample over time. If the goal is to study dynamic behaviors, live-cell data is required, along with alternative ways to generate ground truth. This can include using dyes compatible with live cells, fusion proteins expressed by the cells themselves, or other non-invasive staining method. Such an approach is not always applicable, e.g. when focusing on unique protein expressions or using cells that are difficult to manipulate genetically.

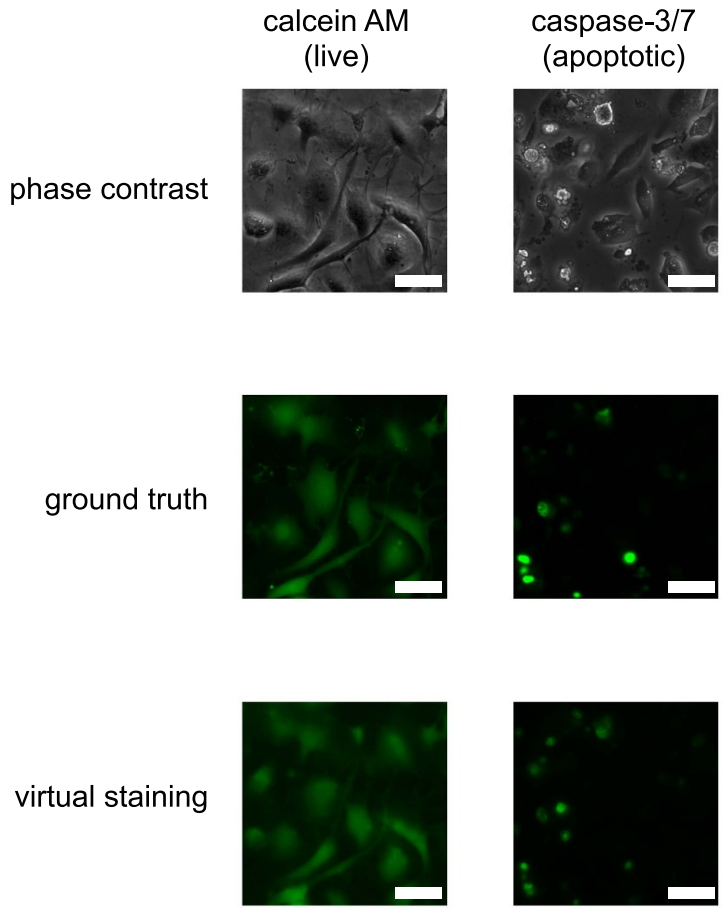

Figure 43. Virtual staining performed on phase-contrast images of endothelial cells. Virtual staining performed using a conditional generative adversarial neural network (cGAN). For each phase-contrast image, there are the corresponding ground-truth (obtained via chemical staining) and virtually-stained images. The left column shows staining for live cells with calcein AM, and the right column, staining with caspase-3/7 to detect apoptotic cells. Virtual staining enables dynamic cell behavioral monitoring on the same cell sample and have great potential for *in vitro* studies using sensitive and, or limited cell samples. The scale bar is 50 μm. Reproduced with permission from [293].

The next hurdle is to image and monitor cells in more complex and physiologically relevant settings, e.g. 3D cell cultures. Such volumetric cultures are often very dense with complex cell–cell interaction patterns, resulting in problematic staining and imaging processes. First, dyes targeting specific cell structures or proteins might not even reach their targets, a common scenario seen in organoid staining where dyes have difficulties penetrating the sample. Second, the dense structure scatters the illumination light, limiting the resolution and focus, which masks the readout. For instance, this is the case when using fluorescent probing in a classical setup of light-sheet-based microscopy, where aberrations, and low-intensity illumination, as a result of scattering, limit the sample thickness range. Similar imaging constraints and penetration challenges are also discussed in the context of MPS in section 23, where deep tissue monitoring without perturbation is critical.

Another relevant challenge is the simultaneous monitoring of dynamic morphological changes on both population and subcellular-scale levels. Such information is relevant for, e.g. understanding the importance of heterogeneous versus synchronized cell responses in cell communication studies or cancer cell screening. In the latter scenario, discriminating cells' behavior based on migration speed and direction, differentiation, and replication frequency over time will be necessary.

Moreover, although new NN model architectures require only a few or even just one image for training, many currently best-working networks require high computational power for training and data analysis. Unfortunately, not all users interested in switching to AI-based software have access to such computational clusters or powerful computers with GPUs.

Advances in science and technology to meet challenges

Meeting these challenges requires progress both in imaging techniques—to provide data rich in information for input into DL models—and on the theoretical side, to develop models capable of accurately

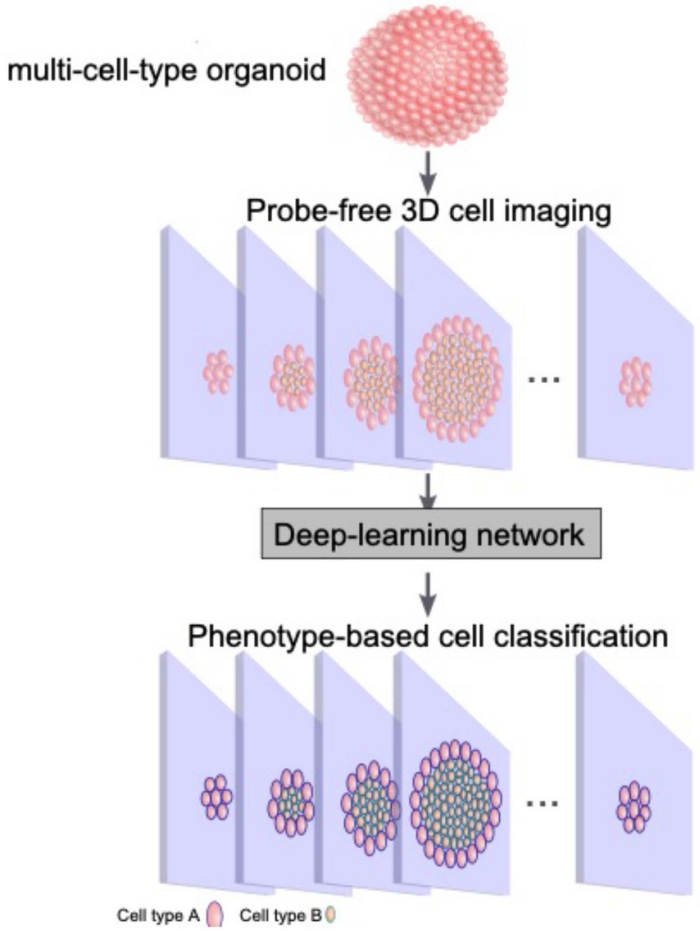

Figure 44. Methodological outline of probe-free cellular phenotype classification using 3D microscopy and deep learning image analysis. An organoid is a 3D *in vitro* cell culture that constitutes a complex, self-assembled cluster of organ-specific stem cell-derived cells (diameter size ranging from approximately tens of micrometers to a few mm). The culture system is a powerful tool in biomedical studies, especially when monitored dynamically and would benefit from live 3D cell imaging and deep learning phenotype identification. Novel imaging modalities are required to provide high temporal and spatial imaging of the dense organoid structure, and the downstream deep learning-assisted image analysis is foreseen to permit extracting quantitative cellular phenotype information.

interpreting the often non-grid structured and interconnected nature of biological information (see figure 44).

Powerful imaging techniques across multiple modalities—such as high-speed light-sheet microscopy for live tissue [295], 3D-structured illumination microscopy for subcellular imaging of thick samples [296], and 3D single-molecule super-resolution microscopy using a tilted light sheet combined with microfluidics [297]—have the potential to overcome current limitations in volumetric imaging, enabling time-resolved monitoring of dynamic cell behavior across scales from large to small. These methodologies could benefit from combining AI with them, such as GNNs, to classify and predict the outcome. Such use of advanced neural architectures for modeling spatially distributed biological phenomena aligns well with recent applications in multimodal motion classification (section 20) and in single-cell dynamics (section 21). Especially in the biomedical field, GNNs have recently gained much attention, due to their ability to, by signal processing, model unstructured and structured relational data (reviewed in [298]).

Recently, a CNN-transformer network, known as a U-net integrated transformer, was recently applied for long-term 4D imaging of zebrafish heartbeat dynamics *in vivo* [299]. It enabled high-resolution imaging across developmental stages with minimal light exposure and acquisition time. The model successfully captured both fine details and global structures, outperforming existing DL methods.

GNNs have also been utilized in an alternative approach where photoactivatable fluorescent probes target annotated cells, and live cell spatial transcriptomics data is attained [300]. The approach provides spatial structure and high-throughput gene expression profiles for individual cells in parallel, successfully clustered by implementing GNNs.

Instead of using expensive and sophisticated microscopy setups for 3D imaging, there is also the possibility of applying different reconstruction approaches that augment, e.g. light-field microscopy images, to circumvent this technology's limitations of non-uniform resolution and slow reconstruction speeds. A so-called view-channel-depth NN [301] yielded artifact-free volumetric image sequences with consistent spatial resolution, allowing real-time reconstruction of biological dynamics.

Regarding the efficacy with which different networks run, improved network architectures will reduce the computational time and, overall, the computational load on local computers. Further, relying on completely unsupervised or semi-supervised ML networks, where a small number of labeled data is mixed with unlabeled data and used for training, will be especially useful when analyzing large data sets that, in turn, will require less computational power.

Concluding remarks

AI for cell biology and biomedicine is a rapidly developing interdisciplinary field that urges tight collaborations across disciplines. The field should focus on developing new tools and providing them in a form that users can modify and retrain the networks for custom image processing with minimal effort. Hence, there is an urgent need for user-friendly software interfaces. Without them, it is very challenging to incorporate AI-aided software in biological laboratories where there are often no programming experts. Reliable, probe-free analysis methods are here to stay and will probably alter many biological laboratories' standard routines. The combination of novel imaging modalities and AI is expected to allow precise 3D single-cell classification as well as time-resolved tracking of complex, environmentally driven multicellular organization.

Acknowledgments

The Swedish Foundation for Strategic Research (ITM17-0384) and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions, DeLiver (No. 766181).

29. Neuroimaging analysis

Mite Mijalkov[1], *Dániel Veréb*[1], *Yu-Wei Chang*[2] *and Joana B Pereira*[1]

[1]Department of Clinical Neuroscience, Division of Neuro, Karolinska Institutet, Stockholm, Sweden
[2]Department of Physics, University of Gothenburg, Gothenburg, Sweden

Status

Neuroimaging techniques such as MRI, CT and positron emission tomography (PET) have been widely used to understand the structure and function of the healthy brain as well as the brains of patients with neurodegenerative and psychiatric disorders. The recent technological advancements in these techniques have led to the collection of larger amounts of higher quality brain imaging data, which has been challenging to analyze using traditional ML approaches due to the complexity of brain anatomy as well as the inherent variability in the images, e.g. due to different scanners or scanning protocols. As a result, it is imperative to create novel, efficient and generalizable methods capable of processing this increased volume of available data [302].

DL models are a promising tool to integrate, assess and make predictions from brain imaging data by using ANNs with multiple layers (figure 45) that extract more meaningful information compared to conventional methods. While their application in clinical practice has not been extensive, there are multiple examples showing their benefits [303], e.g. in the case of assisted reporting, where DL methods are used to pinpoint or quantify pathological changes in medical images. These diagnostic classification strategies share a conceptual foundation with digital pathology (section 25), where DL similarly learns subtle morphological markers to support clinical decisions. In neuroimaging research, DL tools have also been applied to image acquisition and preprocessing by improving the speed of image reconstruction [304], creating higher-resolution images from the originally obtained low-resolution ones or detecting artefacts in the images [302, 305]. Furthermore, one of the most important applications of DL has been in image segmentation, i.e. dividing the image into several regions with comparable properties, allowing for their subsequent quantification [306] (figure 46). Finally, DL has been used in disease diagnosis and prediction, where subjects are classified into different disease groups based on shared behaviors or biomarkers [303].

Given the enormous potential of DL tools and the growing availability of high-quality data, facilitated in part by the creation of many large-scale open-access databases, DL tools may be expected to play a more significant role in future research. Since these tools are not based on any prior assumptions about the data, DL methods can identify novel and unique traits in the data that could provide new knowledge, and therefore allow changing currently established practices, e.g. in studying diseases or categorizing patients.

Current and future challenges

The benefits of using DL methods in data analysis are clear, nevertheless, several limitations hinder their widespread application. Some of these limitations are technical, e.g. the higher dimensional methods needed to analyze the three- or four-dimensional medical images are computationally expensive and computer memory intensive. Furthermore, the lack of understanding of the rigorous mathematical framework and underlying theory by end-users can lead to problems with the interpretability of the models, which could limit their applications and lead to wrong conclusions [302].

One of the most frequently encountered obstacle for DL is obtaining the data needed to test and validate the model [305]. These models need a large number of training data to achieve a good prediction and to prevent overfitting to the training sample. However, especially in the context of supervised learning, producing labeled data is difficult and expensive [303]. Furthermore, the training data needs to be general and representative of the population data that the model is expected to assess. Most implemented models are tested on local datasets, which could have a distribution that is different from real-world data, leading to decreased generalizability and reproducibility of the particular model and difficulties in the comparison of its performance against alternative models. Finally, even if the model's application is limited to a certain imaging modality or disease, there is a variability in the quality of the image, e.g. due to the use of different hardware or imaging protocols, which could again lower the generalizability of the specific model [302, 305].

However, the main issue of DL models is their lack of validation in clinical settings. Model performance has traditionally been evaluated using statistical measures that are frequently obtained from synthetic data. Several visual explanation approaches (e.g. occlusion analysis, gradient-weighted class

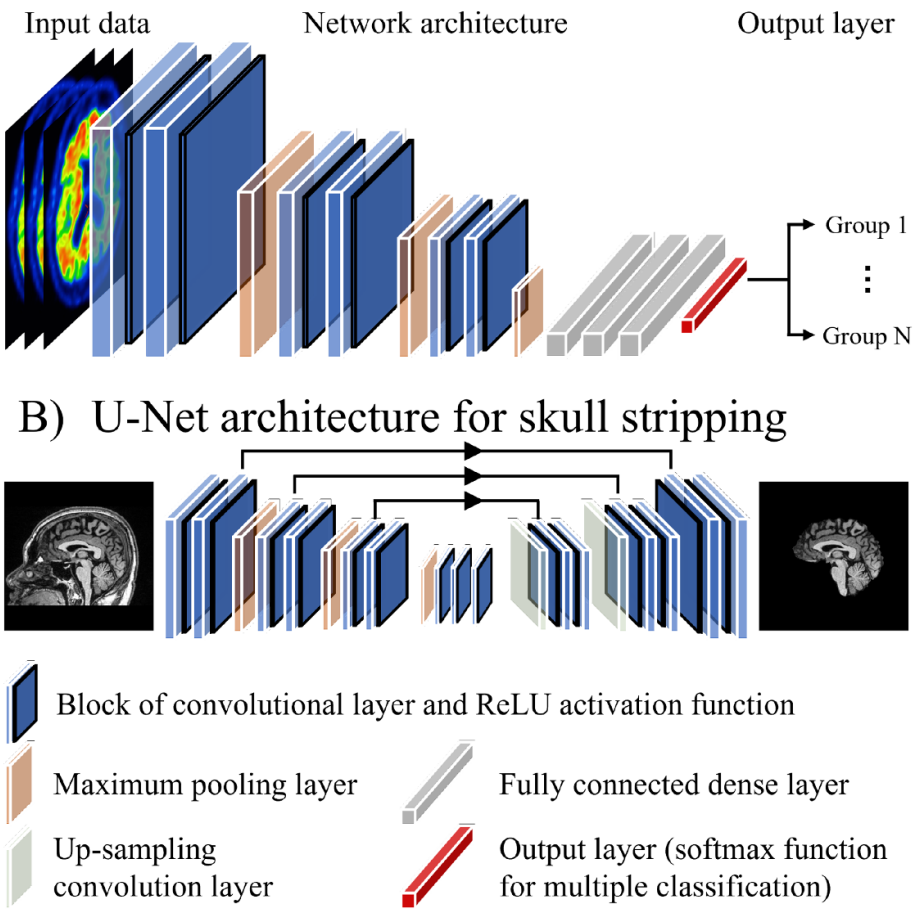

Figure 45. Depiction of commonly used deep learning architectures in neuroimaging studies. Panel (A) demonstrates the structure of a typical convolutional network for accurate binary or multi-class classification using neuroimaging data. Panel (B) represents a U-Net architecture commonly applied in skull-stripping.

activation mapping) have been proposed to obtain the heatmap from DL models, and therefore, further the understanding of the biomarker representation from neuroimaging data via DL. However, in some cases, such metrics can be difficult to interpret by clinical experts and may not be in agreement with experts' predictions, leading to some doubts in the models' predictions. Furthermore, analyzes in research are commonly performed on more standardized, high-quality images or even on synthetic data, which greatly differ from the real-world data acquired from patients in the clinic. This distinction between research-grade and real-world data, and the associated generalization gap, is also emphasized in the context of virtual staining and histopathology transformation (section 26), where sample preparation inconsistencies can derail predictive performance. Hence, this could present an important obstacle for the generalization and practical application of DL models in clinical settings [303, 304].

Advances in science and technology to meet challenges

To overcome the technical challenges, several methods to reduce the dimensionality of the data and, in turn, decrease the computer load have been proposed, e.g. training on only a part of the image, or representing 3D images as a stack of 2D images and use lower dimensionality models to analyze them. The computational time can be further decreased by employing transfer learning [307], a widely used method that entails pre-training the model on a large dataset, recording the obtained weights and finally, applying them to the NN assigned to the current task. Transfer learning and data augmentation strategies are likewise employed in subcellular microscopy tasks (section 21), where labeled datasets are scarce and simulation-based methods are needed to train robust models. The recording of the pre-trained weights allows researchers to train only a subset of layers for the current task, which requires less data; also performing the pre-training on a large data set can lead to more robust results. Another

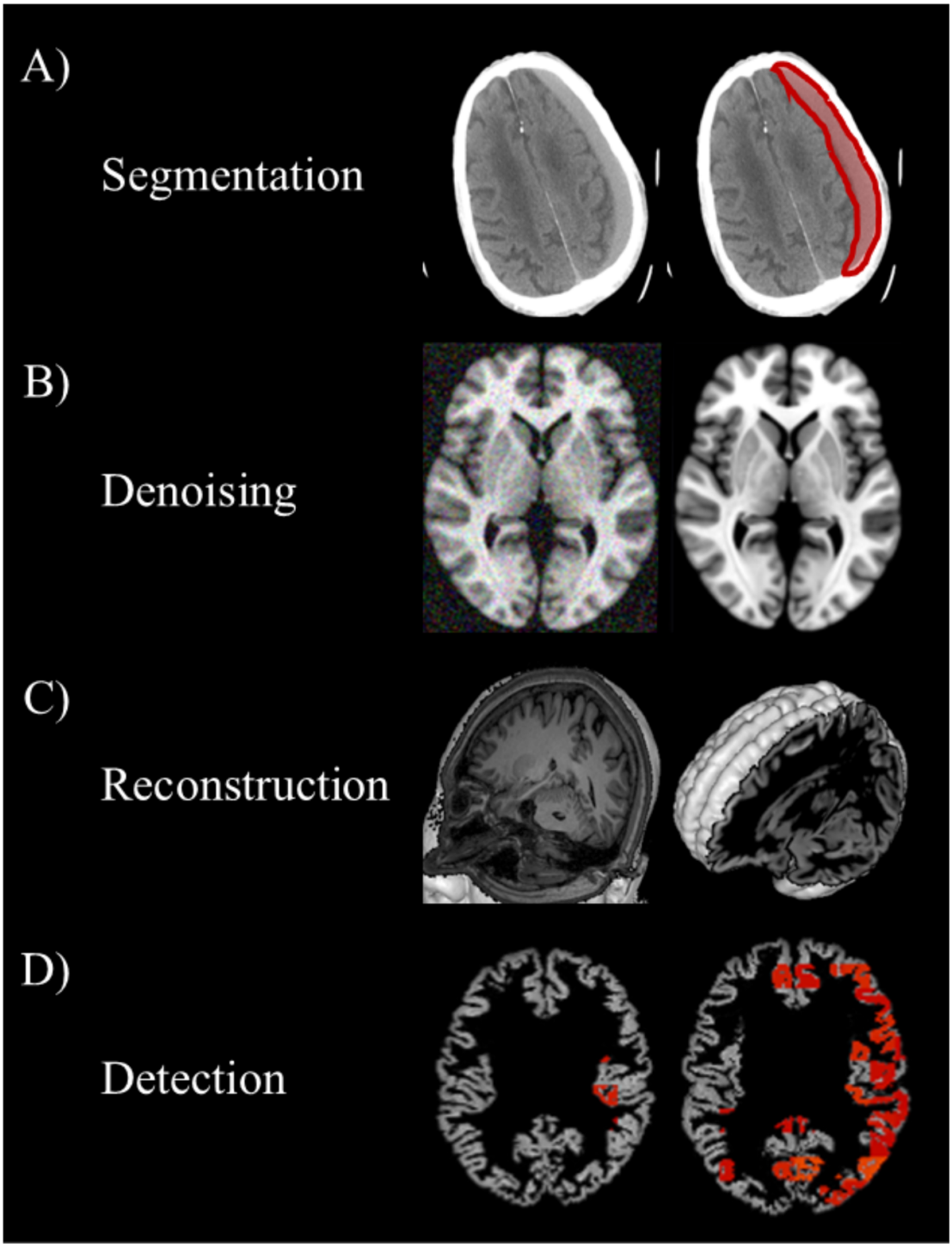

Figure 46. Examples of deep learning applications in neuroimaging pipelines. Deep learning has been utilized in neuroimaging studies extensively due to its ability to detect abstract and complex patterns. This figure provides several common examples of applications, the most common of which are (A) segmentation, (B) denoising, (C) image or 3D reconstruction, and (D) detection of abnormal intensity patterns. Segmentation and detection examples are from the authors' previous work and depict the automatic segmentation of a subdural hematoma in a CT scan and the identification of abnormal tau deposition patterns in a PET scan.

benefit of transfer learning is the ability to share the well-trained DL models between researchers, which can be an important step to improve the generalizability and robustness of these models.

Another broadly used technique to increase the training set and prevent overfitting is training data augmentation [308]. Using this technique, the size of the dataset is effectively increased by introducing some random variations in the data, e.g. random transformation of the image by translation, rotation or deformation, introducing intensity shifts or scaling factors. In addition to augmentation, other techniques have also been developed with the aim of preventing overfitting, including dropout (i.e. randomly

removing nodes from the network at different layers during training) as well as regularization of the network that ascribes weight penalties to different nodes.

To increase the level of use of DL methods in clinical settings, it would be necessary to include the end-user in the process of design and testing the model, enabling the user's understanding of the model and allowing an easier interpretation of the results. Steps have already been taken in that direction, including attempts to integrate the different DL frameworks into the commonly used analysis pipelines and the scanner itself [304]. If successfully integrated, such approaches would allow for multi-site testing and validation for a given method, improving its generalizability.

Finally, there are several ongoing studies aimed at obtaining and providing large, publicly available data sets, e.g. UK Biobank [309], Human Connectome Project [310] or OASIS-3 [311]. The availability of these data sets is extremely important for the design of novel DL methods as they can provide a single setting in which different methods could be compared and benchmarked against each other.

Concluding remarks

Despite the number of challenges that need to be overcome by DL algorithms so that they can be more widely applied to neuroimaging analyzes both in research and clinical practice, there is already overwhelming evidence showing their potential in producing very valuable results. With the current advancement in image acquisition technology, increase in the knowledge of the theoretical underpinnings of the models and understanding of the environment where they would be applied, it is feasible to expect that the current limitations will be addressed in the near future. This will allow researchers and clinicians to agree on the exact protocols of applying these techniques to clinical data, e.g. whether the DL algorithms should be applied as a sole method to tackle a given analysis or if they should be combined with alternative methods or human expertise. Therefore, DL applications in neuroimaging are likely to achieve even more impressive results in the next few years.

Acknowledgments

The authors are funded by a Swedish Research Council Grant (#2022-01108); Alzheimer Foundation; Brain Foundation (#2022-014); by European Union—NextGenerationEU and the Romanian Government, under National Recovery and Resilience Plan for Romania, contract no. 760250/28.12.2023, cod PNRR-C9-I8-CF109/31.07.2023, through the Romanian Ministry of Research, Innovation and Digitalization, within Component 9, Investment I8; Strategic Research Area Neuroscience (StratNeuro); KI Consolidator grant, Center for Medical Innovation (CIMED); Konung Gustaf V:s och Drottning Victorias Stiftelse; Foundation for Geriatric Diseases at Karolinska Institutet; Gamla Tjänarinnor; Stohnes Foundation; Lars Hiertas Memorial Foundation.

