Thesis for the degree of Licentiate of Engineering

Pushing the Boundaries of Biomolecule Characterization through Deep Learning

Henrik Klein Moberg

Department of Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

Pushing the Boundaries of Biomolecule Characterization through Deep Learning

Henrik Klein Moberg

Licentiatavhandlingar vid Chalmers tekniska högskola

Department of Physics

Chalmers University of Technology

SE–412 96 Göteborg, Sweden

Telephone + 46 (0) 31 – 772 1000

Cover: A text-to-image artificial intelligence's interpretation of the phrase "Artificial Intelligence Studies Biological Molecules Through Optical Microscopy". Credit: DALL-E 2 from OpenAI.

Typeset by the author using LaTeX.

Printed by Chalmers Reproservice

Gothenburg, Sweden 2022

to my dear family and friends

Abstract

The importance of studying biological molecules in living organisms can hardly be overstated as they regulate crucial processes in living matter of all kinds. Their ubiquitous nature makes them relevant for disease diagnosis, drug development, and for our fundamental understanding of the complex systems of biology. However, due to their small size, they scatter too little light on their own to be directly visible and available for study. Thus, it is necessary to develop characterization methods which enable their elucidation even in the regime of very faint signals. Optical systems, utilizing the relatively low intrusiveness of visible light, constitute one such approach of characterization. However, the optical systems currently capable of analyzing single molecules in the nano-sized regime today either require the species of interest to be tagged with visible labels like fluorescence or chemically restrained on a surface to be analyzed. Ergo, there exist effectively no methods of characterizing very small biomolecules under naturally relevant conditions through unobtrusive probing. Nanofluidic Scattering Microscopy is a method introduced in this thesis which bridges this gap by enabling the real-time label-free size-and-weight determination of freely diffusing molecules directly in small nano-sized channels. However, the molecule signals are so faint, and the background noise so complex with high spatial and temporal variation, that standard methods of data analysis are incapable of elucidating the molecules' properties of relevance in any but the least challenging conditions. To remedy the weak signal, and realize the method's full potential, this thesis' focus is the development of a versatile deep-learning based computer-vision platform to overcome the bottleneck of data analysis. We find that said platform has considerably increased speed, accuracy, precision and limit of detection compared to standard methods, constituting even a lower detection limit than any other method of label-free optical characterization currently available. In this regime, hitherto elusive species of biomolecules become accessible for study, potentially opening up entirely new avenues of biological research. These results, along with many others in the context of deep learning for optical microscopy in biological applications, suggest that deep learning is likely to be pivotal in solving the complex image analysis problems of the present and enabling new regimes of study within microscopy-based research in the near future.

Keywords: Deep Learning, Artificial Intelligence, Computer Vision, Optical Microscopy, Nanofluidic Scattering Microscopy, Molecule Characterisation, Biological Imaging

List of Appended Papers

This thesis is based on the following appended papers:

Paper I. Barbora Špačková, **Henrik Klein Moberg**, Joachim Fritzsche, Johan Tenghamn, Gustaf Sjösten, Hana Šípová-Jungová, David Albinsson, Quentin Lubart, Daniel van Leeuwen, Fredrik Westerlund, Daniel Midtvedt, Elin K. Esbjörner, Mikael Käll, Giovanni Volpe, and Christoph Langhammer. *Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles* Submitted 2021, Published 2022 in *Nature Methods*.

Paper II. Henrik Klein Moberg, Bohdan Yeroshenko, David Albinsson, Joachim Fritzsche, Daniel Midtvedt, Barbora Špačková, Giovanni Volpe, Christoph Langhammer . *Seeing the Invisible: Deep Learning Optical Microscopy for Label-Free Biomolecule Screening in the sub-10 kDa Regime*, In Manuscript

Paper III. Benjamin Midtvedt, Dana Hassan, **Henrik Klein Moberg**, Jesús Pineda, Jesús Domínguez. *Review: cross modality transforms in biological microscopy enabled by deep learning*, In manuscript, equal author contribution

My Contribution to Appended Papers

Paper I: I developed the deep learning methods used for analyzing proteins, DNA fragments, extracellular vesicles and lipid nanoparticles, as well as carried out all the aforementioned analysis. I Wrote all sections in the paper related to deep learning, as well as some portions of the results section relating to the results of the deep learning analysis.