30. Bio-analytical and diagnostic TEM

Damian Matuszewski[1], *Gustaf Kylberg*[2] *and Ida-Maria Sintorn*[1,2]

[1]Department of Information Technology, Uppsala University, Uppsala, Sweden
[2]VINNOVA Sweden's Innovation Agency, Stockholm, Sweden

Status

TEM offers the only means of directly seeing and analyzing biological nanoparticles and structures at the nano-level. The recent focus on gene therapy and revival of vaccine development has contributed to an increased interest in TEM as an analytical tool of biological nanoparticles in drug development and formulation, in addition to its established use in disease understanding and clinical diagnosis. The advent of automation capabilities in image acquisition in combination with the expectations of DL provide possibilities and promises for making TEM more accessible. An overview of DL used in various EM applications (biology as well as material science) can be found in Treder *et al* [312].

As repeatedly demonstrated in CV applications and many medical and microscopy proof-of-concept studies, DL offers the great possibility to learn relevant information from examples. Thereby, the human's limited capability to instruct the computer on what information to extract can be excluded. It has the general drawback that *a priori* expertise and information not represented in the examples will not be incorporated in the model unless explicitly added. It is also difficult to decipher and validate what information is used to reach a decision. This leads to interpretability and reliability issues which practically limit DL deployment in real-world TEM bio-processing and clinical applications. These applications often face a scenario with a limited amount of training images, lack of or unreliable ground truth, non-representative or too narrow training sets, as well as regulatory requirements. Such limitations also arise in neuroimaging (section 28), where high-dimensional but scarce labeled datasets and clinical validation challenges limit widespread adoption. In addition, for image data in clinical situations, there is also a fear of missing 'other' and unexpected information obvious to the experienced eye.

Commonly used AI networks are very self-confident in their predictions, also when the evidence for a certain decision is dubious. This results in so-called silent failing i.e. misclassifications and unreliability due to too narrow training sets, as well as missing structures in the sample not directly asked for. For certain applications, this is a showstopper. However, in other applications, a model does not need to be extremely accurate or generalize to be useful. They can solve or contribute to a particular step in the analysis pipeline, be interactively fine-tuned by the user for each dataset to be analyzed, or serve as a detection or segmentation step further processed or validated by the user. Thus, in many practical scenarios, tools that reduce rather than remove human input are in demand.

Current and future challenges

Common to all types of microscopy is the need to adapt and modify DL methods to the prerequisites and characteristics of the different imaging techniques. Some general issues that need TEM domain-specific attention are:

(1) **Quality and amount of annotated image data**. The training dataset has to be carefully prepared to represent all variations that can be observed in the application. It is, however, easy to miss rare albeit important instances. In a real setting, the models will not only have to face all varieties of data appearances but also anomalies unlike anything in the training set. Typical supervised models fail when presented with a sample from a different class or appearance than those in the training set. This is exemplified in figure 47. What might be obvious to a human expert requires careful consideration when training and verifying DL models.
(2) **Design of performance and evaluation criteria** that are relevant for the imagery and application. A TEM is a complex scientific instrument used to gain information about nano-scale features. In contrast to many CV applications, looking pleasant is not interesting while being correct is crucial.
(3) **Adaptation of network architectures to focus on the characteristics of TEM**. For example, features at the single-pixel level in combination with large-scale information are usually key for what is studied. The images can in general not be down sampled as is done as a pre-processing step to many standard DL models without severe loss of information.
(4) **Development of TEM-specific augmentation techniques** that incorporate variations in instrument settings and properties, as well as modality-specific artifacts. This is especially important for complex imaging techniques such as TEM since generating real data representing a lot of variation (instrument makes and models, instrument settings, sample preparation, etc) is difficult due to the

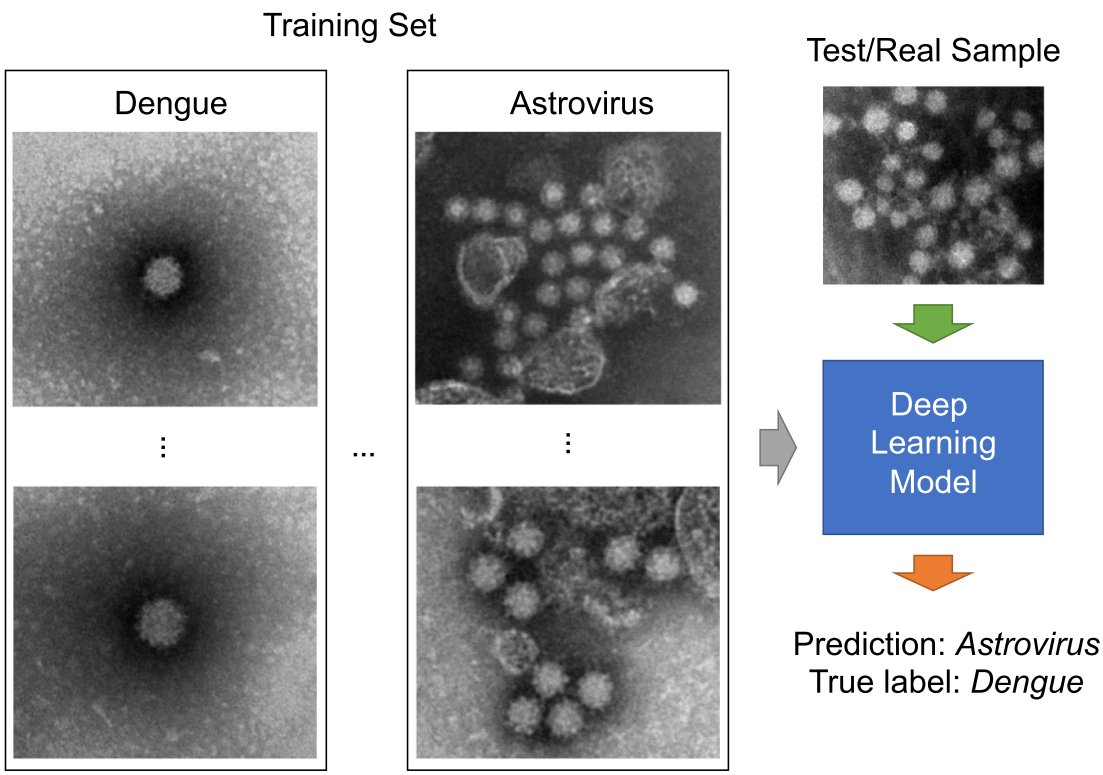

Figure 47. Illustration of the importance of representative training data. Most of the dengue images [315] with more than 2 virus particles were randomly assigned to the test set. They were confused with astrovirus and/or West Nile Fever due to the untrained context. Reproduced from [315]. CC BY 4.0.

low abundance of microscopes and high cost. The value of appropriate augmentation has e.g. been shown in stain normalization in pathology [313]. Comparable efforts to overcome domain shift and limited annotations have been extensively discussed in sections 25 and 26 on digital pathology and virtual staining, respectively.

(5) **Lower the expectations to gain real value**. Shift the focus from trying to apply and 'market' DL as the magic answer to all problems to seeing it as a tool in the toolbox to aid the imaging and analysis process. For example, to automate instrument control as exemplified in [314] to simplify imaging and increase robustness, or to assist the sample analysis and thus reducing the manual input for corrections or verification provided by a human expert.

Advances in science and technology to meet challenges

As described above, many of the challenges are common and general for microscopy and medical imaging techniques but require a specific modality or instrument focus. Therefore, theoretical DL advances, as well as insights gained in other imaging domains and applications, are also of interest to TEM. Incorporating physics and constraints in the DL models and/or loss function is one such technological advancement that has proven useful in other domains [316], but is yet to be widely explored in TEM instrument control and applications.

Another example is the Human-in-the-loop trend [317], meaning that the user or expert guides the model training, evaluation, or use (sometimes referred to as human-in-the-loop data analytics). This is particularly useful in applications with limited (annotated) data, or when fine-tuning or human validation is required for each new sample/dataset. One example of such a human-in-the-loop approach also showcased on EM data is [318]. The user interactively marks the type of objects to search for in a dataset and then validates suggested objects in iterative steps.

The lack of training data and EM benchmarking datasets is being addressed to some extent via competitions or 'challenges' such as the ISBI 2021 challenge 'large-scale Mitochondria 3D Instance Segmentation from Electron Microscopy Images' and the 2012 challenge 'Segmentation of neuronal structures in EM stacks'. This benchmarking and community-based validation echoes initiatives like the AnDi challenge described in sections 21 and 36. In addition, the recent requirements by journals and conferences to publicly share image data will naturally lead to increased availability of also TEM images.

Recently authors Matuszewski and Sintorn published a virus classification data set and DL benchmark performance [315], focusing on discussing the importance of training and test sets representativeness and its impact on performance and reliability in real-world applications.

In recent years, following a large interest in understanding DL decisions, more attention has been put on designing models producing confidence scores [23], anomaly detectors [319], and interpretable models, all to strengthen DL reliability and trustworthiness. A sign of technology maturation of DL in medical imaging along those lines is to apply DL to smaller interpretable parts (cf as a tool in the toolbox) contributing to the decision support with additional sources of information rather than constructing a DL black box total solution [234].

Concluding remarks

Many of the issues with DL deployment in real situations are common to all imaging modalities, e.g. lack of sufficient amounts and variations of realistic and annotated training data, as well as reliability and interpretability of the results. General advances in DL methodology and making more TEM image data publicly available will of course transfer to and benefit users also in the TEM field. In addition, incorporating electron imaging physics and instrumentation properties into DL models and data augmentation properties is a foreseeable adaptation from other imaging fields.

Perhaps somewhat more specific to TEM, we foresee that DL will play a big role in instrument control and automation and hence make this complex imaging technique more accessible and available. We also believe that theoretical DL development focused on the prerequisites in TEM will lead to better and more efficient architectures for handling the large-scale spanned in TEM imaging and analysis. Often the interplay and correlation of image features at multiple scales including information at the finest pixel level is needed to gain insight, and multiple scales need to be traversed to find objects or regions of interest in the huge search space a sample in TEM constitutes.

Acknowledgments

This work has been supported by a post-doc grant from the Swedish e-science collaboration (eSSENCE) to author DM. Author IS has been supported by Uppsala University's AI4Research initiative and the Swedish Foundation for Strategic Research (Grant BD150008).

31. High-content high-throughput screening

Juan C Caicedo[1] *and Beth A Cimini*[2]

[1]Morgridge Institute for Research, University of Wisconsin–Madison, Madison, WI, United States of America
[2]Broad Institute of MIT and Harvard, Boston Cambridge, MA, United States of America

Status

With modern optical, electronic and chemical technology, imaging experiments can be run efficiently in *high throughput* to address pressing research questions in drug discovery and functional genomics. Given the massive amounts of imaging data generated by such experiments, accurate software platforms are also necessary to transform images into *high content* quantitative data for downstream research. Phenotypic high content imaging assays have been used to determine mechanisms of action of drugs, predict toxicity, and even predict the outcomes of other, non-imaging assays (reviewed in [320]). While DL approaches are not currently universal within high content/high throughput screening (HCS/HTS), work in recent years has illustrated several ways in which DL can add value to these assays (figure 48).

The computational workflow to transform images into biological insights typically begins with *segmentation*—that is, providing an exact boundary for each desired object within the image. This area has progressed the furthest in recent years, thanks to foundation models that generalize to new experiments [321], and successful tools becoming widely available to the community [322].

When the objects of interest—typically single cells—have been identified by segmentation, the next challenge is *feature extraction.* This involves quantifying the unique properties of cell state, cell structure, and phenotype observed in the images using multi-dimensional representations [323]. Recent advances in representation learning may automate data-driven optimization of feature representations, instead of requiring analysts to manually create feature sets.

The quality and content of images themselves can now be enhanced and expanded using *image generation and modification strategies.* Generative models (such as some deployed in [164]) can now improve the resolution and SNR of images, which improves the quantitative data obtained. Similarly, networks can predict the staining pattern of many biological stains from unstained images, allowing one to generate measurements of biological structures that were never directly stained (reviewed in [323]). Generative models can also reduce the effects of technical variation (e.g. well or plate effects) by modeling and then removing them; this can allow the researcher to calculate how an observed image would have looked if it was captured in a different batch [276]. Similar batch correction strategies are also explored in virtual staining workflows (section 26), where domain transfer and harmonization are essential for consistency across staining protocols.

Finally, DL has the potential to change *the scale at which assays must be run*: training models to predict the outcome of one assay based on already-collected data from an orthogonal assay has been shown to enhance secondary screening hit rates up to 200X [324].

Current and future challenges

With recent advances in DL research on bioimage, the problem has shifted from assessing whether DL could help to *choosing which aspects of your workflow benefit most from deep learning.* As DL can potentially improve many aspects of the HTS/HCS process (see above), at each step one must assess which options are available, how much they stand to improve the final data, and what the compute, person, and opportunity costs are to implement them.

Compute considerations are critical in HTS/HCS: assays typically involve dozens or hundreds of plates with thousands of fields of view per plate [320]. At this scale, speed matters: each second of analysis time equals ∼1 CPU hour per plate. While DL techniques can be orders of magnitude faster than conventional approaches [325], these increased speeds often rely on access to GPUs, which are at a premium compared to CPUs in most on-premises and cloud environments.

Person and opportunity costs are harder to predict or measure, since one must investigate options (and therefore partially incur these costs) in order to assess the potential benefit. For tasks where tools already have been made for a 'substantially similar' task, care must be taken to assess how similar the external training data is to one's own data; differences invisible to the human eye can still require retraining. If the external training data is sufficiently large and varied, publicly available models may be appropriate without further training; this assessment is often easier if the developers have released performance statistics on public benchmarks. Even when existing tools require no retraining, many require advanced computational skills to use and/or scale; a 2022 conference poll on 'Barriers to entering HCS'

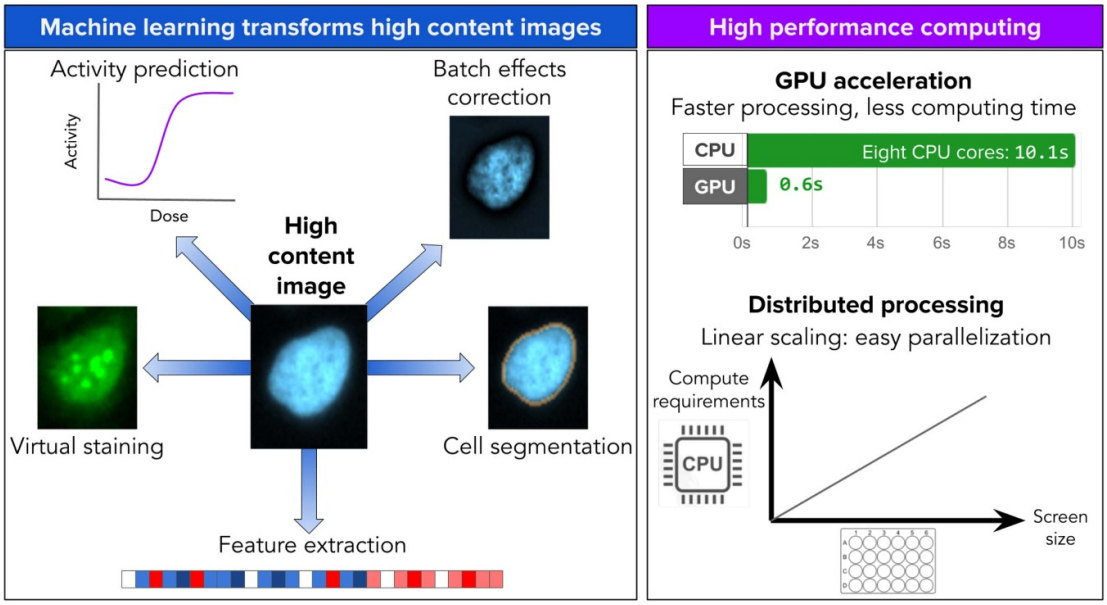

Figure 48. Left panel: five ways deep learning can add value to high content assays. Clockwise from top left: predicting outcomes of other non-imaging assays; removing blur, noise, or batch-induced variation from images; segmenting objects; extracting numerical feature data; and predicting stains not acquired. Methods are not mutually exclusive—any or several may be used on any high content assay. Nuclear images courtesy of the Broad Bioimage Benchmark Collection BBBC022. Right panel: high performance computing is necessary for processing imaging screens. Deep learning models usually benefit from graphical unit processing (GPU) acceleration (top). When processing HCS/HTS data, the processing load can be distributed among multiple machines both during training and deployment of models, resulting in linear scaling of compute requirements as the number of images increases; this is favorable compared to other applications where compute requirements increase exponentially as data size does. (Bottom). Note that GPU acceleration is significantly more expensive than CPU computing, both when purchasing hardware and in the cloud.

identified 'access to knowledge' as the largest barrier and 'Python' as the thing they most wished new employees knew (Cimini, personal communication). This concern over accessibility and tool usability mirrors the call for user-friendly AI interfaces raised in section 27 on cell phenotype determination, particularly for non-programming biological users.

If one decides that no existing tools are appropriate, many more decisions emerge—should one adapt an existing bioimaging tool, or seek out new computational architectures? If using an existing tool, should one train from scratch or fine-tune an existing model? Are there existing biological data sets that you can use for training, or must you generate your own? How should the ground truth be defined, how can it be made efficiently, and what are appropriate metrics of success? These questions are difficult for experts to answer and will take even longer for new users.

Advances in science and technology to meet challenges

The success of DL for natural images is partly explained by the careful curation of datasets with ground truth annotations. Ground truth collection has been slower for microscopy analysis tasks, but the field is making progress for problems such as cell segmentation [321, 322]; more coordinated efforts are necessary to expand the availability of well annotated data for training foundation models under different experimental conditions. Since creation of these sets is costly and time consuming, training strategies that do not require supervision or explicit annotations for learning should be deeply explored. Such semi-supervised and contrastive learning approaches have also gained momentum in SPT (section 21) and neuroimaging (section 28), where labeled data is difficult or costly to obtain. For instance, a family of self-supervised learning techniques, including contrastive learning, use image matching under different transformations (such as cropping or color adjustment) to capture machine-useful (though not human-interpretable) features that make an image unique. These features have shown to be highly informative to recognize objects in natural images and could be used for phenotypic analysis in images of cells or to accelerate training of segmentation and/or classification models by requiring significantly less ground truth to fine-tune.

While more powerful networks are needed, these tools will not gain wide adoption if users find them too challenging to use. Educational materials, tutorials, guides and documentation for non-experts

on how to make decisions (about models, hyperparameters, and more) and use these tools in practice are required to reach the scientists that will ultimately use them to drive biological discovery. This will require tools to create easy to use user interfaces [326] or PnP API libraries that can be loaded in Jupyter Notebooks [164] or similar environments to support the development of quick and reusable processing pipelines.

Finally, once an approach is finalized by the scientist, DL architectures require significant computing resources; while these needs are greatest during training, they may also be significant for processing new datasets in inference mode, which provides an especial challenge for the scale of HTS. New advances in model compression and efficient architectures must be developed and adopted for bioimage analysis. Energy efficient architectures can run on mobile devices and could be useful to accelerate processing where GPUs are available or to allow deployment where they are not. To serve HTS-scale data, tools must also be developed to allow easier parallelization of models across remote servers or cloud computing services [164, 326, 327]. The community also needs guidelines to navigate these computational choices and to make architecture and platform decisions in practice.

Concluding remarks

While it advances rapidly, DL continues to have great potential to revolutionize the entire HCS/HTS process, from assay design to compound selection to segmentation and feature extraction. As new tools are created and/or refined to tackle these and other as-yet-unimagined possibilities, we believe that emphasizing *reusability, interpretability,* and *scalability* will help the entire community. Training on diverse image sets, while more time consuming, leads to reusable tools which require minimal hand configuration [321, 322], and comprehensive documentation increases practical ability to reuse. Tools should maximize interpretability: for some tools, this may involve disclosure of failure modes; for others it may involve adding attention maps so users understand what led the network to a particular conclusion; and for others creation of ground truth benchmarks to better align results across experiments. Finally, reusable and easy ways to generically scale new approaches must be developed to ensure adoption at HTS scale. If such tools are created and made widely available and user-friendly, it is easy to imagine that in a very few years there will be no steps in HCS/HTS that do not routinely incorporate DL models.

Acknowledgments

This work was supported by the Broad Institute Schmidt Fellowship program to JCC, NSF Grant Number 2348683 to JCC, and National Institutes of Health Grant P41 GM135019 to BAC. This project has been made possible in part by Grant Number 2020-225720 to BAC from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation.

32. Ultrasound and photoacoustic image formation

Muyinatu A Lediju Bell

Departments of Electrical & Computer Engineering, Biomedical Engineering, and Computer Science, Johns Hopkins University, Baltimore, MD 21218, United States of America

Status

The success of diagnostic and interventional medical procedures is deeply rooted in the ability of modern imaging systems to deliver clear and interpretable information. One of the most widespread imaging systems available in hospitals and clinics around the world is the ultrasound imaging system. With its 60+ year history, ultrasound imaging has four major benefits in comparison to other medical imaging and microscopy systems: (1) safety, (2) portability, (3) cost-effectiveness, and (4) real-time delivery of images.

While sound is transmitted and received in ultrasound imaging systems, when augmented with lasers and other light sources to create photoacoustic imaging systems, light is transmitted and sound is received. This alternative approach provides optical absorption information, rather than the acoustic reflectivity information that is provided with traditional ultrasound imaging system, while retaining the four major benefits noted above. Ultrasound and photoacoustic images may also be interleaved to improve the overall clinical experience.

In both ultrasound and photoacoustic imaging systems, the sound received by an array of acoustic sensors is typically converted to an interpretable image through the beamforming process, which is often the first line of software defense against poor quality images. However, the beamforming process has historically suffered from recurring limitations that produce poor image quality in a subset of patients. In particular, traditional beamforming procedures rely on assumptions about wave propagation (e.g. sound speed, acoustic pathways) that are not true in the presence of significant inter- and intrapatient variations. For example, a single, direct path from an ultrasound or photoacoustic source to the acoustic receiver is often assumed, but this assumption does not consider the presence of multi-path acoustic scattering or reflections that occur in the presence of acoustic impedance differences.

While DL is impacting ultrasound and photoacoustic image formation through improvements to beamformers, image quality, and diagnostic interpretability [328, 329], no historical advance prior to this impact has combined multiple benefits in a single image formation or signal processing step [330]. As a result, potentially useful information provided by the same raw data is either absent or prolonged, considering that advanced methods may be time consuming to implement and are typically applied in succession, rather than in parallel. Alternatively, a DL approach has the benefit of learning from multiple training examples, rather than relying on flawed assumptions. This benefit opens the door for the delivery of clear, interpretable images that combine multiple benefits in parallel, based on a single input of raw sensor data [330]. An example of this approach is presented in figure 49.

Current and future challenges

Three significant challenges surround the capability of DL to create ultrasound and photoacoustic images that combine multiple image formation and signal processing techniques in a single step. First, ultrasound and photoacoustic sources must be localized and detected in the presence of noise and artifacts. Based on the Hyugens–Fresnel principle [331], this challenge can be reduced to detection of point-like sources, which can represent single scatterers within tissue, needle or catheter tips, photoacoustic signals from an optical fiber tip, or individual microbubbles. When the acoustic response from a point source travels outward from the source to the transducer, the shape of the recorded wavefront is determined by the distance from the source to the acoustic receivers. Thus, sources that are closer to an array of acoustic receivers will have a different recorded wavefront shape than sources that are farther away. This unique shape-to-depth relationship can be learned by DNNs [332, 333]. Advantages include the ability to spatially locate acoustic sources with high precision and accuracy in comparison to images created within the diffraction limits imposed by the beamforming process [333, 334].

The second challenge is accurate segmentation of an imaging target or disease feature of interest. The underlying goal is to produce an image that emphasizes structures of interest for a particular application and deemphasizes surrounding structures. Segmentation is typically performed after image formation, but if the image quality is poor, then the segmentation will also suffer. This is additionally problematic when ultrasound or photoacoustic imaging systems are operated by less experienced users or when automated tasks rely on segmentations to deliver a diagnosis, treatment, or medical assessment. Similar

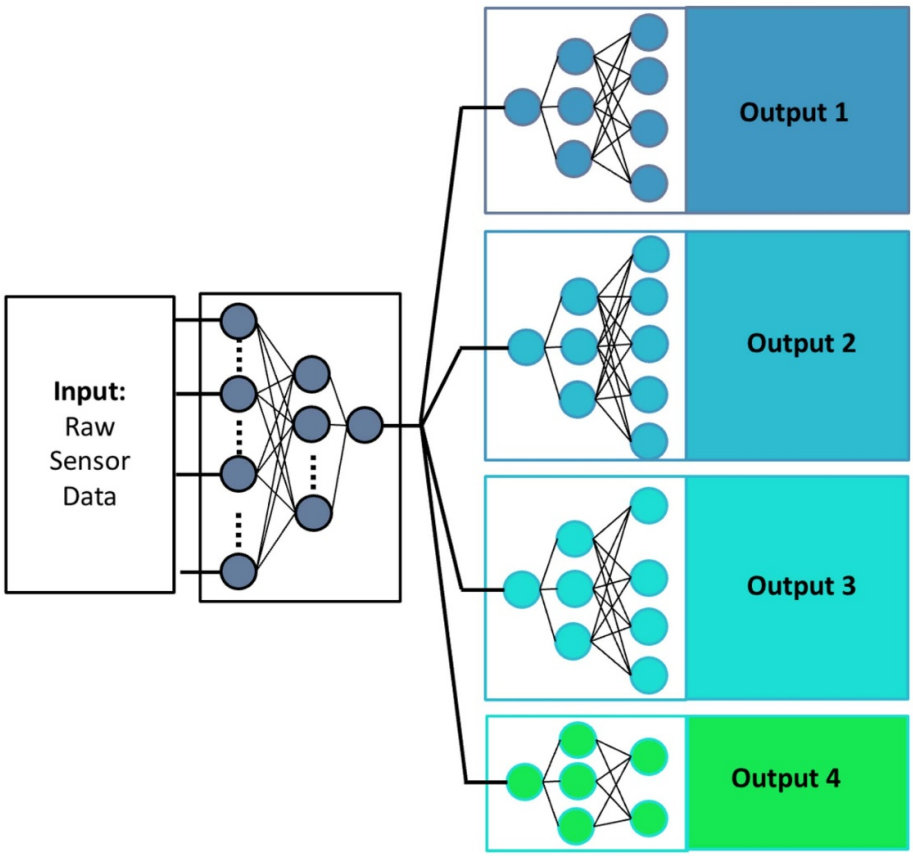

Figure 49. Deep learning enhanced ultrasound or photoacoustic imaging. General concept of supplying a single input of raw ultrasound sensor data, from which multiple outputs are simultaneously produced in parallel with the assistance of deep neural networks (DNNs).

concerns regarding operator variability and automation reliability are echoed in the digital pathology domain (section 25), where inter-observer disagreement can hinder model training and interpretation.

The third challenge is the time required to create images with advanced beamforming methods, such as coherence-based images. For example, the short-lag spatial coherence (SLSC) [335] beamformer successfully reduces acoustic clutter in cases where traditional clutter reduction methods, such as harmonic imaging, fails. SLSC beamforming also has the ability to determine the fluid or solid content of suspicious masses [336]. Despite these benefits, SLSC and similarly advanced beamformers can be computationally intensive to implement, which has hindered integration into existing clinical systems.

Additional challenges that must be overcome to enable widespread future impact include the ability to understand, interpret, and predict expected network outcomes and failure points. This ability will enable systematic development of new DL approaches. In addition, providing the most optimal speed up of advanced beamforming approaches will expand existing potential, and smart integration with robotic approaches will promote future possibilities for fully automated procedures.

Advances in science and technology to meet challenges

DNNs were either applied or created to address the three significant challenges noted above. Point source detection was demonstrated with multiple networks, including AlexNet, Resnet, and VGG-16 [334]. Segmentation and feature detection from raw data was demonstrated with U-Nets [330]. CohereNet [337] was created to calculate coherence functions for the advanced SLSC beamformer. CystNet1 [338, 339] and CystNet2 [339] each consist of one encoder and two decoders, and these networks were built to simultaneously image and segment cysts from raw ultrasound data in parallel, rather than perform the traditional sequential approach. Figure 50 shows the architecture of a selection of these DNNs.

These advances either independently or collectively demonstrate the feasibility of creating multiple outputs from a single input of raw channel data with DNNs. The collective demonstration exists because it is possible to concatenate multiple DNNs in parallel, thus providing an approach to input raw data to each concatenated network and achieve simultaneous outputs from each input. While many of the

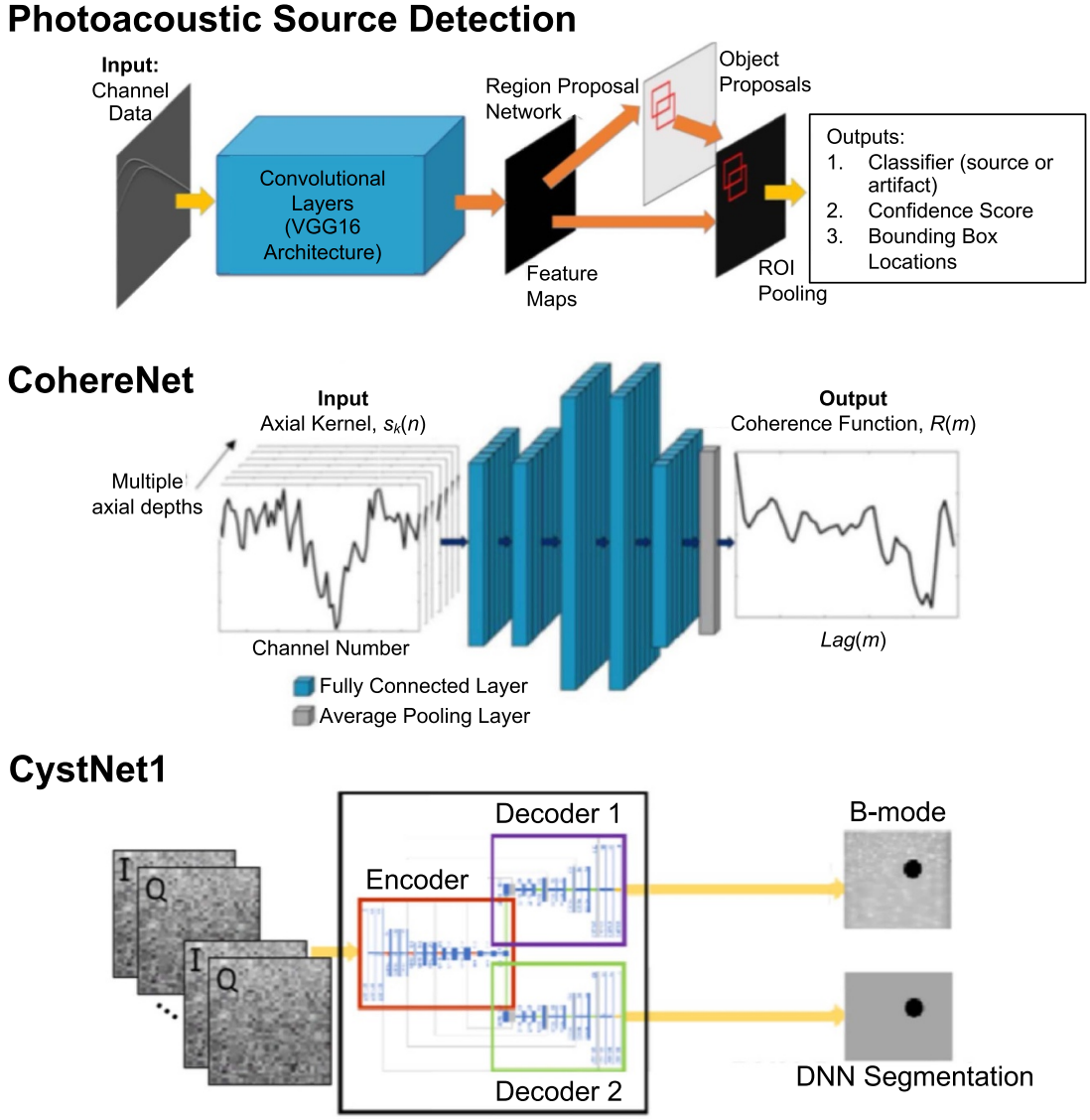

Figure 50. Examples of neural networks applied to photoacoustic and ultrasound imaging. DNN architectures were applied or created to address challenges with photoacoustic source detection, spatial correlation calculations, and segmentation. Each DNN accepts various versions of raw channel data. Reproduced from [333]. CC BY 4.0. Reproduced from [337]. CC BY 4.0. Reproduced with permission from [339].

networks noted above were trained with simulated data that mimicked the physics of wave propagation [334, 338, 339] or *in vivo* breast data which contains variability arising from highly heterogeneous breast tissue [337], it is a critical advance that each DNN operated on sufficiently variable data sources relative to the training data. This generalizability highlights the success of the training process. This physics-informed training paradigm aligns with modeling approaches used in neuroimaging (section 29) and TEM (section 28).

An apparent tradeoff between providing human-interpretable images and integrating advances with robotics has also emerged as a direct outcome of the science and technology implemented to meet existing challenges. For example, a DNN can be created to achieve a specific task for robotic integration, such as implementation of visual servoing to find and stay centered on a target of interest. The output of this DNN can be coordinates, rather than a human-interpretable image that is then used to extract coordinates [340]. On the other hand, a human-interpretable image is useful for supervising the automated procedure, intervening if necessary, and providing interpretable reports if there is a runtime error. Thus, the two seemingly competing approaches between robot and human data formatting have symbiotic advantages when operating in parallel [338].

The availability of resources is anticipated to further advance the field with regard to open-source implementations that lower entry barriers into an otherwise specialized field. One such resource was made possible through the Challenge on Ultrasound Beamforming with DL [341, 342]. Outcomes of

this challenge include freely available datasets, code, and trained network weights, which may collectively be employed to benchmark new approaches.

Concluding remarks

Ultrasound and photoacoustic imaging are two technologies that use the same sensing hardware to make images. After raw sensor data is received by ultrasound and photoacoustic imaging systems, there are multiple signal processing and beamforming steps that can be implemented to address a variety of healthcare challenges across multiple organs, diagnoses, and procedures. DL provides a viable pathway to implement multiple approaches in parallel, possibly in a single signal processing step. This pathway is promising to overcome previous barriers to producing high-quality images for all patients.

Acknowledgments

Thanks to all of my past and present Photoacoustic and Ultrasonics Systems Engineering (PULSE) Lab researchers and clinical collaborators who contributed to this direction of research, as well as to all of my CUBDL co-organizers and data contributors who helped to make open-source resources available. This work was supported in part by the NIH Trailblazer Award (Grant No. R21-EB025621), the NIH Trailblazer Award Supplement (Grant No. R21-EB025621-03S), the NSF CAREER Award (Grant No. ECCS 1751522), the NSF SCH Award (Grant No. NSF IIS-2014088), the Computational Sensing and Medical Robotics Research Experience for Undergraduates Program (Grant No. NSF EEC 1852155), and NIH R01-EB032960. More information is available at https://pulselab.jhu.edu and https://cubdl.jhu.edu.

33. Enabling equitable access to DL solutions

Bruno M Saraiva[1,2], *Guillaume Jacquemet*[3,4,5,6] *and Ricardo Henriques*[1,2,7]

[1]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[2]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
[3]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland
[4]Faculty of Science and Engineering, Cell Biology, Åbo Akademi University, Turku, Finland
[5]Turku Bioimaging, University of Turku and Åbo Akademi University, FI- 20520 Turku, Finland
[6]InFLAMES Research Flagship Center, Åbo Akademi University, FI- 20520 Turku, Finland
[7]MRC-Laboratory for Molecular Cell Biology, University College London, London, United Kingdom

Status

The ML field has been a critical ally in advancing and improving microscopy image analysis, with several ML algorithms automating common tasks such as image segmentation and classification. Recently there has been an incredibly rapid development of new microscopy image analysis approaches, thanks to the current boom in developing DL techniques. A DL algorithm will look at data, labeled in the case of supervised learning or unlabeled in the case of unsupervised learning, and through self-optimization, will infer data features that it can then use to perform the desired task. When correctly implemented, it can provide its users access to expert-level performance at unprecedented speed. However, the implementation of DL approaches is highly dependent on the quality and quantity of data used for training and, typically, on the availability of high-performance computers. When used appropriately, DL techniques have shown incredible performance in several image analysis problems such as segmentation (U-Net [343]; StarDist [344]), classification (YOLOv2 [345]), denoising (Noise2Void [161]), restoration (CARE [8]), and super-resolution, either via super pixelation or improving reconstruction algorithms (DeepSTORM [9]).

Although there is no question about the high impact potential that DL has in image analysis, users are often required to either pay for closed-source tools or have programming knowledge to take advantage of these approaches poses a barrier to the widespread adoption of DL techniques in most research fields. In recent years, several efforts have been made to bring these techniques to the non-expert scientific community to tackle these issues. Notorious examples include DeepImageJ [346], which focuses on implementing pre-trained DL models as plugins for ImageJ/FIJI [347, 348], the most widely used tool for image analysis (figure 51). Another example is ZeroCostDL4Mic [164], which provides easy-to-use Google Colab notebooks that allow users to train and use DL models for their microscopy image analysis without programming knowledge (figure 51). Considering ZeroCostDL4Mic runs on Google Colab, it provides users with a cloud-based solution that eliminates the need for dedicated hardware to use DL approaches for their image analysis tasks. Similar strategies aimed at lowering the technical threshold for adoption are emphasized in sections 30 and 27, which highlight PnP interfaces and user-centered design to broaden access to DL in high-throughput screening and cellular phenotyping. DL4MicEverywhere [349] uses the same philosophy of ZeroCostDL4Mic allowing users to train and test DL models without having to interact with code but using their local resources instead of Google Colab. Besides these three, many other projects make available tools using DL microscopy image analysis tools, such as Cellpose [350], CSBDeep [8], DeepMIB [351], and many others.

Current and future challenges

While DL has been providing researchers with new approaches to enhance their image analysis capabilities, it is still a technology with its shortcomings that makes it not easily accessible to every researcher around the globe. As of February 2025, the number of publications containing 'deep learning microscopy' as search keywords compared to publications with only 'microscopy' as keyword shows that even with access to image acquisition equipment, access and adoption of DL approaches is still not widespread, with US and China leading the way (figure 52).

DL algorithms require high computational power, which is expensive and poses a barrier for researchers with limited access to research funds. ZeroCostDL4Mic, which can be run entirely on Google Colab which currently has a free usage plan. However, this plan limits the amount of computational power and storage space that can be used, and there are no guarantees that this might not change or that DL approaches will not evolve to a point where they require more than what is currently provided for free. DL4MicEverywhere features a no-code interface that allows users to train and test DL models in their own hardware, however this means users are required to have access to hardware powerful enough

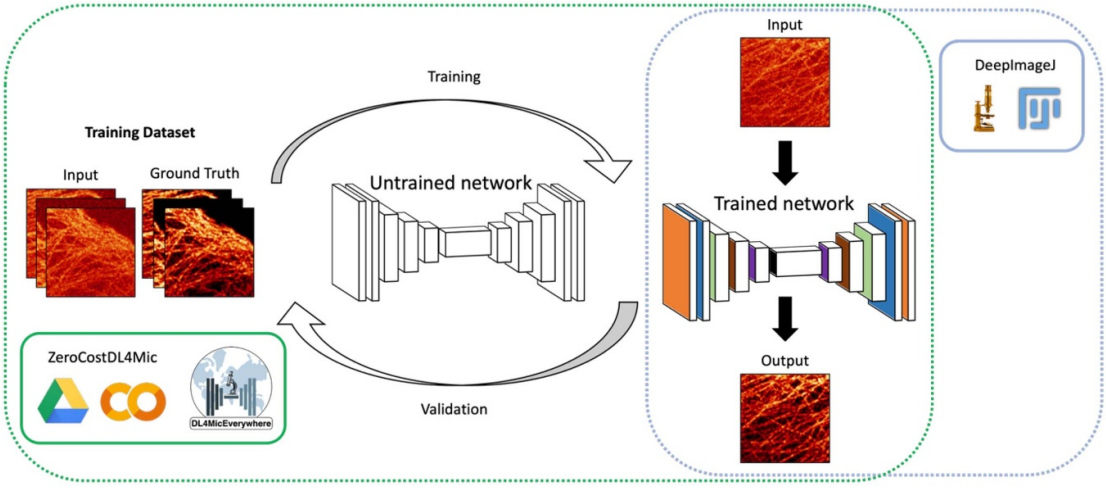

Figure 51. ZeroCostDL4Mic, DL4MicEverywhere and DeepImageJ provide deep learning implementations that require no programming knowledge. ZeroCostDL4Mic, through the combined use of Google Drive and Colab, allows users to upload their data, train their models, and use the trained models in their image analysis workflows. DL4MicEverywhere allows users to train and implement their models using their available hardware. DeepImageJ integrates several pre-trained models as part of an ImageJ/FIJI plugin that users can use without having to train deep learning models with their data.

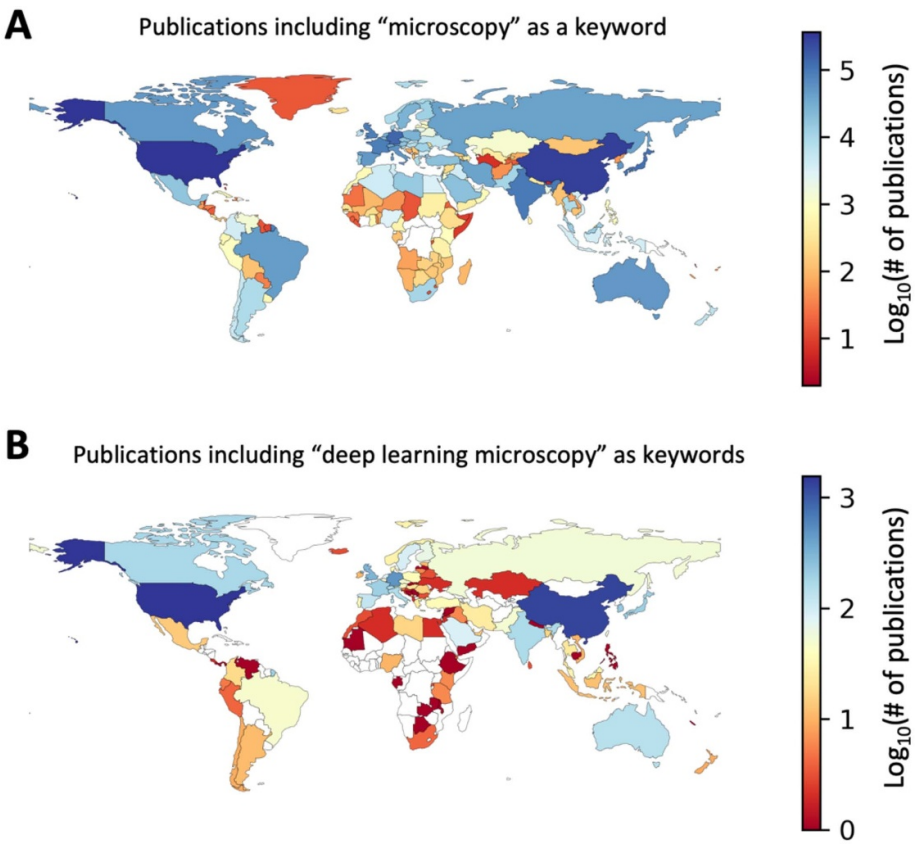

Figure 52. Number of publications including either 'microscopy' or 'deep learning microscopy' as keywords. *Source*: Web of Science for publication records (February 2025) and Geopandas for plotting. White coloring corresponds to no publication records found.

to run these models. In addition to this, there is also the question of environmental impact of implementing and using DL approaches, as not only the energy consumption is considerably high, but there is also the issue of carbon emissions along with hardware production and distribution.

One of the major frameworks for developing DL approaches, TensorFlow, is mainly funded and developed by a private company. While currently it might operate with an open-source model, this might change in the future. This could create a first access problem and further increase the economic

burden of implementing DL solutions. However, it should be noted that this might not be very likely as the company also benefits from keeping TensorFlow open source. Furthermore, as the efforts of DL developments are spread across multiple DL libraries, a major breakthrough in one of them might force users and developers of the alternatives to migrate, leading to additional costs. In addition, to use DL as a tool in their research, researchers must have technical knowledge of programming and image analysis. Although ZeroCostDL4MIC, DL4MicEverywhere and DeepImageJ remove the need for users to know programming, users are still limited to using the included DL models. Recently, the BioImage Model Zoo [352] project, together with its community partners [164, 344, 346, 347, 349, 353–359], have created a standard for deploying DL models for image analysis and a community-driven database of pretrained models. Nonetheless, even with access to pretrained models and frameworks that require little to no programming knowledge, users still need to know when to use specific DL models, how to generate/access and pre-process the necessary data, how to analyze the output of these models, and how to validate the results [360]. This echoes challenges seen in digital pathology (section 25), where clinical users must also grapple with understanding model limitations and interpretation in high-stakes settings. Considering how DL approaches rely on proper datasets for training the models, generating, accessing, and storing data can also be an issue due to the associated costs. All of this can be especially challenging for researchers who are not yet versed in the DL field.

Advances in science and technology to meet challenges

DL is a technology that, although very powerful, carries with it a considerable economic burden to its potential users. Due to this, it is not yet accessible to researchers worldwide, especially those where funds for scientific research are not easily obtained. Projects like DeepImageJ [346] help alleviate this burden by providing users with the means to use already established pretrained models. However, researchers must have their own solutions if they want to work or use new DL models. Google Colab, as used by ZeroCostDL4Mic [164], is an option. However, its free usage plan has limitations and is still entirely reliant on a service provided by a private company with its own financial interests. As such, creating publicly funded cloud-based solutions that can be commonly used by researchers worldwide will be key to making DL accessible to every researcher. The European project AI4Life is focused on bringing sustainable quality research infrastructure and services to enable life sciences researchers to access DL image analysis tools by creating a bridge between life and computer sciences. The Chan Zuckerberg Initiative is also contributing to bring DL to life sciences researchers by funding several projects that focus on implementing DL approaches in Napari [361], a Python based open-source image processing tool. As a key factor in DL, data access can also be a constraint in using DL in microscopy. Ensuring that the acquired data follow the FAIR principles (findable, accessible, interoperable, reusable) can promote data sharing to the whole scientific community and enable DL solutions to research groups that might not have the means to generate the required data. An example of how sharing data can have a real impact is the work developed by Abdurahman *et al* [362], in which by using a publicly available dataset they were able to implement a DL strategy to detect malaria parasites in thick blood smear microscopic images. Sharing data and pre-trained models will also be key in reducing the carbon footprint inherent to the need for high-computational power required for DL. This emphasis on FAIR data principles and sustainability also aligns with community-driven model hubs like DeepImageJ and BioImage Model Zoo (sections 33 and 35).

In addition to the economic burden that implementing DL approaches entails, the knowledge required to take advantage of this technology hinders the adoption of DL for image analysis. The image.sc forum is an example of a community-driven knowledge network that can help and guide new users who want to deepen their knowledge in image analysis, including with DL implementations. Online courses and training sessions will also be fundamental to bringing DL to every researcher, as they inherently have fewer associated costs, making them more inclusive than courses requiring in-person attendance, and reducing the carbon footprint associated with travelling to in-person events.

Concluding remarks

There is no question that DL revolutionized the field of microscopy image analysis. DL approaches have outperformed many classical image analysis tasks, providing researchers with state-of-the-art performance at unprecedented speeds. However, due to the need for specific knowledge and equipment to implement a DL approach, it is still a tool that is not easily accessible to all researchers. Several recent projects have contributed to make DL approaches more accessible by removing the need for programming knowledge to use DL for microscopy image analysis. Nevertheless, there are more challenges to be solved. Having access to quality data is a fundamental prerequisite for DL model training. However, accessing the data required for a DL approach is not always possible for all research groups. Understanding when

and how to use DL and what pre-processing is needed can also be a limiting factor that could be solved by creating a community-driven knowledge network and training in using these approaches. As a scientific community, we need to join efforts to develop and implement strategies that can make DL equitable and available to the global scientific community.