Paper II: I developed all theoretical results concerning the limits of the deep learning methods, as well as wrote the manuscript and produced all figures.

Paper III: I wrote the section on molecular imaging and produced its corresponding figure.

Related Publications Not Included in The Thesis

David Tomecek, **Henrik Klein Moberg**, Iwan Darmadi, Írem Tanyelí, Elin Langhammer, Olof Andersson, Christoph Langhammer. *A nanoplasmonic hydrogen sensor with sub-1000 ppm limit of detection in humid air* , In manuscript

Laura Waller, Kristina Monakhova, Jason Smith, Wang Hao, Lei Tian, Natasa Sladoje, Christoph Langhammer, Barbora Špačková, **Henrik Klein Moberg**, Aydogan Ozcan, Carlo Manzo, Saga Helgadottir, Aykut Argun, Mite Mijalkov, Daniel Vereb, Yu-wei Chang, Joana Pereira, Daniel Midtvedt, Ida-Maria Sintorn, Benjamin Midtvedt, Jesús Pineda, Sage Daniel, Estibaliz Gómez-de-Marisca, Wei Ouyang, Ebba Josefson, Carolina Wälhby, Lei Tian, Giovanni Volpe *Roadmap on Machine Learning for Microscopy*, In manuscript

List of Acronyms

ADAM	–	Adaptive Moment Estimation
AI	–	Artificial Intelligence
ANNA-PALM	–	Artificial Neural Network Accelerated PALM
DL	–	Deep Learning
BNP	–	Biological Nanoparticle
BSA	–	Bovine Serum Albumin
CNN	–	Convolutional Neural Network
COCO	–	Common Objects in Context
CV	–	Computer Vision
DINO	–	Self-Distillation with No labels
DL	–	Deep Learning
EV	–	Extracellular Vesicle
FCNN	–	Fully Convolutional Neural Network
FPN	–	Feature Pyramid Network
FWHM	–	Full-width at Half-maximum
GAN	–	Generative Adversarial Network
GPU	–	Graphics Processing Unit
iOC	–	Integrated Optical Contrast
iSCAT	–	Interferometric Scattering Microscopy
LoD	–	Limit of Detection
LodeSTAR	–	Low-shot deep Symmetric Tracking And Regression
LoQ	–	Limit of Quantization
ML	–	Machine Learning
MT	–	Microtubules
MW	–	Molecular Weight
NA	–	Numerical Aperture
NPC	–	Nuclear Pore Complexes
NSM	–	Nanofluidic Scattering Microscopy
PALM	–	Photo-Activated Localization Microscopy
PSF	–	Point Spread Function
RL	–	Reinforcement Learning
SA	–	Standard Analysis
SLB	–	Supported Lipid Bilayer

SMLM	–	Single-Molecule Localization Microscopy
SNR	–	Signal-to-Noise Ratio
STORM	–	Stochastic Optical Reconstruction Microscopy
TPU	–	Tensor Processing Unit
YOLO	–	You Only Look Once

Contents

Chapter 1

Introduction

Through the paradigm of deep learning (DL), recent progress in computer vision (CV) has been truly astounding; quickly evolving from struggling to recognize animals in relatively simple images to leading a revolution in autonomous vehicles [1], image-based medical diagnosis [2], material science [3], astronomy [4], weather prediction [5], waste disposal [6] and many other fields in the span of just a few years. The keys underpinning the success lie in the fruitful marriage of a few seemingly disparate technological advancements; the development of computationally tractable self-learning neural network-based algorithms [7], the ever-growing availability and quality of digital acquisition devices enabling high-throughput experimentation, and the explosive growth of the financially lucrative personal computer market leading to ever-faster parallellizable computing devices. The combination of large amounts of high-quality images, effectively 2-dimensional vector spaces containing locally dependent features, with convolution-based DL architectures known under the umbrella term of *convolutional neural networks* (CNNs), has been particularly potent. Self-learning CNN-based algorithms learning general short-range dependencies between locally correlated data-points (i.e. pixels in images), has been the workhorse of most recent progress [8]. This generality means that algorithms developed in the context of e.g. animal classification or facial recognition can often directly be applied to other image-based CV tasks with relatively minor tweaks, as a result of the generic self-learning nature of DL-based algorithms. This generality has effectively formed a chain reaction within quantitative image-based analytical research, in which lessons learnt in one application of DL-based CV can often directly be applied to every other application.