Acknowledgments

R.H. and B.S. are supported by Gulbenkian Foundation and received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 101001332), and the European Union through the Horizon Europe program (AI4LIFE project, grant agreement 101057970-AI4LIFE), the European Molecular Biology Organization (EMBO) Installation Grant (EMBO-2020-IG4734), and the Chan Zuckerberg Initiative Visual Proteomics Grant (vpi-0000000044). G. J. is supported by the Academy of Finland (G.J. 338537), the Cancer Society of Finland (G.J.), Åbo Akademi University Research Foundation (G.J., CoE CellMech), the Solution for Health strategic funding to Åbo Akademi University (G.J.) and the InFLAMES Flagship Programme of the Academy of Finland (G.J., 337531).

34. Deployment of DL applications

Wei Ouyang[1] *and Trang Le*[2]

[1]Science for Life Laboratory, Department of Applied Physics, KTH—Royal Institute of Technology, 171 65 Stockholm, Sweden
[2]Department of Bioengineering, Stanford University, Stanford, CA 94305, United States of America

Status

Due to their extraordinary performance, DL models have gained rapid popularity in microscopy image analysis. While DNNs are performant in a wide range of tasks, they pose unprecedented challenges in software design and deployment due to the amount of data and computation required for running them, especially in the training phase. Under the hood, most of the DL methods are implemented using one of the few popular DL frameworks including PyTorch [363] and Tensorflow [364], which are implemented in Python. Source code for these methods is wrapped into repositories and shared via online platforms such as GitHub under permissive licenses. The openness of the DL field greatly contributes to the wide spread of popular models and further development, it has become a common practice.

In addition to the raw python code, Jupyter notebooks [365], which split the source code into sections and are surrounded with explanatory text in an executable document, are often provided for demonstrating the usage of the interface functions in the companion code repository. Notebooks are widely used for educational purposes and provided in hands-on workshops and tutorials in microscopy image analysis. To improve the reproducibility, container-based cloud execution services such as Binder [366] and Google Colaboratory are provided for free and become valuable resources in the deployment of DL applications. Colaboratory in particular, provides free GPU which is ideally suited for distributing DL tools and it is in fact used by ZeroCost4Mic [164], which provides a collection of curated notebooks for running DL based microscopy image analysis configured via interactive user interface elements in the Colaboratory notebooks. This aligns with accessibility-focused platforms discussed in section 32, where deployment strategies aim to overcome hardware limitations for global adoption.

While it is flexible to use Python source code or Jupyter notebooks, users with little programming experience can easily adapt the tool according to instructions provided by the developers. For users without programming skills, it is still challenging to adapt the code or notebook to work with conventional software, e.g. ImageJ [367], in a more complete analysis workflow. Despite it being challenging to run DL in programming languages other than Python, Java solutions such as DeepImageJ [346] have been proposed. To address the need, napari8, which is a trending Python-based image analysis software supported by Chan Zuckerberg Initiative, is being developed and gaining traction in the community. The aim of the project is to provide a fast, interactive, multi-dimensional image viewer for browsing, annotating, and analyzing large multi-dimensional images.

Furthermore, due to the data- and computation-hungry nature of DL methods, web and cloud computing has become increasingly important for the further scaling of the applications to handle massive datasets with almost unlimited access to data storage and compute power. On that front, web based platforms such as ImJoy [353], CDeep3M [368], and DeepCell Kiosk [369] are developed for supporting DL applications running on the server side. Since the computations are carried out on remote servers that are maintained by IT experts or developers, users can use these tools with little or no setup, in most cases, using a web browser to access the user interface. This type of deployment approach is more scalable compared to the conventional desktop software, however, it requires transmission of potentially large amounts of data to remote servers which can be limited by the bandwidth of the internet connection, it poses challenges on server-side data confidentiality and privacy concerns. In practice, the user will also require more feature-rich software running fully in the web browser or in the cloud to avoid moving data between local vs remote in a more complete analysis workflow.

Current and future challenges

Currently, it is a pressing need for the community to work together to solve the challenges in the deployment of DL tools. For desktop software, the challenges include figuring out how to ship the software packages, complex dependencies, making it easier to install, and reliably work under the several types of mainstream operating systems. Web and server-based deployment options are becoming more common. It alleviates the deployment issues by delegating the task to IT experts for setting up the complex software environment and accessing cloud storage and computational resources. At the same time, sustainable funding models and long-term political commitment are needed to keep these cloud resources openly available to the research community.

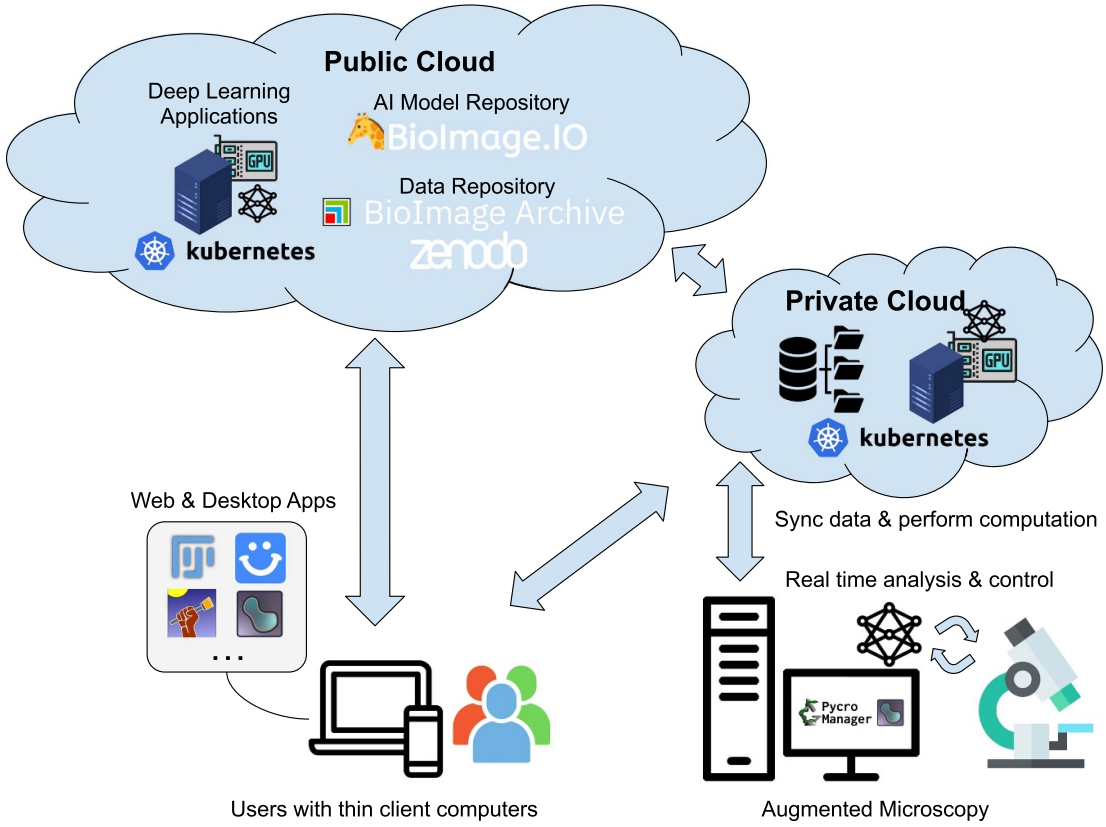

Figure 53. Scalable web-powered computational systems for deploying deep learning tools for microscopy. To build and deploy deep learning tools for handling massive data with computational heavy models, we need to design distributed systems in a public or private cloud environment. Local workstations are used to control the microscope and acquire data, potentially in an augmented microscopy setting where deep learning tools (built with, e.g. Pycro-Manager or napari) are used in real-time to perform real-time image analysis and generate feedback signals for controlling the microscope. Within the same institutional network, data will be synchronized into central data management systems and further analyzed using deep learning tools deployed as containers in the private computing cluster managed by Kubernetes. Data accompanied with publications can be further published to the public cloud (the internet) in data repositories such as BioImage Archive or Zenodo, and the corresponding deep learning models can be published to the BioImage Model Zoo (at BioImage.IO). Tools are deployed to public cloud servers to provide services for data management and annotation, deep learning model training and inference etc. Users with a thin client computer including portable laptops and mobile devices use web or desktop applications to access the web services running in public or private cloud servers. Due to the fact that heavy computation and massive data are handled by servers, users can use a web browser to do data management, annotation and analysis tasks, e.g. in ImJoy.

Future-proof AI systems for microscopy require a scalable human-compatible framework. To work with the ever-increasing amount of data, it is inevitable to utilize a centralized computing cluster hosted by an institutional IT department or in the cloud. In the meantime, users tend to use a 'thin' client such as a laptop, tablet, or mobile devices for accessing the services. Different from conventional desktop software with user interface and the computational code coupled to the same software module, cloud-facing software requires a major change in the design pattern that separates the user interface and the compute parts. As shown in figure 53, while the user interface parts run in the user's web browser or a desktop client, the main parts contain DL models and other heavy computation runs in one or multiple remote servers in a public or private computing cluster. The two parts need to be synchronized via communication over the internet, and often implemented in different programming languages, e.g. HTML/CSS/JavaScript for the interface and Python for the compute part. To make the transition smoother for the next generation of tool developers for microscopy image analysis, we will likely need coordinated efforts in the community to create tools and platforms, and produce educational materials to simplify the process.

In addition, reducing the computational costs and making ML training more environmentally friendly is an important aspect to consider when distributing DL tools. Meanwhile, for patient related microscopy images, data confidentiality and privacy represent another dimension of challenges in the deployment of DL tools.

Advances in science and technology to meet challenges

To address the challenges in shipping desktop software with complex dependencies, conda-like virtual environments or container-based solutions are used as an alternative with the price of an increased package size. For example, to take advantage of the existing Java-based developers in the ImageJ community, there is an ongoing effort of building bridges between Python and Java in the pyimagej project to allow easy integration of DL models in ImageJ. Pycro-Manager is another software which connects python with micro-manager and further enables real-time DL powered image analysis and feedback control during image acquisition (figure 53). In the meantime, since it contains binary compiled for a specific operation system, it does not guarantee that it works for all the cases. For cloud based solutions, Container orchestration platforms represented with Kubernetes are becoming the de-facto standard for institutional IT and academic cloud providers to provide a managed environment for developers to deploy their DL tools. AI model training and serving software such as KubeFlow and Nvidia Triton inference server makes it easier to serve AI models for production and executed remotely via HTTP-based interface. For data confidentiality, while private computing clusters (as shown in figure 53) may address some of the challenges, federated learning [370] enables the training of powerful models while keeping data private in their own data lakes.

To facilitate the sharing of AI models, repositories such as the BioImage Model Zoo [352] (https://bioimage.io) are developed to facilitate the sharing of pre-trained models, and softwares is joined by a consortium to define common model formats to enable cross-compatibility. These efforts make it easier to distribute DL models and can be used in multiple software. Meanwhile, re-using existing models either as is or a warm start for training can greatly reduce efforts in producing models and further contribute to climate change. These model-sharing initiatives are central to the sustainability and democratization goals also emphasized in sections 32 and 36. In addition, model compression techniques such as knowledge distillation are used to reduce the model size and accelerate the model execution in e.g. augmented microscopy to provide real-time feedback.

In the browser, building a rich and powerful user interface is becoming easier and more reliable. WebAssembly, which enables compiling and running foundational scientific software packages written in C/C++, Rust, etc in the browser. It makes it possible to reuse Python libraries such as numpy, scipy, pandas and scikit-image for loading and processing images directly in the web browser and paves the way for creating powerful in-browser image analysis tools with easy-to-use user interface. ImJoy is a framework built for taking advantage of the web ecosystem and providing a remote procedure call layer to connect plugins running in the browser or in a remote server. On top of ImJoy, ImageJ.JS (https://ij.imjoy.io) is a tool we developed by comping ImageJ in Java to JavaScript which is now being used by ∼1000 users per day.

Concluding remarks

With massive natural language models such as OpenAI GPT-3, Codex [371] or ChatGPT, it allows generating executable source code in various programming languages such as Python and JavaScript from plain English. This opens a new door for future DL tool developers to create simplified voice or chat-bot like interfaces for reducing the complexity of user interface design and making the tools more flexible and human-compatible. However, the sheer size and computation required to run massive models virtually rejects the access for low-budgeted research entities, and it makes tech giants become the 'natural monopoly'. The wider AI community will need to join efforts and explore the way forward.

Overall, the wide adoption of AI solutions in microscopy imaging is leading to the paradigm shift to a future of augmented microscopy powered by human-in-the-loop AI, and generating profound changes in the way we understand biology and contribute to precision medicine and healthcare. Ultimately, the 'democratization' of DL microscopy will hinge as much on political will as on code quality. A concerted global effort—uniting researchers, national facilities, and funding agencies—can create a federated, publicly accountable cloud that rivals commercial offerings and guarantees that scientific data remain a public good.

Acknowledgments

This work is funded by the Knut and Alice Wallenberg Foundation (Grant Nos. 2016–0204 and 2018.0172).

35. DeepTrack 2

Benjamin Midtvedt and Jesus Pineda

University of Gothenburg, Department of Physics, Sweden

Status

The ability to quantitatively analyze the microscopic world was enabled by two relatively recent advances [94, 372]. First, digital video microscopy has allowed researchers to numerically represent visual data and record it for later analysis. Second, the explosive growth of computing power has made the expensive, high-dimensional analysis of video recordings possible.

Recently, we have seen a new wave of developments in the analysis of microscopy video data, thanks to the power of DL. Indeed, DL has shown remarkable performance on many common tasks in microscopy, such as cell counting, object detection, cell morphometry, trajectory reconstruction, super-resolution microscopy, diagnostics, object classification, and cross-modality transformations [373, 374]. However, DL has not yet been broadly adopted as an analysis tool in digital microscopy, mainly because of the significant barrier-to-entry to the development of custom DL solutions for microscopy data.

Here, we present DeepTrack 2, an open-source Python library equipped with all the necessary tools to produce a full, E2E microscopy-analysis pipeline [94]. A common example of such a pipeline designed to extract mechanistic and spatio-temporal information from biological data entails the following steps:

1. The image is prepared for analysis.
2. Objects of interest are detected and measured to extract morphological and intensity information.
3. Detections in different video frames are connected into trajectories.
4. Time-resolved information is combined to gain both measures of both local and global properties.

DeepTrack 2 provides DL solutions for each step of this processing pipeline, allowing users with varying programming experience to use and optimize the available solutions to their data, goals, and challenges. For step 1., DeepTrack 2 integrates tools for data normalization, noise suppression, augmentations, and even more advanced methods such as virtual staining to produce high-contrast and high-specificity images (figure 54(a)). For step 2., DeepTrack 2 ships with state-of-the-art models ready for training, such as LodeSTAR [208] for label-free particle detection, U-Net for semantic segmentation, and YOLO for simultaneous detection and classification (figure 54(b)). These capabilities directly support workflows described in section 21 (cell dynamics) and section 14 (single-molecule localization), where tracking and segmentation across time are essential for capturing cellular behavior. For step 3. and 4., DeepTrack 2 uses MAGIK [46], a state-of-the-art graph-based network which both connects observations into trajectories and extracts information about the dynamics of the system from spatio-temporal data (figure 54(c)).

Recent research has demonstrated the benefit of DeepTrack 2 to analyze microscopy data. For example, figure 55(a) shows how virtual staining can unveil high-quality visualizations of complex biological systems from cheap-to-capture brightfield images using conditional generative adversarial neural networks (cGANs) [98]; figure 55(b) demonstrates that close-to-perfect object detection can be achieved from just a single unannotated image using LodeSTAR [208]; and figure 55(c) shows how GNNs such as MAGIK can be used to connect detections into traces even if the cells divide [46]. Furthermore, DeepTrack 2 has been used for the characterization of microplanktons from Holographic microscopy images [375] and the monitoring of active droploids, a new class of active matter systems [376].

Importantly, DeepTrack 2 provides a hardware-agnostic API so any analysis pipeline is instantly reusable, reproducible and future-proof across microscopes and laboratories. These are critical advantages necessary for any microscopy-oriented DL frameworks.

Current and future challenges

The transition to DL-enhanced analysis has not been without challenges. Firstly, the high variability between imaging modalities and object samples has made publicly available datasets an unreliable source of training data. The dataset would need to contain the same object of interest, imaged through a near-identical optical device, and annotated with the desired ground truth. The likelihood of these coinciding is slim-to-none. Comparable difficulties in sourcing compatible training data are encountered in sections 29 and 31 (TEM and ultrasound/photoacoustic imaging), where instrumentation-specific constraints create barriers to generalized model reuse.

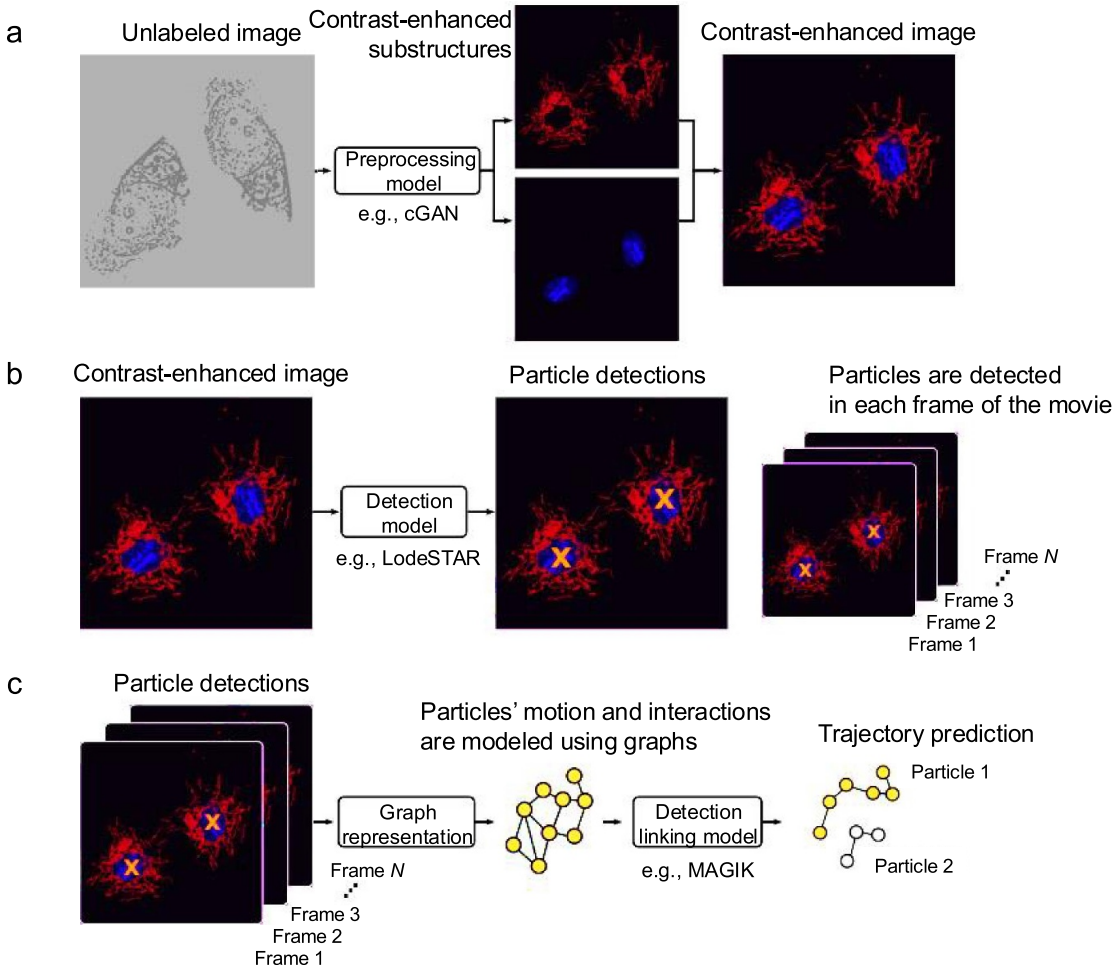

Figure 54. DeepTrack 2 integrates a deep learning ecosystem for end-to-end microscopy data analysis. (a) An input image is optionally processed by an image-to-image neural network to, for example, highlight regions of interest. (b) Structures of interest (here, the nuclei of the cell) within each frame of a video are detected and marked (orange markers). (c) The detections in each frame are linked to form trajectories over time using graph neural networks.

Secondly, for many applications, the expected quality of analysis exceeds human precision. Examples include sub-pixel localization of objects for super-resolution microscopy and force calibrations, object detection in noisy images, 3D-microscopy and more. Consequently, even high-quality human annotations are insufficient for training these NNs.

Thirdly, NNs have deservedly gained a reputation as black-box functions, calling into question their reliability for clinical and industrial use. Consequently, DL has struggled to gain traction beyond a research tool. A similar concern about black-box models hindering clinical adoption is expressed in section 28 (neuroimaging), where interpretability is a key requirement for clinical deployment. Fundamentally, this is because NNs only yield answers with little explanation of how that answer was acquired.

Finally, the rapid development of methods and model architectures has resulted in a scattered field without a unified interface. Consequently, comparing methods is prohibitively difficult. Moreover, reusing DL methods for new data often requires re-implementing them from scratch, which is a daunting task for non-experts.

These four challenges combined have resulted in a slow adoption of new DL methods, despite their significant advantage in performance. DeepTrack 2 attempts to address these issues through a three-pronged approach:

1. The use of synthetic data for training NNs to help reduce the reliance on annotated data while generating perfect ground truths [94, 375].
2. The development of label-free, low-shot, and interpretable models to promote low-cost methods with less opaque NNs [46, 208].

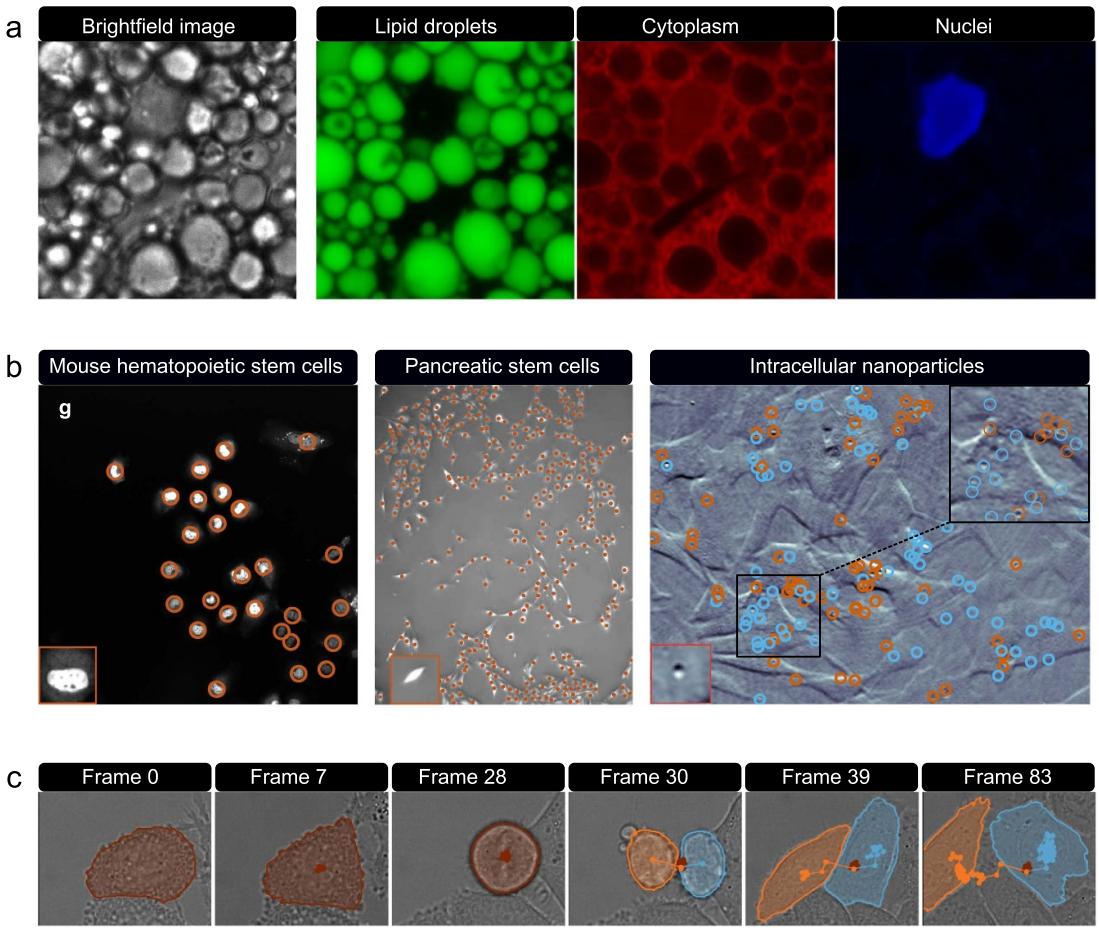

Figure 55. DeepTrack 2 analyzes experimental data. (a) Brightfield images of adipocytes are virtually stained to highlight the lipid droplets, cytoplasm and nuclei. Reproduced from [98]. CC BY 4.0. (b) LodeSTAR learns unsupervised to detect mouse stem cells (despite the high variability between cells), pancreatic stem cells (despite the densely packed sample), and intracellular polystyrene particles (orange markers), and biological structures (blue markers), despite the low signal-to-noise ratio. Reproduced from [208]. CC BY 4.0. In all cases, LodeSTAR was trained on the single crop shown in the respective inset and then applied to the whole image. (c) MAGIK learns to track Hela cells (despite being shape-varying cells), even when cells are dividing. Reproduced from [46]. CC BY 4.0.

3. The unification of methods into a consistent interface, by continuously implementing state-of-the-art methods and encouraging authors to contribute their methods [377].

Advances in science and technology to meet challenges

We identify three avenues of research essential to bring out the full potential of DL for quantitative microscopy:

1. Further development of unsupervised or self-supervised methods, particularly for tasks such as segmentation, trajectory reconstruction, and classification. While the scaffolding exists for these developments through methods such as self-distillation and contrastive learning, they have yet to be optimized for the specific challenges of microscopy.
2. Development of interpretable NNs. The recently popularized attention-based NNs have allowed much more transparent NNs than previously possible. Attentive NNs can highlight the parts of the data that lead to a particular conclusion, providing insight into the reason behind an answer. This echoes interpretability efforts in human-in-the-loop imaging systems, as discussed in section 27, where AI is expected to augment user trust through transparent decision-making. Currently, this is mainly used for classification tasks. However, we see no reason why attention-like mechanisms cannot be incorporated into NNs designed for other tasks in microscopy, significantly increasing the interpretability and thereby the trustworthiness of the NNs.
3. Design of advanced simulation methods for various optical devices. Unlike the impossibly complex macroscopic world, the physics of the microscopic world can feasibly be fully simulated. As such, the

need for annotated data can be done away with entirely by developing faster and more accurate simulation techniques.

Concluding remarks

DL is undeniably a powerful tool for quantitative microscopy. Nonetheless, it has remained a research tool instead of reaching the hands of clinicians. We identify four key challenges that need to be overcome to reach the full potential of DL for quantitative microscopy. These are: a reliable source of training data that is general enough to match the specific problem and experimental device of the end user, a method of training NNs beyond the limit of human accuracy for annotation, the development of less opaque models, and a unified interface for using these models.

We also consider the field in a good state to tackle these problems. Self-supervised methods are on the rise, and interpretable layers such as attention layers are taking over the field. Targeted development of these two approaches for microscopy may lead to the widespread adoption of reliable DL methods, revealing physical and biological insights encoded in the data in an unsupervised manner.

36. DL in ImageJ

Estibaliz Gómez-de-Mariscal[1,2,3], *Daniel Sage*[4] *and Arrate Muñoz-Barrutia*[5,6]

[1]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[2]Gulbenkian Institute for Molecular Medicine, Oeiras, Portugal
[3]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
[4]Biomedical Imaging Group and EPFL Center for Imaging, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[5]Bioengineering Department, Universidad Carlos III de Madrid, Leganés, Spain
[6]Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

Status

As in other scientific fields, DL has demonstrated outstanding performance for many microscopy image processing tasks [378, 379]. Its potential to reveal visual features in complex images has changed the game in microscopy image analysis; especially the case when the visual features are unexplainable by humans, like for image restoration (e.g. CARE [8]), for dense nuclei detection (e.g. StarDist [380]), for super-resolution localization microscopy (e.g. DECODE [177]), or for correlative microscopy imaging. Although recent contributions translating DL to microscopy imaging have empowered researchers with the capacity to build powerful pipelines, the IT barrier is still too high for most end-users. It requires technical competencies and programming expertise to fully exploit this new technology. Similar concerns about accessibility and user interface limitations were highlighted in sections 32 and 33, where the need for GUI-based and cloud-integrated frameworks like DeepImageJ and ImJoy is seen as key to equitable deployment. Life scientists are usually uncomfortable with Python, the preferred language for DL frameworks (TensorFlow, PyTorch). Instead, they significantly prefer the friendly user interface of Java-based software like ImageJ [367], Icy [357], or QuPath [358]. ImageJ is currently the most used software in cell imaging, as it provides a rich palette of bioimage analysis tools. Moreover, it can interoperate with many other platforms in a unique ecosystem, including drivers for microscopy or data analysis packages. Nevertheless, the hardware restrictions of Java have prevented its use as a reference tool for bioimage analysis with ML. Hence, the users of ImageJ have been restricted from deploying DL technology in their image analysis pipelines.

In the last few years, several initiatives have already been taken to integrate DL features in ImageJ. Several teams (DeepClass4Bio [381], CSBDeep [8], deepImageJ [346], JDLL [382]) are working to provide access to pre-trained DL models through ImageJ plugins or ImageJ macros. In the context of ImageJ, deepImageJ is the generic consumer of pre-trained DL models. On the other hand, well-established ImageJ plugins have also integrated DL in their pipelines. For instance, TrackMate [383] has included cell segmentation with CellPose or StarDist to extend the versatility of the tracking pipeline. Finally, the BioImage Model Zoo [352] is a community-based initiative that gives effective access to DL technology. It proposes a standardized format tailored for life scientists to share and deploy trained DL models across open bioimage analysis software in a reproducible manner (e.g. ImageJ, Ilastik [354], ImJoy [353], ZeroCostDL4Mic [164], QuPath [358], biaPy [326], DL4MicEverywhere [349]).

Current and future challenges

We identify challenges at two levels for the full integration of DL in microscopy image analysis. Current DL approaches have the capacity to boost the limits of microscopy image acquisition [8, 177]. Still, their performance relies on the quality and the quantity of data, both acquired raw images and annotated or reference images (figure 56(c)). Usually, scientific images are acquired to answer specific biological questions. Unfortunately, researchers look for a suitable method to analyze once all the images are acquired. We envision that integrating the data analysis at the initial steps of the project could significantly improve the imaging workflow's robustness (figure 56(a)). This is particularly demanding for data-centric approaches where the images are the heart of the image-analysis task. The acquisition protocol should be appropriately designed for this purpose, including a large variability of examples to reduce the time and effort needed to obtain scientific results.

The deployment of DL methods routinely is not straightforward. Important technical niches in the ImageJ ecosystem (figure 56(b)) are (1) the lack of user-guided training, (2) the easy use of DL models on large images, for which approaches such as BigDataViewer [384] could potentially be a powerful solution, (3) the connection to high-performance computing resources (GPU) in line with Fiji HPC Workflow Manager and HPC-Parallel Tools, and (4) the interactive insertion of DL models in complex

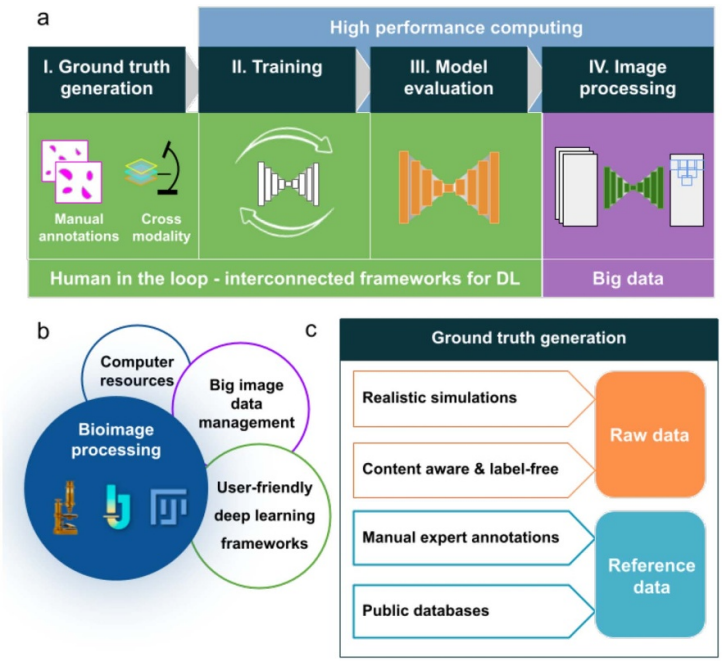

Figure 56. Image analysis workflows and missing niches. (a) General workflow to get a functional DL model and the corresponding steps of development. (b) Identified niches missing and areas to push for DL-empowered bioimage analysis in ImageJ: Computer resources to exploit steps II to IV; big data management to enable additional functionalities in step IV; user-friendly deep learning frameworks to enable steps II–IV. (c) Strategies to obtain the data needed for DL model training: realistic simulations by using a physical mo content-aware approaches to design the microscopy acquisition; tedious expert annotations or crowdsourcing annotation; publicly available datasets through image-analysis challenges and open publications.

bioimage analysis pipelines. Likewise, more optimal integration of DL into bioimage analysis pipelines relies mainly on the education of the practitioners in ML and image analysis. In practice, it is essential to train the researchers on the appropriate use cases and warn them about the limitations and risks of the DL technology. The pandemic significantly increased participation in recorded online courses (NEUBIAS Academy, EMBO, and EMBL) and the edition of user guides of good practices. The community needs to keep the momentum for transferring this knowledge. Therefore, this is a crucial challenge to tailor the current DL knowledge and technology to users in terms of data acquisition and usable computational tools for analysis.

As part of AI technology, DL is a very active field. Preparedness for the integration of upcoming new methods in ImageJ is still a bottleneck. Some still missing techniques are (1) human-in-the-loop, (2) auto ML, (3) self-supervised or weakly supervised training, and (4) tracking of biological particles. Ultimately, the next generation of DL approaches seeks smart workflows capable of embedding the prior knowledge (e.g. with prompt engineering) to a guided analysis through largely generalizable foundation models [321], as well as sparse information to run weakly supervised and few-shot training. All these developments require a deep understanding of both DL and bioimaging and, therefore, the close collaboration between developers and end-users. This mirrors the collaborative development challenges described in DeepTrack 2 (section 34), which emphasizes standardization, interpretability, and the integration of community-driven contributions. Thus, it is a great challenge to push DL to its full potential for smart imaging in biomedical research.

Advances in science and technology to meet challenges

DL is data-hungry, and its reusability can only be accomplished through model training or fine-tuning. Aware of this, the focus on acquiring FAIR data is notoriously growing in academia. The number of publicly available databases tailored for DL model training is increasing more quickly in CV than in microscopy (Data Science Bowl [379], MONAI (https://monai.io)). Data generators and simulators are another source of microscopy images for training; they are specifically suitable when the physical laws of the image formation model are well identified.

The performance of the NNs depends on the amount of data available or on the accessibility to validated trained models. For example, the resolution and SNR needed for a specific measurement will vary according to the structures to analyze. Some works provide guidelines about these features so researchers

can optimize their image acquisition [8, 177, 380]. Furthermore, recent approaches focus on adjusting the content in the image—content aware approaches [8]. The aim is to push the limits of data acquisition and virtually recover biologically relevant information from simpler or more sample-friendly acquisitions [8, 177, 385]. Ultimately, the ability of DL algorithms to encode a large amount of information from the images in latent spaces enables the discovery of different biological behaviors.

Existing user-friendly software for DL typically targets a specific step of a general DL-based image analysis workflow. For example, ZeroCostDL4Mic has democratized the training procedure and the assessment of several NN architectures. It allows most of the steps in figure 56(a) using the free cloud computing Google Collaboratory. ImJoy [349] proposes a framework to interact with virtual content beyond DL directly in the browser without the need for any technical installation steps. This is a big step towards connecting with cloud computing. Ilastik is the reference tool for image segmentation using ML, and now, it can also run DL-trained models. For ImageJ, we found third-party plugins, mainly for prediction, DeepClass4Bio [381], CSBDeep [8], StarDist [380] and deepImageJ [346]. The harmonization of pre-trained models across these tools strongly connects to efforts in the BioImage Model Zoo (sections 26 and 30), where reproducibility and accessibility are central design goals. The latter has enabled the generic use of trained models of the Bioimage Model Zoo and their combination to deploy advanced bioimage analysis pipelines [386].

As the need for more powerful computational hardware increases with new AI-based approaches, cloud computing or network-embedded infrastructures are becoming essential to ensure translational technology in microscopy imaging [349]. This direction is aligned with cloud-enabled deployment strategies described in section 33, where browser-based and containerized models are the foundation of modern DL scalability. Shared infrastructures may be the key to reducing the carbon footprint of energy-intensive training. The new European infrastructure AI4Life (www.eu-openscreen.eu/projects/ai4life.html) aims to provide sustainable, intuitive, and highest-quality research services and infrastructures that will enable all life scientists to exploit ML to improve the utility and interpretability of image data.

Concluding remarks

Integrating DL in ImageJ as the initial steps provided by DeepImageJ is crucial for the computational microscopy imaging area. Friendly access (up to zero code) and accessibility to DL tools are required for democratized access to powerful DL solutions. By open access to pre-trained models and related data, the users can test them and understand their potential and the limitations of this new technology. Such testing and reusability workflows are also exemplified in section 30 (high-content screening), where pre-trained architectures are adapted with minimal training for broad biological utility.

Community-driven initiatives, such as the Bioimage Model Zoo, are significantly expanding the reach of DL in open-source bioimage analysis software. Among these efforts, deepImageJ plays a pivotal role, bridging the gap between pre-trained DL models and ImageJ users and facilitating a more seamless integration of DL capabilities into the ImageJ ecosystem.

The BioImage Model Zoo [352] initiative is actively addressing key challenges outlined in the previous section. In particular, standardizing model formats enhances cross-compatibility across different platforms, contributing to the broader dissemination of robust DL models. The growing availability of advanced DL tools supports the adoption of holistic approaches in life-sciences research. As technological advances continue to drive innovation in this field, we are approaching the next major breakthrough-one that could be as transformative for microscopy imaging as AlphaFold has been for structural biology.

Acknowledgments

We acknowledge the support of Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, MCIN/AEI/10.13039/501100011033/, under Grant PID2023-152631OB-I00*, co-financed by European Regional Development Fund (ERDF), 'A way of making Europe' and the European Commission through the Horizon Europe program (AI4LIFE project, Grant Agreement 101057970-AI4LIFE). E.G.M. was supported by the Gulbenkian Foundation and the European Molecular Biology Organization (EMBO) Installation Grant (EMBO-2020-IG-4734) (Granted to the Optical Cell Biology laboratory at Instituto Gulbenkian de Ciência) and Postdoctoral Fellowship (EMBO ALTF 174–2022).*

37. Hackathons to spur innovation

Ebba Josefson Lindqvist and Johanna Bergman

AI Sweden, Gothenburg, Sweden

Status

Research and innovation do not happen in isolation. Existing and acting within an ecosystem, as illustrated in figure 57, is becoming increasingly important for organizations as AI is becoming a key technology in almost all industries [387]. Few organizations have all the competence, data, and infrastructure needed to fully apply ML at scale. To address real world research and innovation challenges, both extensive domain knowledge and ML expertise is needed. Hackathons are one way of addressing this issue and facilitate collaborations.

The ML community has a long tradition of hackathons and competitions [388]. There are several platforms, the most famous being Kaggle and Codalab [389, 390]. Hackathons are traditionally focused on either a specific problem within ML or a problem formulation around a dataset. AI Sweden, the Swedish national center for applied AI has hosted several hackathons focusing on industry data and problem formulations from industry partners [391]. The key success factors have been, top management commitment from industry, a clear problem formulation, close collaboration with domain experts, accessible data, well defined evaluation metrics, and access to the computational resources needed.

One example within microscopy was the Adipocyte Cell Imaging Challenge, hosted together with AstraZeneca. The problem formulation and data was provided by domain experts at AstraZeneca. AI Sweden invited researchers and AI experts, provided a collaborative platform, access to data and computational infrastructure. The task was to utilize ML to predict the content of fluorescence images from the corresponding bright field images [392] (see figure 58 for example images). This task closely parallels the goals of virtual staining discussed in section 26, where DL models are trained to predict high-information content from label-free microscopy data. Results from this specific hackathon resulted in new industry-academia collaborations, publications [393], and most importantly solutions that could be directly applied by AstraZeneca benefitting their research.

Current and future challenges

There are several challenges that need to be addressed in order to make open hackathons more beneficial for the application of ML to microscopy. Namely:

Data access

Access to data is key for developing ML applications for microscopy. In many cases the data needed is collected within industry. Such data is rarely accessible for academic researchers, startups or other companies. At the same time, these organizations typically hold a lot of the ML competence needed but still may not be able to use all the resources needed to make the full potential use of the data collected.

The main reasons for industry data to not be shared are e.g. the traditional way of thinking of the data as proprietary and too valuable to share openly as, most of the time, considerable investments have been made in collecting the data. Sharing data also opens up for questions concerning legal issues, uncertainty regarding business models for data, and how to know the value of your data. Legal considerations may include questions regarding Intellectual Property, General Data Protection Regulation, and what is the appropriate copyright license to be used. In general the more openly a dataset may be used the more it opens up for new advances in research. To open up for access to a dataset may also be crucial to enable publications in ML and microscopy to enable peer-reviews. This tension between proprietary datasets and the need for open access to enable reproducibility and reuse has been discussed in section 32 on equitable access to DL solutions. This requires an open and new mindset for many industry actors.

Domain knowledge and cross-disciplinary research

To understand the data, problem formulation, and interpret the results, extensive domain knowledge is needed. During a hackathon, access to domain knowledge and possibility to work in cross-functional teams is of critical importance. From the Adipocyte Cell Imaging Challenge, we could see that teams with cross-disciplinary expertise performed better than teams coming solely from the ML field. This despite the fact that all teams had the opportunity to have daily contact with the domain experts from AstraZeneca.

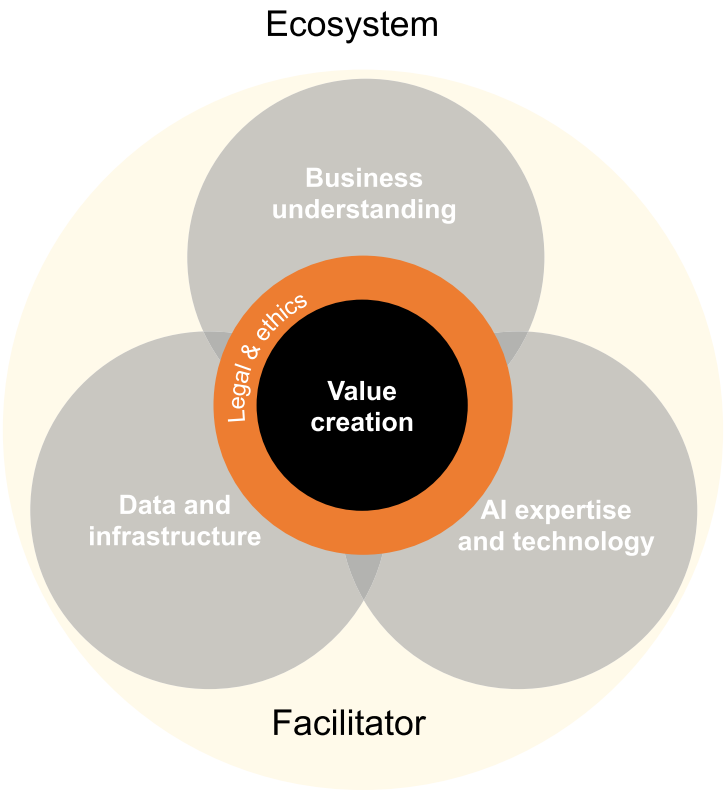

Figure 57. Innovation ecosystem. Illustration of the components needed for successfully applying machine learning to research fields such as microscopy.

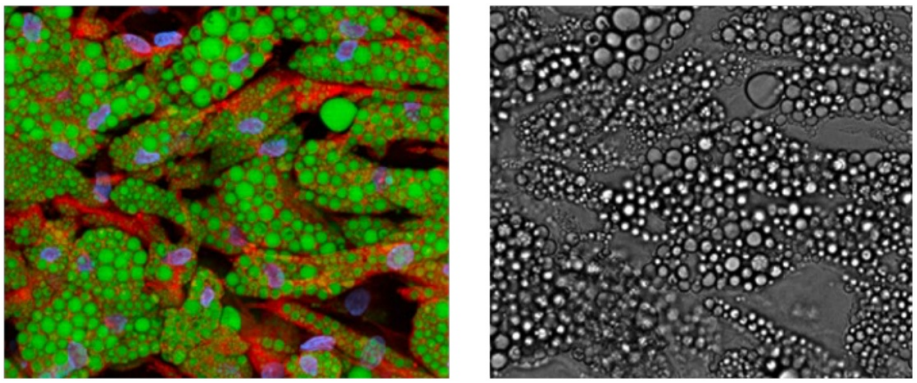

Figure 58. Adipocyte cell imaging challenge. Example images from the Adipocyte Cell Imaging Challenge hosted by AI Sweden and AstraZeneca in 2021.

Traditionally pharmaceutical companies have a relatively long-time perspective for research and focus a lot on their internal resources and established academic research collaborations.

The trends in applied ML research are diverging in two directions. One direction towards being as open as possible with strong community efforts, collaborations, and grass-root research initiatives where data, code, and models are typically shared openly and where open source is the default option. On the other hand many advancements in the field are seen within big companies with a lot of resources, in this case models and data are typically not shared. Common for both directions is the speed and amount of resources needed (both in terms of researchers, data, and computational resources).

Naturally, research that also has a large component of data collection or even physical experiments (for example microscopy) have a longer time frame. As ML is becoming an important part of other research fields with the potential to lead to disruptive advances, the different traditions, trends, and time perspectives need to be addressed. This cultural gap between fast-paced ML development and domain-specific biomedical research was also highlighted in sections 28 and 34, where clinical interpretability and reproducible pipelines remain ongoing challenges.

Advances in science and technology to meet challenges

We see two main topics for advancements to further benefit from hackathons and competitions in the intersection of ML and microscopy research, especially around facilitating collaboration between industry, startups, and academia, and the broader ML community.

Data sharing & data business models

Industry data sharing is challenging for several reasons. Our experience is that the organization of the hackathon itself forces the participating parties to come up with pragmatic solutions for data access, from both a technical and a legal perspective, a process that is often lengthy in other projects. Taking a different and more open approach in such a process potentially accelerates the development of new best practices for data sharing. Similar community-driven strategies have been key to the success of platforms like the BioImage Model Zoo and DeepImageJ (sections 30 and 35), which also promote reproducibility through standardized models and shared benchmarks.

In addition to data access we see a strong need for developing business models and legal frameworks to share data and code to create lasting effects of hackathons and competitions. It is equally important to change the mindset of the data being proprietary and create an understanding of the value of sharing data more openly.

Community building and cross-disciplinary research

To succeed with hackathons, building a strong community and platform for collaboration is key. Building a strong community of researchers from different disciplines and organizations will benefit the research field and development of ML applications for microscopy. Organizing hackathons could be one way of building such a community.

Concluding remarks

To conclude, we see that hackathons have potential to further integrate the microscopy research field with ML as one way of exploring new collaborations, methods, and ways of working. In addition, we see an opportunity to build a strong research community around microscopy and ML. To enable such a community, collaborative platforms, data access, legal frameworks and business models around data are needed. Furthermore, cross-disciplinary collaboration and strong domain expertise is key to success when ML is applied to research questions in other fields.

Shorter challenge driven competitions or hackathons should be seen as a way of exploring and initiating new (sometimes unexpected) research collaborations that have the potential to lead to continued long term collaborations. This could also be one way of addressing and challenging the different time perspectives and research traditions of industry and academia, accelerate data sharing, and invite a broader community to take part in applied research.

Collaborative research will be increasingly important as ML becomes an important tool for microscopy research. Based on the experience from the Adipocyte Cell Imaging Challenge, hackathons could be one enabler for accelerating access to industry data developing the collaborations needed within ML research for microscopy. They also serve as testbeds for prototype deployment strategies described in section 33, where cloud-based solutions and containerized workflows are central to real-world scalability. Neutral organizations, such as AI Sweden, or dedicated challenge platforms (e.g. Kaggle, Codalab), can play a facilitating role and reach a broader community.

Acknowledgments

We want to direct a special thank you to AstraZeneca for giving us the opportunity to co-host the Adipocyte Cell Imaging Challenge as well as all the participating teams proving that hackathons are a powerful tool for building new collaborations and accelerating applied research.

ORCID iDs

Giovanni Volpe 0000-0001-5057-1846
Carolina Wählby 0000-0002-4139-7003
Michael Hecht 0000-0001-9214-8253
Artur Yakimovich 0000-0003-2458-4904
Ivo F Sbalzarini 0000-0003-4414-4340
Christopher A Metzler 0000-0001-6827-7207
Mingyang Xie 0009-0000-9405-7937

Kevin Zhang 0000-0002-8798-2443
Isaac C D Lenton 0000-0002-5010-6984
Halina Rubinsztein-Dunlop 0000-0002-8332-2309
Daniel Brunner 0000-0002-4003-3056
Bijie Bai 0000-0002-4744-996X
Aydogan Ozcan 0000-0002-0717-683X
Nataša Sladoje 0000-0002-6041-6310
Joakim Lindblad 0000-0001-7312-8222
Jason T Smith 0000-0001-6675-5252
Margarida Barroso 0000-0002-0407-3181
Li-Yu Yu 0000-0003-3986-8416
Yongtao Liu 0000-0003-0152-1783
Maxim A Ziatdinov 0000-0003-2570-4592
Uri Manor 0000-0002-9802-1955
Elias Nehme 0000-0003-1759-1751
Ofri Goldenberg 0000-0003-1242-5404
Yoav Shechtman 0000-0001-8498-5203
Barbora Špačková 0000-0002-8442-5641
Frank Cichos 0000-0002-9803-4975
Stefano Bo 0000-0002-2738-867X
Carlo Manzo 0000-0002-8625-0996
Harshith Bachimanchi 0000-0001-9497-8410
Erik Selander 0000-0002-2579-0841
Antoni Homs-Corbera 0000-0001-7889-273X
Caroline Beck Adiels 0000-0002-2403-2213
Mite Mijalkov 0000-0002-9844-4413
Dániel Veréb 0000-0003-2077-5252
Yu-Wei Chang 0000-0001-9598-2278
Joana B Pereira 0000-0002-4604-2711
Gustaf Kylberg 0000-0002-0055-4412
Beth A Cimini 0000-0001-9640-9318
Muyinatu A Lediju Bell 0000-0002-8394-4482
Bruno M Saraiva 0000-0002-9151-5477
Daniel Sage 0000-0002-1150-1623
Arrate Muñoz-Barrutia 0000-0002-1573-1661

References

[1] Tolles W E and Bostrom R C 1956 Automatic screening of cytological smears for cancer: the instrumentation *Ann. New York Acad. Sci.* **163** 1211–8
[2] Hounsfield G N 1979 Computed medical imaging *Nobel Lecture*
[3] Witte R, Andriasyan V, Georgi F, Yakimovich A and Greber U F 2018 Concepts in light microscopy of viruses *Viruses* **10** 202
[4] Weissenberger G, Henderikx R J M and Peters P J 2021 Understanding the invisible hands of sample preparation for cryo-EM *Nat. Methods* **18** 463–71
[5] Galimov E and Yakimovich A 2022 A tandem segmentation-classification approach for the localization of morphological predictors of C. elegans lifespan and motility *Aging* **14** 1665–77
[6] Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* pp 234–41
[7] Schmidt U, Weigert M, Broaddus C and Myers G 2018 Cell detection with star-convex polygons *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018* pp 265–73
[8] Weigert M *et al* 2018 Content-aware image restoration: pushing the limits of fluorescence microscopy *Nat. Methods* **15** 1090–7
[9] Nehme E, Weiss L E, Michaeli T and Shechtman Y 2018 Deep-STORM: super-resolution single-molecule microscopy by deep learning *Optica* **5** 458
[10] Wang H, Rivenson Y, Jin Y, Wei Z, Gao R, Günaydın H, Bentolila L A, Kural C and Ozcan A 2019 Deep learning enables cross-modality super-resolution in fluorescence microscopy *Nat. Methods* **16** 103–10
[11] Antun V, Renna F, Poon C, Adcock B and Hansen A C 2020 On instabilities of deep learning in image reconstruction and the potential costs of AI *Proc. Natl Acad. Sci.* **117** 30088–95
[12] Li R, Della Maggiora G, Andriasyan V, Petkidis A, Yushkevich A, Deshpande N, Kudryashev M and Yakimovich A 2024 Microscopy image reconstruction with physics-informed denoising diffusion probabilistic model *Commun. Eng.* **3** 186
[13] Gottschling N M, Antun V, Adcock B and Hansen A C 2020 The troublesome kernel: why deep learning for inverse problems is typically unstable (arXiv:2001.01258)
[14] Jain S B, Shao Z, Veettil S K T and Hecht M 2022 Adversarial attacks for machine learning denoisers and how to resist them *Proc. SPIE* **12204** 1220402

[15] Sitzmann V, Diamond S, Peng Y, Dun X, Boyd S, Heidrich W, Heide F and Wetzstein G 2018 End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging *ACM Trans. Graph.* **37** 1–13
[16] Sinha A, Lee J, Li S and Barbastathis G 2017 Lensless computational imaging through deep learning *Optica* **4** 1117–25
[17] Monakhova K, Yurtsever J, Kuo G, Antipa N, Yanny K and Waller L 2019 Learned reconstructions for practical mask-based lensless imaging *Opt. Express* **27** 28075–90
[18] Yanny K, Monakhova K, Shuai R W and Waller L 2022 Deep learning for fast spatially varying deconvolution *Optica* **9** 96–99
[19] Kellman M R, Bostan E, Repina N A and Waller L 2019 Physics-based learned design: optimized coded-illumination for quantitative phase imaging *IEEE Trans. Comput. Imaging* **5** 344–53
[20] Monakhova K, Tran V, Kuo G and Waller L 2021 Untrained networks for compressive lensless photography *Opt. Express* **29** 20913–29
[21] Batson J and Royer L 2019 Noise2self: blind denoising by self-supervision *Int. Conf. on Machine Learning* pp 524–33
[22] Monakhova K, Richter S R, Waller L and Koltun V 2022 Dancing under the stars: video denoising in starlight *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 16241–51
[23] Abdar M *et al* 2021 A review of uncertainty quantification in deep learning: techniques, applications and challenges *Inf. Fusion* **76** 243–97
[24] Sulam J, Papyan V, Romano Y and Elad M 2018 Multilayer convolutional sparse modeling: pursuit and dictionary learning *IEEE Trans. Signal Process.* **66** 4090–104
[25] Keller P J, Schmidt A D, Wittbrodt J and Stelzer E H 2008 Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy *Science* **322** 1065–9
[26] Reynaud E, Peychl J, Huisken J and Tomancak P 2015 Guide to light-sheet microscopy for adventurous biologists *Nat. Methods* **12** 30–34
[27] Kremer J, Stensbo-Smidt K, Gieseke F, Pedersen K S and Igel C 2017 Big universe, big data: machine learning and image analysis for astronomy *IEEE Intell. Syst.* **32** 16–22
[28] Rebollo-Neira L and Bowley J 2013 Sparse representation of astronomical images *J. Opt. Soc. Am. A* **30** 758–68
[29] Pratley L 2019 Radio astronomy image reconstruction in the big data era *PhD Thesis* University College London
[30] Korkut E H and Surer E 2023 Visualization in virtual reality: a systematic review *Virtual Real.* **27** 1447–80
[31] Pires F, Costa C and Dias P 2021 On the use of virtual reality for medical imaging visualization *J. Digit. Imaging* **34** 1034–48
[32] El Beheiry M, Doutreligne S, Caporal C, Ostertag C, Dahan M and Masson J-B 2019 Virtual reality: beyond visualization *J. Mol. Biol.* **431** 1315–21
[33] Günther U, Pietzsch T, Gupta A, Harrington K I S, Tomancak P, Gumhold S and Sbalzarini I F 2019 Scenery: flexible virtual reality visualization on the Java VM *IEEE Conf. on Visualization* pp 166–70
[34] Günther U, Harrington K I S, Dachselt R and Sbalzarini I F 2020 Bionic tracking: using eye tracking to track biological cells in virtual reality *European Conf. on Computer Vision—ECCV 2020 Workshops* vol 12535 pp 280–97
[35] Tiemann J, McGinity M, Sbalzarini I F and Günther U 2024 Live and interactive 3D photomanipulation under the microscope using virtual reality *CHI EA'24: Extended Abstracts of the 2024 CHI Conf. on Human Factors in Computing Systems, article 228* pp 1–7
[36] Segovia-Miranda F *et al* 2019 Three-dimensional spatially resolved geometrical and functional models of human liver tissue reveal new aspects of NAFLD progression *Nat. Med.* **25** 1885–93
[37] Maddu S, Cheeseman B L, Sbalzarini I F and Müller C L 2022 Stability selection enables robust learning of differential equations from limited noisy data *Proc. R. Soc. A* **478** 20210916
[38] Maddu S, Cheeseman B L, Müller C L and Sbalzarini I F 2021 Learning physically consistent differential equation models from data using group sparsity *Phys. Rev. E* **103** 042310
[39] Brunton S L, Proctor J L and Kutz J N 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems *Proc. Natl Acad. Sci. USA* **113** 3932–7
[40] Cheeseman B L, Günther U, Gonciarz K, Susik M and Sbalzarini I F 2018 Adaptive particle representation of fluorescence microscopy images *Nat. Commun.* **9** 5160
[41] Amat F, Myers E W and Keller P J 2012 Fast and robust optical flow for time-lapse microscopy using super-voxels *Bioinformatics* **29** 373–80
[42] Jonsson J, Cheeseman B L, Maddu S, Gonciarz K and Sbalzarini I F 2022 Parallel discrete convolutions on adaptive particle representations of images *IEEE Trans. Image Process.* **31** 4197–212
[43] Scholler J, Jonsson J, Jordá-Siquier T, Gantar I, Batti L, Cheeseman B, Pagès S, Sbalzarini I F and Lamy C M 2023 Efficient image analysis for large-scale next generation histopathology using pAPRica *bioxRiv Preprint* (https://doi.org/10.1101/2023.01.27.525687) (Posted 28 January 2023)
[44] Jonsson J, Cheeseman B L and Sbalzarini I F 2025 APR-CNN: convolutional neural networks for the adaptive particle representation of large microscopy images (Transactions on Machine Learning Research) 5qKI2dkrjL
[45] Mallat S G 1989 A theory for multiresolution signal decomposition: the wavelet representation *IEEE Trans. Pattern Anal. Mach. Intell.* **11** 674–93
[46] Pineda J *et al* 2022 Geometric deep learning reveals the spatiotemporal fingerprint of microscopic motion (arXiv:2202.06355)
[47] Kobayashi H, Cheveralls K C, Leonetti M D and Royer L A 2022 Self-supervised deep learning encodes high-resolution features of protein subcellular localization *Nat. Methods* **19** 995–1003
[48] Lotfollahi M *et al* 2023 Predicting cellular responses to complex perturbations in high-throughput screens *Mol. Syst. Biol.* **19** e11517
[49] Zhang X, Venkatachalapathy S, Paysan D, Schaerer P, Tripodo C, Uhler C and Shivashankar G V 2024 Unsupervised representation learning of chromatin images identifies changes in cell state and tissue organization in DCIS *Nat. Commun.* **15** 6112
[50] Helmuth J A, Paul G and Sbalzarini I F 2010 Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images *BMC Bioinform.* **11** 372
[51] Midtvedt B 2024 Annotation-free deep learning for quantitative microscopy *PhD Thesis* University of Gothenburg
[52] Ulman V, Oremus Z and Svoboda D 2015 TRAgen: a tool for generation of synthetic time-lapse image sequences of living cells *Int. Conf. on Image Analysis and Processing (ICAP)* pp 623–34
[53] Lange M *et al* 2024 A multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors *Cell* **187** 6742–59
[54] Bragantini J *et al* 2025 Ultrack: pushing the limits of cell tracking across biological scales *Nat Methods* **22** 2423–36

[55] Venkatakrishnan S V, Bouman C A and Wohlberg B 2013 Plug-and-play priors for model based reconstruction *2013 IEEE Global Conf. on Signal and Information Processing*
[56] Ahmad R, Bouman C A, Buzzard G T, Chan S, Liu S, Reehorst E T and Schniter P 2020 Plug-and-play methods for magnetic resonance imaging: using denoisers for image recovery *IEEE Signal Process. Mag.* **37** 105–16
[57] Gregor K and LeCun Y 2010 Learning fast approximations of sparse coding *Proc. 27th Int. Conf. on Machine Learning*
[58] Xie M, Liu J, Sun Y, Gan W, Wohlberg B and Kamilov U 2021 Joint reconstruction and calibration using regularization by denoising with application to computed tomography *Proc. IEEE/CVF Int. Conf. on Computer Vision Workshops* (*Montreal*)
[59] Gilton D, Ongie G and Willett R 2021 Model adaptation for inverse problems in imaging *IEEE Trans. Comput. Imaging* **7** 661–74
[60] Pellizzari C J, Spencer M F and Bouman C A 2020 Coherent plug-and-play: digital holographic imaging through atmospheric turbulence using model-based iterative reconstruction and convolutional neural networks *IEEE Trans. Comput. Imaging* **6** 1607–21
[61] Shastri S K, Ahmad R, Metzler C A and Schniter A P 2022 Denoising generalized expectation-consistent approximation for MR image recovery *IEEE J. Sel. Areas Inf. Theory* **3** 528–42
[62] Wei K, Aviles-Rivero A, Liang J, Fu Y, Schönlieb C-B and Huang H 2020 Tuning-free plug-and-play proximal algorithm for inverse imaging problems *Int. Conf. on Machine Learning* (PMLR)
[63] Kellman M, Zhang K, Markley E, Tamir J, Bostan E, Lustig M and Waller L 2020 Memory-efficient learning for large-scale computational imaging *IEEE Trans. Comput. Imaging* **6** 1403–14
[64] Liu J, Sun Y, Gan W, Xu X, Wohlberg B and Kamilov U S 2021 SGD-Net: efficient model-based deep learning with theoretical guarantees *IEEE Trans. Comput. Imaging* **7** 598–610
[65] Dholakia K, Reece P and Gu M 2008 Optical micromanipulation *Chem. Soc. Rev.* **37** 42–55
[66] Ashkin A, Dziedzic J M, Bjorkholm J E and Chu S 1986 Observation of a single-beam gradient force optical trap for dielectric particles *Opt. Lett.* **11** 288–90
[67] Polimeno P *et al* 2018 Optical tweezers and their applications *J. Quant. Spectrosc. Radiat. Transfer* **218** 131–50
[68] Bui A A M, Stilgoe A B, Lenton I C D, Gibson L J, Kashchuk A V, Zhang S, Rubinsztein-Dunlop H and Nieminen T A 2017 Theory and practice of simulation of optical tweezers *J. Quant. Spectrosc. Radiat. Transfer* **195** 66–75
[69] Hanna S and Morling T S 2022 Packing and dynamics of particle clusters in high order Bessel beams *Proc. Complex Light and Optical Forces XVI* PC120170N
[70] Lenton I C D, Volpe G, Stilgoe A B, Nieminen T A and Rubinsztein-Dunlop H 2020 Machine learning reveals complex behaviours in optically trapped particles *Mach. Learn.: Sci. Technol.* **1** 045009
[71] Ciriza D B, Magazzù A, Callegari A, Barbosa G, Neves A A R, Iatì M A, Volpe G and Maragò O M 2022 Faster and more accurate geometrical-optics optical force calculation using neural networks *ACS Photonics* **10** 234–41
[72] Helgadottir S, Argun A and Volpe G 2019 Digital video microscopy enhanced by deep learning *Optica* **6** 506–13
[73] Argun A, Thalheim T, Bo S, Cichos F and Volpe G 2020 Enhanced force-field calibration via machine learning *Appl. Phys. Rev.* **7** 041404
[74] Volpe G *et al* 2022 Roadmap for optical tweezers *J. Phys. Photon.* **5** 022501
[75] Farhat N H, Psaltis D, Prata A and Paek E 1985 Optical implementation of the Hopfield model *Appl. Opt.* **24** 1469
[76] Li H S, Qiao Y and Psaltis D 1993 Optical network for real-time face recognition *Appl. Opt.* **32** 5026
[77] Shen Y *et al* 2017 Deep learning with coherent nanophotonic circuits *Nat. Photon.* **11** 441
[78] Lin X, Rivenson Y, Yardimci N T, Veli M, Luo Y, Jarrahi M and Ozcan A 2018 All-optical machine learning using diffractive deep neural networks *Science* **361** 1004
[79] Rafayelyan M *et al* 2020 Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction *Phys. Rev. X* **10** 041037
[80] Zhou T, Lin X, Wu J, Chen Y, Xie H, Li Y, Fan J, Wu H, Fang L and Dai Q 2021 Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit *Nat. Photon.* **15** 367–73
[81] Moughames J, Porte X, Thiel M, Ulliac G, Larger L, Jacquot M, Kadic M and Brunner D 2020 Three-dimensional waveguide interconnects for scalable integration of photonic neural networks *Optica* **7** 640
[82] Porte X, Skalli A, Haghighi N, Reitzenstein S, Lott J A and Brunner D 2021 A complete, parallel and autonomous photonic neural network in a semiconductor multimode laser *J. Phys. Photon.* **3** 024017
[83] Mengu D and Ozcan A 2022 All-optical phase recovery: diffractive computing for quantitative phase imaging *Adv. Opt. Mater.* **10** 2200281
[84] Sakib Rahman M S and Ozcan A 2021 Computer-free, all-optical reconstruction of holograms using diffractive networks *ACS Photonics* **8** 3375–84
[85] Luo Y, Zhao Y, Li J, Çetintaş E, Rivenson Y, Jarrahi M and Ozcan A 2022 Computational imaging without a computer: seeing through random diffusers at the speed of light *eLight* **2** 4
[86] Bai B, Luo Y, Gan T, Hu J, Li Y, Zhao Y, Mengu D, Jarrahi M and Ozcan A 2022 To image, or not to image: class-specific diffractive cameras with all-optical erasure of undesired objects *eLight* **2** 14
[87] Porte X, Dinc N U, Moughames J, Panusa G, Juliano C, Kadic M, Moser C, Brunner D and Psaltis D 2021 Direct (3+ 1) D laser writing of graded-index optical elements *Optica* **8** 1281
[88] Grabulosa A *et al* (3+ 1) D-printed adiabatic 1-to-N broadband couplers (arXiv:2208.09767)
[89] Momeni A *et al* Training of physical neural networks (arXiv:2406.03372)
[90] Young G *et al* 2018 Quantitative mass imaging of single biological macromolecules *Science* **360** 423–7
[91] Altman L E and Grier D G 2020 CATCH: characterizing and tracking colloids holographically using deep neural networks *J. Phys. Chem. B* **124** 1602–10
[92] Midtvedt B, Olsén E, Eklund F, Höök F, Adiels C B, Volpe G and Midtvedt D 2021 Fast and accurate nanoparticle characterization using deep-learning-enhanced off-axis holography *ACS Nano* **15** 2240–50
[93] Manifold B, Men S, Hu R and Fu D 2021 A versatile deep learning architecture for classification and label-free prediction of hyperspectral images *Nat. Mach. Intell.* **3** 306–15
[94] Midtvedt B, Helgadottir S, Argun A, Pineda J, Midtvedt D and Volpe G 2021 Quantitative digital microscopy with deep learning *Appl. Phys. Rev.* **8** 011310
[95] Bohren C F and Huffman D R 2007 Rayleigh-Gans theory *Absorption and Scattering of Light by Small Particles*
[96] Lin H *et al* 2021 Microsecond fingerprint stimulated Raman spectroscopic imaging by ultrafast tuning and spatial-spectral learning *Nat. Commun.* **12** 3052

[97] Remer I, Shaashoua R, Shemesh N, Ben-Zvi A and Bilenca A 2020 High-sensitivity and high-specificity biomechanical imaging by stimulated Brillouin scattering microscopy *Nat. Methods* **17** 913–6
[98] Helgadottir S, Midtvedt B, Pineda J, Sabirsh A, Adiels C B, Romeo S, Midtvedt D and Volpe G 2021 Extracting quantitative biological information from bright-field cell images using deep learning *Biophys. Rev.* **2** 031401
[99] Rivenson Y, Zhang Y, Günaydın H, Teng D and Ozcan A 2018 Phase recovery and holographic image reconstruction using deep learning in neural networks *Light Sci. Appl.* **7** 17141
[100] Deng M, Li S, Goy A, Kang I and Barbastathis G 2020 Learning to synthesize: robust phase retrieval at low photon counts *Light Sci. Appl.* **9** 36
[101] Wang F, Bian Y, Wang H, Lyu M, Pedrini G, Osten W, Barbastathis G and Situ G 2020 Phase imaging with an untrained neural network *Light Sci. Appl.* **9** 77
[102] Bostan E, Heckel R, Chen M, Kellman M and Waller L 2020 Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network *Optica* **7** 559–62
[103] Xue Y, Cheng S, Li Y and Tian L 2019 Reliable deep-learning-based phase imaging with uncertainty quantification *Optica* **6** 618–29
[104] Goy A, Rughoobur G, Li S, Arthur K, Akinwande A I and Barbastathis G 2019 High-resolution limited-angle phase tomography of dense layered objects using deep neural networks *Proc. Natl Acad. Sci.* **116** 19848–56
[105] Matlock A, Zhu J and Tian L 2023 Multiple-scattering simulator-trained neural network for intensity diffraction tomography *Opt. Express* **31** 4094–107
[106] Kang I, Goy A and Barbastathis G 2021 Dynamical machine learning volumetric reconstruction of objects' interiors from limited angular views *Light Sci. Appl.* **10** 74
[107] Liu R, Sun Y, Zhu J, Tian L and Kamilov U 2022 Recovery of continuous 3D refractive index maps from discrete intensity-only measurements using neural fields *Nat. Mach. Intell.* **4** 781–91
[108] Cao R, Divekar N S, Nuñez J K, Upadhyayula S and Waller L 2024 Neural space–time model for dynamic multi-shot imaging *Nat. Methods* **21** 2336–41
[109] Walter A, Mannheim J G and Caruana C J (ed) 2021 *Imaging Modalities for Biological and Preclinical Research: A Compendium* vol 1 (IOP Publishing) (https://doi.org/10.1088/978-0-7503-3059-6)
Walter A, Mannheim J G and Caruana C J (ed) 2021 *Imaging Modalities for Biological and Preclinical Research: A Compendium* vol 2 (IOP Publishing) (https://doi.org/10.1088/978-0-7503-3747-2)
[110] Walter A *et al* 2020 Correlated multimodal imaging in life sciences: expanding the biomedical horizon *Front. Phys.* **8** 47
[111] Lei Y, Wang T, Curran W J, Liu T and Yang X 2020 Deep learning in medical image registration: a review *Phys. Med. Biol.* **65** 20TR01
[112] Boveiri H R, Khayami R, Javidan R and Mehdizadeh A 2020 Medical image registration using deep neural networks: a comprehensive review *Comput. Electr. Eng.* **87** 106767
[113] Haskins G, Kruger U and Yan P 2020 Deep learning in medical image registration: a survey *Mach. Vis. Appl.* **31** 8
[114] Pielawski N, Wetzer E, Öfverstedt J, Lu J, Wählby C, Lindblad J and Sladoje N 2020 CoMIR: contrastive multimodal image representation for registration *Advances in Neural Information Processing Systems* vol 33 (NeurIPS) pp 18433–44
[115] Lu J, Öfverstedt J, Lindblad J and Sladoje N 2022 Is image-to-image translation the panacea for multimodal image registration? A comparative study *PLoS One* **17** e0276196
[116] Öfverstedt J, Lindblad J and Sladoje N 2022 Fast computation of mutual information in the frequency domain with applications to global multimodal image alignment *Pattern Recognit. Lett.* **159** 196–203
[117] Paul-Gilloteaux P, Heiligenstein X, Belle M, Domart M-C, Larijani B, Collinson L, Raposo G and Salamero J 2017 eC-CLEM: flexible multidimensional registration software for correlative microscopies *Nat. Methods* **14** 102–3
[118] Wu G, Nowotny T, Zhang Y, Yu H-Q and Li D D-U 2016 Artificial neural network approaches for fluorescence lifetime imaging techniques *Opt. Lett.* **41** 2561
[119] Smith J T, Yao R, Sinsuebphon N, Rudkouskaya A, Un N, Mazurkiewicz J, Barroso M, Yan P and Intes X 2019 Fast fit-free analysis of fluorescence lifetime imaging via deep learning *Proc. Natl Acad. Sci. USA* **116** 24019–30
[120] Zickus V *et al* 2020 Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation *Sci. Rep.* **10** 20986
[121] Xiao D, Sapermsap N, Chen Y and Li D D U 2023 Deep learning enhanced fast fluorescence lifetime imaging with a few photons *Optica* **10** 944–51
[122] Kapsiani S *et al* 2025 Deep learning for fluorescence lifetime predictions enables high-throughput *in vivo* imaging *bioRxiv Preprint* https://doi.org/10.1101/2025.02.20.639036 (Posted 26 February 2025)
[123] Shi J *et al* 2024 Self-supervised deep learning enables robust fluorescence lifetime estimation with limited photons *Proc. SPIE* PC1285719
[124] Smith J *et al* 2022 *In vitro* and *in vivo* NIR fluorescence lifetime imaging with a time-gated SPAD camera *Optica* **9** 532–44
[125] Ochoa M, Rudkouskaya A, Yao R, Yan P, Barroso M and Intes X 2020 High compression deep learning based single-pixel hyperspectral macroscopic fluorescence lifetime imaging in vivo *Biomed. Opt. Express* **11** 5401
[126] Smith J T, Ochoa M and Intes X 2020 UNMIX-ME: spectral and lifetime fluorescence unmixing via deep learning *Biomed. Opt. Express* **11** 3857
[127] Chen M, Jin L, Xuan X, Yang D, Cheng Y and Zhang J 2025 Interpretable deep learning illuminates multiple structures fluorescence imaging: a path toward trustworthy artificial intelligence in microscopy (arXiv:2501.05490)
[128] Chen Y-I *et al* 2022 Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells *Commun. Biol.* **5** 18
[129] Zang Z, Xiao D, Wang Q, Li Z, Xie W, Chen Y and Li D D U 2022 Fast fluorescence lifetime imaging analysis via extreme learning machine (arXiv:2203.13754)
[130] Erbas I, Pandey V, Amarnath A, Wang N, Swaminathan K and Radev S T 2024 Intes X. Compressing recurrent neural networks for FPGA-accelerated implementation in fluorescence lifetime imaging (arXiv:2410.00948)
[131] Erbas I, Amarnath A, Pandey V, Swaminathan K, Wang N and Intes X 2024 Unlocking real-time fluorescence lifetime imaging: multi-pixel parallelism for FPGA-accelerated processing (arXiv:2410.07364)
[132] Pandey V, Erbas I, Michalet X, Ulku A, Bruschini C, Charbon E, Barroso M and Intes X 2024 Deep learning-based temporal deconvolution for photon time-of-flight distribution retrieval *Opt. Lett.* **49** 6457–60

[133] Cahill L C, Giacomelli M G, Yoshitake T, Vardeh H, Faulkner-Jones B E, Connolly J L, Sun C-K and Fujimoto J G 2018 Rapid virtual hematoxylin and eosin histology of breast tissue specimens using a compact fluorescence nonlinear microscope *Lab. Invest.* **98** 150–60

[134] Walsh A J, Mueller K P, Tweed K, Jones I, Walsh C M, Piscopo N J, Niemi N M, Pagliarini D J, Saha K and Skala M C 2020 Classification of T-cell activation via autofluorescence lifetime imaging *Nat. Biomed. Eng.* **5** 77–88

[135] Zheng C, Park J K, Yildirim M, Boivin J R, Xue Y, Sur M, So P T C and Wadduwage D N 2021 De-scattering with excitation patterning enables rapid wide-field imaging through scattering media *Sci. Adv.* **7** eaay5496

[136] Guan H *et al* 2022 Deep-learning two-photon fiberscopy for video-rate brain imaging in freely-behaving mice *Nat. Commun.* **13** 1–9

[137] McAleer S, Fast A, Xue Y, Seiler M J, Tang W C, Balu M, Baldi P and Browne A W 2021 Deep learning–assisted multiphoton microscopy to reduce light exposure and expedite imaging in tissues with high and low light sensitivity *Transl. Vis. Sci. Technol.* **10** 30

[138] Fan W, Chen T, Gil E, Zhu S, Yakovlev V, Wang D W and Zhang D 2021 Second harmonic imaging enhanced by deep learning decipher *ACS Photonics* **8** 1562–8

[139] Han J, Liu K, Isaacson K B, Monakhova K, Griffith L G and You S 2025 System- and sample-agnostic isotropic three-dimensional microscopy by weakly physics-informed, domain-shift-resistant axial deblurring *Nat. Commun.* **16** 745

[140] Rivenson Y *et al* 2019 Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning *Nat. Biomed. Eng.* **3** 466–77

[141] Sun Y, You S, Du X, Spaulding A, George Liu Z, Chaney E J, Spillman D R, Marjanovic M, Tu H and Boppart S A 2020 Real-time three-dimensional histology-like imaging by label-free nonlinear optical microscopy *Quant. Imaging Med. Surg.* **10** 2177

[142] Bower A J, Psaltis D, Borhani N and Boppart S A 2019 Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy *Biomed. Opt. Express* **10** 1339–50

[143] You S, Chaney E J, Tu H, Sun Y, Sinha S and Boppart S A 2021 Label-free deep profiling of the tumor microenvironment *Cancer Res.* **81** 2534–44

[144] Sarti M *et al* 2022 Deep learning for automated analysis of cellular and extracellular components of the foreign body response in multiphoton microscopy images *Front. Bioeng. Biotechnol.* **9** 1497

[145] Hollon T C *et al* 2020 Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks *Nat. Med.* **26** 52–58

[146] Huttunen M J, Hassan A, McCloskey C W, Fasih S, Upham J, Vanderhyden B C, Boyd R W and Murugkar S 2018 Automated classification of multiphoton microscopy images of ovarian tissue using deep learning *Proc. SPIE* **23** 066002

[147] Shi J, Tu H, Park J, Marjanovic M, Higham A M, Luckey N N, Cradock K A, Liu Z G and Boppart S A 2022 Weakly supervised deep learning for human breast cancer-related optical biomarker discovery in label-free virtual histopathology *bioRxiv Preprint* https://doi.org/10.1101/2022.05.24.493356 (Posted 17 November 2022)

[148] You S, Sun Y, Yang L, Park J, Tu H, Marjanovic M, Sinha S and Boppart S A 2019 Real-time intraoperative diagnosis by deep neural network driven multiphoton virtual histology *npj Precis. Oncol.* **3** 1–8

[149] Binnig G and Rohrer H 1983 Scanning tunneling microscopy *Surf. Sci.* **126** 236–44

[150] Binnig G, Quate C F and Gerber C 1986 Atomic force microscope *Phys. Rev. Lett.* **56** 930

[151] Gerber C and Lang H P 2006 How the doors to the nanoworld were opened *Nat. Nanotechnol.* **1** 3–5

[152] Vasudevan R K, Jesse S, Kim Y, Kumar A and Kalinin S V 2012 Spectroscopic imaging in piezoresponse force microscopy: new opportunities for studying polarization dynamics in ferroelectrics and multiferroics *MRS Commun.* **2** 61–73

[153] Balke N, Jesse S, Kim Y, Adamczyk L, Tselev A, Ivanov I N, Dudney N J and Kalinin S V 2010 Real space mapping of Li-ion transport in amorphous Si anodes with nanometer resolution *Nano Lett.* **10** 3420–5

[154] Kannan R, Ievlev A V, Laanait N, Ziatdinov M A, Vasudevan R K, Jesse S and Kalinin S V 2018 Deep data analysis via physically constrained linear unmixing: universal framework, domain examples, and a community-wide platform *Adv. Struct. Chem. Imaging* **4** 20

[155] Kelley K P, Ziatdinov M, Collins L, Susner M A, Vasudevan R K, Balke N, Kalinin S V and Jesse S 2020 Fast scanning probe microscopy via machine learning: non-rectangular scans with compressed sensing and Gaussian process optimization *Small* **16** 2002878

[156] Liu Y *et al* 2022 Exploring physics of ferroelectric domain walls in real time: deep learning enabled scanning probe microscopy *Adv. Sci.* **9** 2203957

[157] Liu Y, Kelley K P, Vasudevan R K, Funakubo H, Ziatdinov M A and Kalinin S V 2022 Experimental discovery of structure–property relationships in ferroelectric materials via active learning *Nat. Mach. Intell.* **4** 341–50

[158] Fang L *et al* 2021 Deep learning-based point-scanning super-resolution imaging *Nat. Methods* **18** 406–16

[159] Buchholz T-O, Krull A, Shahidi R, Pigino G, Jékely G and Jug F 2019 Chapter 13—Content-aware image restoration for electron microscopy T Müller-Reichert and G Pigino ed *Methods in Cell Biology* vol 152 (Academic Press) pp 277–89

[160] Johnson J, Alahi A and Fei-Fei L 2016 Perceptual losses for real-time style transfer and super-resolution (arXiv:1603.08155)

[161] Krull A, Buchholz T-O and Jug F 2019 Noise2Void—learning denoising from single noisy images (arXiv:1811.10980)

[162] Broaddus C, Krull A, Weigert M, Schmidt U and Myers G 2020 Removing structured noise with self-supervised blind-spot networks *2020 IEEE 17th Int. Symp. on Biomedical Imaging (ISBI)* pp 159–63

[163] Prakash M, Delbracio M, Milanfar P and Jug F 2022 Interpretable unsupervised diversity denoising and artefact removal *Int. Conf. on Learning Representations* (*1 November 2022*)

[164] von Chamier L *et al* 2021 Democratising deep learning for microscopy with ZeroCostDL4Mic *Nat. Commun.* **12** 2276

[165] Prigent S, Valades-Cruz C A, Leconte L, Maury L, Salamero J and Kervrann C 2022 BioImageIT: open-source framework for integration of image data management with analysis *Nat. Methods* **19** 1328–30

[166] Lugmayr A, Danelljan M and Timofte R 2019 Unsupervised learning for real-world super-resolution (arXiv:1909.09629)

[167] Stites H C and Manor U 2025 PSSR2: a user-friendly Python package for democratizing deep learning-based point-scanning super-resolution microscopy *BMC Methods* **2** 1

[168] Lelek M, Gyparaki M T, Beliu G, Schueder F, Griffié J, Manley S, Jungmann R, Sauer M, Lakadamyali M and Zimmer C 2021 Single-molecule localization microscopy *Nat. Rev. Methods Primers* **1** 39

[169] Von Diezmann L, Shechtman Y and Moerner W E 2017 Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking *Chem. Rev.* **117** 7244–75

[170] Möckl L, Roy A R and Moerner W E 2020 Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments [Invited] *Biomed. Opt. Express* **11** 1633

[171] Hershko E, Weiss L E, Michaeli T and Shechtman Y 2019 Multicolor localization microscopy and point-spread-function engineering by deep learning *Opt. Express* **27** 6158
[172] Nehme E, Freedman D, Gordon R, Ferdman B, Weiss L E, Alalouf O, Naor T, Orange R, Michaeli T and Shechtman Y 2020 DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning *Nat. Methods* **17** 734–40
[173] Nehme E, Ferdman B, Weiss L E, Naor T, Freedman D, Michaeli T and Shechtman Y 2021 Learning optimal wavefront shaping for multi-channel imaging *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 2179–92
[174] Barth R, Fourel G and Shaban H A 2020 Dynamics as a cause for the nanoscale organization of the genome *Nucleus* **11** 83–98
[175] Naor T, Nogin Y, Nehme E, Ferdman B, Weiss L E, Alalouf O and Shechtman Y 2022 Quantifying cell-cycle-dependent chromatin dynamics during interphase by live 3D tracking *iScience* **25** 104197
[176] Ouyang W, Aristov A, Lelek M, Hao X and Zimmer C 2018 Deep learning massively accelerates super-resolution localization microscopy *Nat. Biotechnol.* **36** 460–8
[177] Speiser A *et al* 2021 Deep learning enables fast and dense single-molecule localization with high accuracy *Nat. Methods* **18** 9
[178] Kim T, Moon S and Xu K 2019 Information-rich localization microscopy through machine learning *Nat. Commun.* **10** 1996
[179] Boyd N, Jonas E, Babcock H and Recht B 2018 DeepLoco: fast 3D localization microscopy using neural networks (https://doi.org/10.1101/267096)
[180] Zhang P, Liu S, Chaurasia A, Ma D, Mlodzianoski M J, Culurciello E and Huang F 2018 Analyzing complex single-molecule emission patterns with deep learning *Nat. Methods* **15** 913–6
[181] Wu T, Lu P, Rahman M A, Li X and Lew M D 2022 Deep-SMOLM: deep learning resolves the 3D orientations and 2D positions of overlapping single molecules with optimal nanoscale resolution *Opt. Express* **30** 36761
[182] Jouchet P, Roy A R and Moerner W E 2023 Combining deep learning approaches and point spread function engineering for simultaneous 3D position and 3D orientation measurements of fluorescent single molecules *Opt. Commun.* **542** 129589
[183] Möckl L, Roy A R, Petrov P N and Moerner W E 2020 Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network BGnet *Proc. Natl Acad. Sci. USA* **117** 60–67
[184] Goldenberg O, Ferdman B, Nehme E, Ezra Y S and Shechtman Y 2022 Learning optimal multicolor PSF design for 3D pairwise distance estimation *Intell. Comput.* **2022** 0004
[185] Ikoma H, Kudo T, Peng Y, Broxton M and Wetzstein G 2021 Deep learning multi-shot 3D localization microscopy using hybrid optical–electronic computing *Opt. Lett.* **46** 6023
[186] Saguy A, Alalouf O, Opatovski N, Jang S, Heilemann M and Shechtman Y 2023 DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning *Nat. Methods* **20** 1939–48
[187] Xiao D, Kedem Orange R, Opatovski N, Parizat A, Nehme E, Alalouf O and Shechtman Y 2024 Large-FOV 3D localization microscopy by spatially variant point spread function generation *Sci. Adv.* **10** eadj3656
[188] Fu S *et al* 2023 Field-dependent deep learning enables high-throughput whole-cell 3D super-resolution imaging *Nat. Methods* **20** 459–68
[189] Zhou Z, Wu J, Wang Z and Huang Z-L 2023 Deep learning using a residual deconvolutional network enables real-time high-density single-molecule localization microscopy *Biomed. Opt. Express* **14** 1833
[190] Abdehkakha A, Madani S M N and Snoeyink C 2025 Enhancing single-molecule localization microscopy with deep learning *Appl. Opt.* **64** A19
[191] Zhang P *et al* 2023 Deep learning-driven adaptive optics for single-molecule localization microscopy *Nat. Methods* **20** 1748–58
[192] Sage D *et al* 2019 Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software *Nat. Methods* **16** 387–95
[193] Miller H, Zhou Z, Shepherd J, Wollman A J M and Leake M C 2018 Single-molecule techniques in biophysics: a review of progress in methods and applications *Rep. Prog. Phys.* **81** 024601
[194] Taylor R W and Sandoghdar V 2019 Interferometric scattering microscopy: seeing single nanoparticles and molecules via Rayleigh scattering *Nano Lett.* **19** 4827–35
[195] Špačková B *et al* 2022 Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles *Nat. Methods* **19** 751–8
[196] Midtvedt B *et al* 2022 Single-shot self-supervised object detection in microscopy *Nat. Commun.* **13** 7492
[197] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2023 Attention is all you need
[198] Bronstein M M, Bruna J, LeCun Y, Szlam A and Vandergheynst P 2017 Geometric deep learning: going beyond Euclidean data *IEEE Signal Process. Mag.* **34** 18–42
[199] Pineda J, Midtvedt B, Bachimanchi H, Noé S, Midtvedt D, Volpe G and Manzo C 2023 Geometric deep learning reveals the spatiotemporal fingerprint of microscopic motion *Nat. Mach. Intell.* **5** 71–82
[200] Xu Y, Shinomiya M and Harada A 2016 Soft matter-regulated active nanovalves locally self-assembled in femtoliter nanofluidic channels *Adv. Mater.* **28** 2209–16
[201] Krishnan M, Mojarad N, Kukura P and Sandoghdar V 2010 Geometry-induced electrostatic trapping of nanometric objects in a fluid *Nature* **467** 692–5
[202] Xu Y 2018 Nanofluidics: a new arena for materials science *Adv. Mater.* **30** 1702419
[203] Perrin J 1910 Mouvement brownien et molécules *J. Phys. Theor. Appl.* **9** 5–39
[204] Crocker J C and Grier D G 1996 Methods of digital video microscopy for colloidal studies *J. Colloid Interface Sci.* **179** 298–310
[205] Parthasarathy R 2012 Rapid, accurate particle tracking by calculation of radial symmetry centers *Nat. Methods* **9** 724–6
[206] Shorten C and Khoshgoftaar T M 2019 A survey on image data augmentation for deep learning *J. Big Data* **6** 1–48
[207] Ershov D *et al* 2021 Bringing TrackMate into the era of machine-learning and deep-learning *bioRxiv Preprint* https://doi.org/10.1101/2021.09.03.458852 (Posted 20 September 2021)
[208] Midtvedt B *et al* 2022 Single-shot self-supervised particle tracking (arXiv:2202.13546)
[209] Ulman V *et al* 2017 An objective comparison of cell-tracking algorithms *Nat. Methods* **14** 1141–52
[210] Dosovitskiy A *et al* 2021 An image is worth 16x16 words: transformers for image recognition at scale (arXiv:2010.11929)
[211] Jones P H, Maragò O M and Volpe G 2015 *Optical Tweezers: Principles and Applications* (Cambridge University Press)
[212] Gavrilov M, Jun Y and Bechhoefer J 2014 Real-time calibration of a feedback trap *Rev. Sci. Instrum.* **85** 095102
[213] García L P, Pérez J D, Volpe G, Arzola A V and Volpe G 2018 High-performance reconstruction of microscopic force fields from Brownian trajectories *Nat. Commun.* **9** 1–9
[214] Wu P, Huang R, Tischer C, Jonas A and Florin E L 2009 Direct measurement of the nonconservative force field generated by optical tweezers *Phys. Rev. Lett.* **103** 108101

[215] Braun M, Bregulla A P, Günther K, Mertig M and Cichos F 2015 Single molecules trapped by dynamic inhomogeneous temperature fields *Nano Lett.* **15** 5499–505
[216] Su X, Fischer A and Cichos F 2021 Towards measuring the Maxwell—Boltzmann distribution of a single heated particle *Front. Phys.* **9** 669459
[217] Argun A, Thalheim T, Bo S, Cichos F and Volpe G 2020 DeepCalib (available at: https://github.com/softmatterlab/DeepCalib)
[218] Argun A, Volpe G and Bo S 2021 Classification, inference and segmentation of anomalous diffusion with recurrent neural networks *J. Phys. A* **54** 294003
[219] Chenouard N *et al* 2014 Objective comparison of particle tracking methods *Nat. Methods* **11** 281–9
[220] Vestergaard C L, Blainey P C and Flyvbjerg H 2014 Optimal estimation of diffusion coefficients from single-particle trajectories *Phys. Rev. E* **89** 022726
[221] Muñoz-Gil G *et al* 2021 Objective comparison of methods to decode anomalous diffusion *Nat. Commun.* **12** 1
[222] Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y and Shechtman Y 2019 Single-particle diffusion characterization by deep learning *Biophys. J.* **117** 185–92
[223] Bo S, Schmidt F, Eichhorn R and Volpe G 2019 Measurement of anomalous diffusion using recurrent neural networks *Phys. Rev. E* **100** 010102
[224] Jamali V, Hargus C, Ben-Moshe A, Aghazadeh A, Ha H D, Mandadapu K K and Alivisatos A P 2021 Anomalous nanoparticle surface diffusion in LCTEM is revealed by deep learning-assisted analysis *Proc. Natl Acad. Sci.* **118** e2017616118
[225] Pacheco-Pozo A and Krapf D 2024 Effectively detecting anomalous diffusion via deep learning *Nat. Comput. Sci.* **4** 731–2
[226] Samek W, Montavon G, Lapuschkin S, Anders C J and Müller K-R 2021 Explaining deep neural networks and beyond: a review of methods and applications *Proc. IEEE* **109** 247–78
[227] Vaswani A *et al* 2017 Attention is all you need *Advances in Neural Information Processing Systems* vol 30
[228] Fernández-Fernández G, Manzo C, Lewenstein M, Dauphin A and Muñoz-Gil G 2024 Learning minimal representations of stochastic processes with variational autoencoders *Phys. Rev. E* **110** L012102
[229] Muñoz-Gil G, Guigo I Corominas G and Lewenstein M 2021 Unsupervised learning of anomalous diffusion data: an anomaly detection approach *J. Phys. A: Math. Theor.* **54** 504001
[230] Zhang Y, Zhu J, Xie H and He Y 2025 Physics-informed deep learning for stochastic particle dynamics estimation *Proc. Natl Acad. Sci.* **122** e2418643122
[231] Brückner D B, Arlt N, Fink A, Ronceray P, Rädler J O and Broedersz C P 2021 Learning the dynamics of cell–cell interactions in confined cell migration *Proc. Natl Acad. Sci.* **118** e2016602118
[232] Zhang K, Zhu Y, Wang J and Zhang J 2020 Adaptive structural fingerprints for graph attention networks *Int. Conf. on Learning Representations*
[233] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O and Hassabis D 2021 Highly accurate protein structure prediction with AlphaFold *Nature* **596** 583–9
[234] Rudin C 2019 Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead *Nat. Mach. Intell.* **1** 206–15
[235] Midtvedt B, Helgadottir S, Argun A, Pineda J, Midtvedt D and Volpe G DeepTrack 2 (available at: https://github.com/softmatterlab/DeepTrack2) (Accessed February 2023)
[236] Yang Y, Qiu J, Song M, Tao D and Wang X 2020 Distilling knowledge from graph convolutional networks *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 7074–83
[237] Tailor S A, Opolka F, Lio P and Lane N D 2021 Do we need anisotropic graph neural networks? *Int. Conf. on Learning Representations*
[238] Manzo C and Garcia-Parajo M F 2015 A review of progress in single particle tracking: from methods to biophysical insights *Rep. Prog. Phys.* **78** 124601
[239] Newby J M, Schaefer A M, Lee P T, Forest M G and Lai S K 2018 Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D *Proc. Natl Acad. Sci. USA* **115** 9026–31
[240] Kowalek P, Loch-Olszewska H and Szwabiński J 2019 Classification of diffusion modes in single-particle tracking data: feature-based versus deep-learning approach *Phys. Rev. E* **100** 32410
[241] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 Single trajectory characterization via machine learning *New J. Phys.* **22** 013010
[242] Muñoz-Gil G *et al* 2025 Quantitative evaluation of methods to analyze motion changes in single-particle experiments *Nat. Commun.* **16** 6749
[243] Nisbet E G and Sleep N H 2001 The habitat and nature of early life *Nature* **409** 1083–91
[244] Lyons T W, Reinhard C T and Planavsky N J 2014 The rise of oxygen in Earth's early ocean and atmosphere *Nature* **506** 307–15
[245] Field C B, Behrenfeld M J, Randerson J T and Falkowski P 1998 Primary production of the biosphere: integrating terrestrial and oceanic components *Science* **281** 237–40
[246] Sunagawa S *et al* 2020 Tara Oceans: towards global ocean ecosystems biology *Nat. Rev. Microbiol.* **18** 8
[247] Ma P, Li C, Rahaman M M, Yao Y, Zhang J, Zou S, Zhao X and Grzegorzek M 2022 A state-of-the-art survey of object detection techniques in microorganism image analysis: from classical methods to deep learning approaches *Artif. Intell. Rev.* **56** 1627–98
[248] Li Y, Guo J, Guo X, Hu Z and Tian Y 2021 Plankton detection with adversarial learning and a densely connected deep learning model for class imbalanced distribution *J. Mar. Sci. Eng.* **9** 636
[249] Bachimanchi H, Midtvedt B, Midtvedt D, Selander E and Volpe G 2022 Microplankton life histories revealed by holographic microscopy and deep learning *eLife* **11** e79760
[250] Caron M *et al* 2021 Emerging properties in self-supervised vision transformers (arXiv:2104.14294)
[251] Zarrintaj P, Saeb M R, Stadler F J, Yazdi M K, Nezhad M N, Mohebbi S, Seidi F, Ganjali M R and Mozafari M 2022 Human organs-on-chips: a review of the state-of-the-art, current prospects, and future challenges *Adv. Biol.* **6** 2000526
[252] Lee S, Ko J, Park D, Lee S-R, Chung M, Lee Y and Jeon N L 2018 Microfluidic-based vascularized microphysiological systems *Lab Chip* **18** 2686–709
[253] Roth A 2021 Human microphysiological systems for drug development *Science* **373** 1304–6
[254] Young A T, Rivera K R, Erb P D and Daniele M A 2019 Monitoring of microphysiological systems: integrating sensors and real-time data analysis toward autonomous decision-making *ACS Sens.* **4** 1454–64
[255] Kavand H, Nasiri R and Herland A 2022 Advanced materials and sensors for microphysiological systems: focus on electronic and electrooptical interfaces *Adv. Mater.* **34** 2107876
[256] Peel S and Jackman M 2021 Imaging microphysiological systems: a review *Am. J. Physiol. Cell Physiol.* **320** C669–C80

[257] Li J, Chen J, Bai H, Wang H, Hao S, Ding Y, Peng B, Zhang J, Li L and Huang W 2022 An overview of organs-on-chips based on deep learning *AAAS Res.* **2022** 9869518
[258] Youhanna S *et al* 2022 Organotypic and microphysiological human tissue models for drug discovery and development—current state-of-the-art and future perspectives *Pharmacol. Rev.* **74** 141–206
[259] Ewart L *et al* 2018 Application of microphysiological systems to enhance safety assessment in drug discovery *Annu. Rev. Pharmacol. Toxicol.* **58** 65–82
[260] Kim J *et al* 2020 Three-dimensional human liver-chip emulating premetastatic niche formation by breast cancer-derived extracellular vesicles *ACS Nano* **14** 14971–88
[261] Galan E A, Zhao H, Wang X, Dai Q, Huck W T S and Ma S 2020 Intelligent microfluidics: the convergence of machine learning and microfluidics in materials science and biomedicine *Matter* **3** 1893–922
[262] Würger A 2010 Thermal non-equilibrium transport in colloids *Rep. Prog. Phys.* **73** 2010
[263] Fränzl M and Cichos F 2022 Hydrodynamic manipulation of nano-objects by optically induced thermo-osmotic flows *Nat. Commun.* **13** 656
[264] Weinert F M, Kraus J A, Franosch T and Braun D 2008 Microscale fluid flow induced by thermoviscous expansion along a traveling wave *Phys. Rev. Lett.* **100** 164501
[265] Lin L, Peng X, Mao Z, Wei X, Xie C and Zheng Y 2017 Interfacial-entropy-driven thermophoretic tweezers *Lab Chip* **17** 3061–70
[266] Fränzl M, Thalheim T, Adler J, Huster D, Posseckardt J, Mertig M and Cichos F 2019 Thermophoretic trap for single amyloid fibril and protein aggregation studies *Nat. Methods* **16** 611–4
[267] Braun M and Cichos F 2013 Optically controlled thermophoretic trapping of single nano-objects *ACS Nano* **7** 11200–8
[268] Fränzl M and Cichos F 2020 Active particle feedback control with a single-shot detection convolutional neural network *Sci. Rep.* **10** 12571
[269] Muiños-Landin S, Fischer A, Holubec V and Cichos F 2021 Reinforcement learning with artificial microswimmers *Sci. Robot.* **6** eabd9285
[270] Bengtsson E, Danielsen H, Treanor D, Gurcan M N, MacAulay C and Molnár B 2017 Computer-aided diagnostics in digital pathology *Cytometry A* **91** 551–4
[271] Kartasalo K *et al* 2021 Artificial intelligence for diagnosis and Gleason grading of prostate cancer in biopsies—current status and next steps *Eur. Urol. Focus* **7** 687–91
[272] Bulten W *et al* (the PANDA challenge consortium) 2022 Gleason grading of prostate cancer: the PANDA challenge *Nat. Med.* **28** 154–63
[273] Ström P *et al* 2020 Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study *Lancet Oncol.* **21** 222–32
[274] Bejnordi B E, Litjens G, Timofeeva N, Otte-Holler I, Homeyer A, Karssemeijer N and van der Laak J A 2015 Stain specific standardization of whole-slide histopathological images *IEEE Trans. Med. Imaging* **35** 404–15
[275] Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter A M, Esposito I and Navab N 2016 Structure-preserving color normalization and sparse stain separation for histological images *IEEE Trans. Med. Imaging* **35** 1962–71
[276] Qian W W, Xia C, Venugopalan S, Narayanaswamy A, Dimon M, Ashdown G W, Baum J, Peng J and Ando D M 2020 Batch equalization with a generative adversarial network *Bioinformatics* **36** 875–83
[277] Arvidsson I, Overgaard N C, Krzyzanowska A, Marginean F-E, Simoulis A, Bjartell A, Aström K and Heyden A 2020 Domain-adversarial neural network for improved generalization performance of Gleason grade classification *Proc. SPIE* **11320** 1132016
[278] Mignardi M, Ishaq O, Qian X and Wählby C 2017 Bridging histology and bioinformatics—computational analysis of spatially resolved transcriptomics *Proc. IEEE* **105** 530–41
[279] Chen R J *et al* 2024 Towards a general-purpose foundation model for computational pathology *Nat. Med.* **30** 850–62
[280] Sarwar S, Dent A, Faust K, Richer M, Djuric U, Van Ommeren R and Diamandis P 2019 Physician perspectives on integration of artificial intelligence into diagnostic pathology *npj Digit. Med.* **2** 28
[281] Rivenson Y, Liu T, Wei Z, Zhang Y, Haan K and Ozcan A 2019 PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning *Light Sci. Appl.* **8** 23
[282] Li J *et al* 2021 Biopsy-free in vivo virtual histology of skin using deep learning *Light Sci. Appl.* **10** 1
[283] de Haan K *et al* 2021 Deep learning-based transformation of H&E stained tissues into special stains *Nat. Commun.* **12** 4884
[284] Levy J J, Jackson C R, Sriharan A, Christensen B C and Vaickus L J 2020 Preliminary evaluation of the utility of deep generative histopathology image translation at a mid-sized NCI cancer center *Proc. 13th Int. Joint Conf. on Biomedical Engineering Systems and Technologies (BIOSTEC 2020)* vol 3 (*Valletta, Malta, 24–26 February 2020*) (Bioinformatics) pp 302–11
[285] Rivenson Y, de Haan K, Wallace W D and Ozcan A 2020 Emerging advances to transform histopathology using virtual staining *BME Frontiers* **2020** 9647163
[286] Cohen J P, Luck M and Honari S 2018 Distribution matching losses can hallucinate features in medical image translation *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018* pp 529–36
[287] Zhang Y, de Haan K, Rivenson Y, Li J, Delis A and Ozcan A 2020 Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue *Light Sci. Appl.* **9** 1
[288] Kapil A, Wiestler T, Lanzmich S, Silva A, Steele K, Rebelatto M, Schmidt G and Brieu N 2019 DASGAN—joint domain adaptation and segmentation for the analysis of epithelial regions in histopathology PD-L1 images (available at: https://openreview.net/forum?id=SkxOZOh2gr) (Accessed 15 Auguest 2020)
[289] Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, Liu Y, Topol E, Dean J and Socher R 2021 Deep learning-enabled medical computer vision *npj Digit. Med.* **4** 5
[290] Chai B, Efstathiou C, Yue H and Draviam V M 2024 Opportunities and challenges for deep learning in cell dynamics research *Trends Cell Biol.* **34** 955–67
[291] Sadanandan S K, Ranefall P, Le Guyader S and Wählby C 2017 Automated training of deep convolutional neural networks for cell segmentation *Sci. Rep.* **7** 7860
[292] Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q and Li D 2021 Evaluation and development of deep neural networks for image super-resolution in optical microscopy *Nat. Methods* **18** 194–202
[293] Korczak Z *et al* 2022 Dynamic live/apoptotic cell assay using phase-contrast imaging and deep learning *bioRxiv Preprint* https://doi.org/10.1101/2022.07.18.500422 (Posted 19 July 2022)
[294] Zhu Y, Huang R, Wu Z, Song S, Cheng L and Zhu R 2021 Deep learning-based predictive identification of neural stem cell differentiation *Nat. Commun.* **12** 2614

[295] Patel K B et al 2022 High-speed light-sheet microscopy for the in-situ acquisition of volumetric histological images of living tissue Nat. Biomed. Eng. 6 569–83
[296] Lin R, Kipreos E T, Zhu J, Khang C H and Kner P 2021 Subcellular three-dimensional imaging deep through multicellular thick samples by structured illumination microscopy and adaptive optics Nat. Commun. 12 3148
[297] Saliba N, Gagliano G and Gustavsson A K 2024 Whole-cell multi-target single-molecule super-resolution imaging in 3D with microfluidics and a single-objective tilted light sheet Nat. Commun. 15 10187
[298] Shroff H, Testa I, Jug F and Manley S 2024 Live-cell imaging powered by computation Nat. Rev. Mol. Cell Biol. 25 443–63
[299] Zhang M, Li R, Fu S, Kumar S, Mcginty J, Qin Y and Chen L 2025 Deep learning enhanced light sheet fluorescence microscopy for in vivo 4D imaging of zebrafish heart beating Light Sci. Appl. 14 92
[300] Li J, Chen S, Pan X, Yuan Y and Shen H-B 2022 Cell clustering for spatial transcriptomics data with graph neural networks Nat. Comput. Sci. 2 399–408
[301] Wang Z et al 2021 Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning Nat. Methods 18 551–6
[302] Lundervold A S and Lundervold A 2019 An overview of deep learning in medical imaging focusing on MRI Z. Med. Phys. 29 102–27
[303] Ching T et al 2018 Opportunities and obstacles for deep learning in biology and medicine J. R. Soc. Interface 15 20170387
[304] Montalt-Tordera J, Muthurangu V, Hauptmannand A and Steeden J A 2021 Machine learning in magnetic resonance imaging: image reconstruction Phys. Med. 83 79–87
[305] Mazurowski M A, Buda M, Saha A and Bashir M R 2019 Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI J. Magn. Reson. Imaging 49 939–54
[306] Akkus Z, Galimzianova A, Hoogi A, Rubin D L and Erickson B J 2017 Deep learning for brain MRI segmentation: state of the art and future directions J. Digit. Imaging 30 449–59
[307] Weiss K, Khoshgoftaar T M and Wang D 2016 A survey of transfer learning J. Big Data 3 1–40
[308] Shorten C and M T Khoshgoftaar 2019 A survey on image data augmentation for deep learning J. Big Data 6 1–48
[309] Miller K L et al 2016 Multimodal population brain imaging in the UK Biobank prospective epidemiological study Nat. Neurosci. 19 1523–36
[310] Van Essen D C, Smith S M, Barch D M, Behrens T E, Yacoub E, Ugurbil K and Consortium W-M H C P 2013 The WU-Minn human connectome project: an overview NeuroImage 80 62–79
[311] LaMontagne P J, Benzinger T L, Morris J C, Keefe S, Hornbeck R, Xiong C, Hassenstab E G, Moulder K, Vlassenko A G and Raichle M E 2019 OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease MedRxiv Preprint https://doi.org/10.1101/2019.12.13.19014902 (Posted 15 December 2019)
[312] Treder K P, Huang C, Kim J S and Kirkland A I 2022 Applications of deep learning in electron microscopy Microscopy 71 i100–i15
[313] Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst J-M, Ciompi F and van der Laak J 2019 Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology Med. Image Anal. 58 101544
[314] Rotunno E, Tavabi A H, Rosi P, Frabboni S, Tiemeijer P, Dunin-Borkowski R E and Grill V 2021 Alignment of electron-optical beam shaping elements using a convolutional neuralnetwork (arXiv:2101.04358)
[315] Matuszewski D J and Sintorn I 2021 TEM virus images: benchmark dataset and deep learning classification Comput. Methods Programs Biomed. 209 106318
[316] Willard J, Jia X, Xu S, Steinbach M and Kumar V 2020 Integrating physics-based modeling with machine learning: a survey (arXiv:2003.04919v2)
[317] Budd S, Robinson E C and Kainz B 2021 A survey on active learning and human-in-the-loop deep learning for medical image analysis Med. Image Anal. 71 102062
[318] Gupta A, Wählby C, Sabirsch A and Sintorn I 2022 SimSearch: a human-in-the-loop learning framework for fast detection of regions of interest in microscopy images (arXiv)
[319] Pang G, Shen C, Cao L and Hengel A V D 2021 Deep learning for anomaly detection: a review ACM Comput. Surv. 54 1–38
[320] Chandrasekaran S N, Ceulemans H, Boyd J D and Carpenter A E 2021 Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat. Rev. Drug Discovery 20 145–59
[321] Archit A et al 2025 Segment anything for microscopy Nat. Methods 22 579–91
[322] Stringer C and Pachitariu M 2025 Cellpose3: one-click image restoration for improved cellular segmentation Nat. Methods 22 592–9
[323] Pratapa A, Doron M and Caicedo J C 2021 Image-based cell phenotyping with deep learning Curr. Opin. Chem. Biol. 65 9–17
[324] Simm J et al 2018 Repurposing high-throughput image assays enables biological activity prediction for drug discovery Cell Chem. Biol. 25 611–618.e3
[325] Caicedo J C, McQuin C, Goodman A, Singh S and Carpenter A E 2018 Weakly supervised learning of single-cell feature embeddings Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2018 9309–18
[326] Franco-Barranco D et al 2024 BiaPy: a unified framework for versatile bioimage analysis with deep learning bioRxiv p 2024.02.03.576026
[327] Weisbart E and Cimini B A 2023 Distributed-something: scripts to leverage AWS storage and computing for distributed workflows at scale Nat. Methods 20 1120–1
[328] Luijten B, Chennakeshava N, Eldar Y C, Mischi M and van Sloun R J G 2023 Ultrasound signal processing: from models to deep learning Ultrasound Med. Biol. 49 677–98
[329] Gröhl J, Schellenberg M, Dreher K and Maier-Hein L 2021 Deep learning for biomedical photoacoustic imaging: a review Photoacoustics 22 100241
[330] Bell M A L 2022 Applications of ultrasound imaging in the deep learning age Proc. SPIE 12204 1220403
[331] Kraus H G 1989 J. Opt. Soc. Am. A 6 1196
[332] Reiter A and Bell M A L 2017 A machine learning approach to detect point sources in photoacoustic data Proc. SPIE 100643J
[333] Allman D, Reiter A and Bell M A L 2018 Photoacoustic source detection and reflection artifact removal enabled by deep learning IEEE Trans. Med. Imaging 37 1464–77
[334] Bell M A L 2019 Deep learning the sound of light to guide surgeries Proc. SPIE 10868 108680G
[335] Lediju M A, Trahey G E, Byram B C and Dahl J J 2011 Spatial coherence of backscattered echoes: imaging characteristics IEEE Trans. Ultrason. Ferroelectr. Freq. Control 58 1377–88

[336] Wiacek A, Oluyemi E, Myers K, Mullen L and Bell M A L 2020 Coherence-based beamforming increases the diagnostic certainty of distinguishing fluid from solid masses in breast ultrasound exams *Ultrasound Med. Biol.* **46** 1380–94
[337] Wiacek A, González E and Bell M A L 2020 CohereNet: a deep learning architecture for ultrasound spatial correlation estimation and coherence-based beamforming *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **67** 2574–83
[338] Nair A A, Washington K, Tran T, Reiter A and Bell M A L 2020 Deep learning to obtain simultaneous ultrasound image and segmentation outputs from a single input of raw channel data *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **67** 2493–509
[339] Bhatt M, Nair A, Kempski K M and Bell M A L *Multi-task learning for ultrasound image formation and segmentation directly from raw in vivo data Proc. 2020 IEEE Int. Ultrasonics Symp.* (*Virtual*, *6–11 September 2020*)
[340] Gubbi M R and Bell M A L Deep learning-based photoacoustic visual servoing: using outputs from raw sensor data as inputs to a robot controller *IEEE Int. Conf. on Robotics and Automation (ICRA)* (*Xi'an, China*, *30–June 5 May*)
[341] Bell M A L, Huang J, Hyun D, Eldar Y C, van Sloun R and Mischi M Challenge on ultrasound beamforming with deep learning (CUBDL) *Proc. 2020 IEEE Int. Ultrasonics Symp.* (*6–11 September 2020*)
[342] Hyun D *et al* 2021 Deep learning for ultrasound image formation: CUBDL evaluation framework & open datasets *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **68** 3466–83
[343] Falk T *et al* 2019 U-Net: deep learning for cell counting, detection, and morphometry *Nat. Methods* **16** 67–70
[344] Schmidt U, Weigert M, Broaddus C and Myers G 2018 Cell detection with star-convex polygons pp 265–73 (arXiv:1806.03535)
[345] Redmon J and Farhadi A 2016 YOLO9000: better, faster, stronger (arXiv:1612.08242)
[346] Gómez-de-Mariscal E, García-López-de-Haro C, Ouyang W, Donati L, Lundberg E, Unser M, Muñoz-Barrutia A and Sage D 2021 DeepImageJ: a user-friendly environment to run deep learning models in ImageJ *Nat. Methods* **18** 1192–5
[347] Schindelin J *et al* 2012 Fiji: an open-source platform for biological-image analysis *Nat. Methods* **9** 676–82
[348] Schindelin J, Rueden C T, Hiner M C and Eliceiri K W 2015 The ImageJ ecosystem: an open platform for biomedical image analysis *Mol. Reprod. Dev.* **82** 518–29
[349] Hidalgo-Cenalmor I *et al* 2024 DL4MicEverywhere: deep learning for microscopy made flexible, shareable and reproducible *Nat. Methods* **21** 925–7
[350] Pachitariu M and Stringer C 2022 Cellpose 2.0: how to train your own model *Nat. Methods* **19** 1634–41
[351] Belevich I and Jokitalo E 2020 DeepMIB: user-friendly and open-source software for training of deep learning network for biological image segmentation *bioRxiv* p 2020.07.13.200105 (Accessed 28 February 2025)
[352] Ouyang W *et al* 2022 BioImage model zoo: a community-driven resource for accessible deep learning in bioimage analysis *bioRxiv* p 2022.06.07.495102 (Accessed 18 September 2022)
[353] Ouyang W, Mueller F, Hjelmare M, Lundberg E and Zimmer C 2019 ImJoy: an open-source computational platform for the deep learning era *Nat. Methods* **16** 1199–200
[354] Berg S *et al* 2019 ilastik: interactive machine learning for (bio)image analysis *Nat. Methods* **16** 1226–32
[355] Pontén F, Jirström K and Uhlen M 2008 The Human Protein Atlas—a tool for pathology *J. Pathol.* **216** 387–93
[356] The Human Protein Atlas (available at: www.proteinatlas.org/) (Accessed 28 February 2025)
[357] de Chaumont F *et al* 2012 Icy: an open bioimage informatics platform for extended reproducible research *Nat. Methods* **9** 690–6
[358] Bankhead P *et al* 2017 QuPath: open source software for digital pathology image analysis *Sci. Rep.* **7** 16878
[359] Franco-Barranco D, Andrés-San Román J A, Gómez-Gálvez P, Escudero L M, Muñoz-Barrutia A and Arganda-Carreras I 2023 BiaPy: a ready-to-use library for bioimage analysis pipelines *2023 IEEE 20th Int. Symp. on Biomedical Imaging (ISBI)* (*28 February 2025*) (available at: https://ieeexplore.ieee.org/abstract/document/10230593)
[360] Laine R F, Arganda-Carreras I, Henriques R and Jacquemet G 2021 Avoiding a replication crisis in deep-learning-based bioimage analysis *Nat. Methods* **18** 1136–44
[361] Sofroniew N *et al* 2022 napari: a multi-dimensional image viewer for Python *Zenodo* (available at: https://zenodo.org/record/7276432) (Accessed 29 May 2023)
[362] Abdurahman F, Fante K A and Aliy M 2021 Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models *BMC Bioinform.* **22** 112
[363] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library (arXiv:1912.01703)
[364] Abadi M *et al* TensorFlow: large-scale machine learning on heterogeneous distributed systems
[365] Kluyver T 2016 Jupyter Notebooks—a publishing format for reproducible computational workflows *Positioning and Power in Academic Publishing: Players, Agents and Agendas* ed F Loizides and B Schmidt *et al* 87–90 (IOS Press)
[366] Jupyter P *et al* 2018 Binder 2.0—reproducible, interactive, sharable environments for science at scale *Proc. 17th Python Science Conf.* pp 113–20
[367] Schneider C A, Rasband W S and Eliceiri K W 2012 NIH Image to ImageJ: 25 years of image analysis *Nat. Methods* **9** 671–5
[368] Haberl M G *et al* 2018 CDeep3M-plug-and-play cloud-based deep learning for image segmentation *Nat. Methods* **15** 677–80
[369] Bannon D *et al* 2021 DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes *Nat. Methods* **18** 43–45
[370] Sheller M J *et al* 2020 Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data *Sci. Rep.* **10** 12598
[371] Chen M *et al* 2021 Evaluating large language models trained on code (arXiv:2107.03374)
[372] Patterson D *et al* 2021 Carbon emissions and large neural network training (arXiv:2104.10350)
[373] Moen E, Bannon D, Kudo T, Graf W, Covert M and Van Valen D 2019 Deep learning for cellular image analysis *Nat. Methods* **16** 1233–46
[374] Midtvedt D, Mylnikov V, Stilgoe A, Käll M, Rubinsztein-Dunlop H and Volpe G 2022 Deep learning in light–matter interactions *Nanophotonics* **11** 3189–214
[375] Bachimanchi H *et al* 2022 Microplankton life histories revealed by holographic microscopy and deep learning (arXiv:2202.09046)
[376] Grauer J, Schmidt F, Pineda J, Midtvedt B, Löwen H, Volpe G and Liebchen B 2021 Active droploids *Nat. Commun.* **12** 6005
[377] Midtvedt B, Helgadottir S, Argun A, Pineda J, Midtvedt D and Volpe G 2020 Deeptrack-2.0 (available at: https://github.com/softmatterlab/DeepTrack-2.0)
[378] Meijering E 2020 A bird's-eye view of deep learning in bioimage analysis *Comput. Struct. Biotechnol. J.* **18** 2312–232
[379] Caicedo J C *et al* 2019 Nucleus segmentation across imaging experiments: the 2018 data science bowl *Nat. Methods* **16** 1247–53
[380] Schmidt U *et al* 2018 Cell detection with start-convex polygons *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Springer Cham) pp 265–73

[381] Inés A, Domínguez C, Heras J, Mata E and Pascual V 2019 DeepClas4Bio: connecting bioimaging tools with deep learning frameworks for image classification *Comput. Biol. Med.* **108** 49–56
[382] García López de Haro C, Dallongeville S, Musset T, Gómez-de-Mariscal E, Sage D, Ouyang W, Muñoz-Barrutia A, Tinevez J-Y and Olivo-Marin J-C 2024 JDLL: a library to run deep learning models on Java bioimage informatics platforms *Nat. Methods* **21** 7–8
[383] Ershov D *et al* 2022 TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines *Nat. Methods* **19** 829–32
[384] Pietzsch T, Saalfeld S, Preibisch S and Tomancak P 2015 BigDataViewer: visualization and processing for large image data sets *Nat. Methods* **12** 481–3
[385] Gómez-de-Mariscal E, Del Rosario M, Pylvänäinen J W, Jacquemet G and Henriques R 2024 Harnessing artificial intelligence to reduce phototoxicity in live imaging *J. Cell Sci.* **137** jcs261545
[386] Fuster-Barceló C, García-López-de-Haro C, Gómez-de-Mariscal E, Ouyang W, Olivo-Marin J-C, Sage D and Muñoz-Barrutia A 2024 Bridging the gap: integrating cutting-edge techniques into biological imaging with deepImageJ *Biol. Imaging* **4** e14
[387] Yams N B, Richardson V, Shubina G E, Albrecht S and Gillblad D 2020 Integrated AI and innovation management: the beginning of a beautiful friendship *Technol. Innov. Manage. Rev.* **10** 5–18
[388] Goodfellow I J *et al* 2013 Challenges in representation learning: a report on three machine learning contests *Int. Conf. on Neural Information Processing*
[389] Kaggle 2022 (available at: www.kaggle.com/) (Accessed 21 March 2022)
[390] Codalab 2022 (available at: https://codalab.org/) (Accessed 21 March 2022)
[391] AI Sweden (available at: www.ai.se/en) (Accessed 21 March 2022)
[392] Adipocyte Cell Imaging Challenge 2020 (available at: www.ai.se/en/project/adipocyte-cell-imaging-challenge) (Accessed 28 February 2025)
[393] Dance A 2021 AI spots cell structures that humans can't *Nature* **592** 154–5