One field where this potential is becoming increasingly clear is optical microscopy, in which data analysis is often a limiting step because of images which are both highly complex and have low signal-to-noise-ratio (SNR). Promising proof-of-principle results on super-resolving fluorescence microscopy [9], DL-enhanced holography [10] and virtual chemical staining [11] to name a few examples, suggests that DL is beginning to lead a revolution also in optical microscopy as it has in aforementioned fields.

In life sciences, optical microscopy has proven instrumental for fast and accurate characterization of biomolecule properties in- and outside of complex biofluids. Historically, in the regime of small biomolecules (∼100s of kDa), sensitive imaging techniques such as fluorescence microscopy and interferometric scattering microscopy (iSCAT) fill this characterization role, though they suffer from problems arising from fluorescent labelling [12] and necessity of surface binding, respectively. To remedy this, we have in paper I introduced Nanofluidic Scattering Microscopy (NSM), which relies on the sub-wavelength interference of light scattered from a nanofluidic channel and a nano-object, such as a biomolecule, inside it. However, the combination of generally low SNR resulting in typically "blinking" trajectories, changing imaging conditions, and noise which is complex in both the spatial and temporal dimensions, makes it very difficult to reliably calculate the biomolecules' properties through standard heuristic-based methods alone. However, to borrow the words of Dr. Hassabis of Google's DeepMind, "AI is perfect at taking weak, messy signals and making sense of them." [13]

This forms the essential background against which the works within this thesis are justified; we show that by employing DL-based CV algorithms to analyze biomolecule trajectories measured from within our in-house NSM system, we not only manage to analyze complex bio-fluids in the form of extracellular vesicles, but also surpass state-of-the-art results on nano-scale biomolecule weight-and-size screening and trajectory tracking. We show that a custom CNN-based architecture trained on simulated biomolecule trajectories atop simulated noise surpasses the performance of standard algorithms in terms of tracking and determining the molecular weight (MW) and hydrodynamic radius (R_s) of biomolecules in the low-kDa regime. Furthermore, said model can accurately track and measure MW and R_s of biomolecules even below the 10-kDa regime, constituting approximately an order of magnitude improvement in limit of detection (LoD) over current state-of-the-art. In this regime, hitherto elusive species of biomolecules become accessible for label-free study using NSM, including cytokines ($\approx 5 - 25$ kDa) important for cancer research and the protein hormone insulin (≈ 5.6 kDa), potentially opening up entirely new avenues of biological research [14].

In this thesis, I delve deeper into the complexities of DL-based CV which has enabled all this progress in a short time, and argue how we might employ it to our in-house microscopy method NSM to analyze single biomolecule trajectories and potentially break through the state-of-the-art in microscopy in certain respects.

Looking further ahead, the ability to optically monitor particles in the sub-10 kDa regime is a tantalizing prospect, as it may enable a quick and cost-efficient analysis of subtle binding events, such as protein-DNA bindings [15] or nanovesicle corona formations [16]. Beyond the fields of life science and biotechnology, it may potentially also enable the analysis of currently invisible contaminator particles in air [17] and catalytic reactions of single particles [18], suggesting significant research potential beyond the immediate consequences such a decrease in LoD may bring for the field of biology.

Chapter 2

Background

This chapter is dedicated to introducing the key concepts required for understanding the results produced in this thesis. It is not meant as an exhaustive review (see e.g. [19]); rather, its main purpose is to provide an intuitive understanding behind the most important tools of optical microscopy and deep learning which lie at the heart of this work. Specifically, the first section of this chapter introduces the basics of optical microscopy for molecular imaging, the method of nano-channel-based optical microscopy introduced in Paper I, important features of a select few biological molecules studied in this thesis, as well as touching upon the growing need for advanced data science within biotechnology and life science applications. The second section of this chapter introduces deep learning, specifically in the context of computer vision in general and for biological applications in particular. Finally, the chapter ends with a short perspective on the future trajectory of deep learning, and the prospective consequences which may follow in terms of what new types of advanced analysis might soon become possible, as well as a short summary of review Paper III.

2.1 Optical Microscopy for Molecular Imaging

In general, an optical microscope is a system which leads light in the visible spectra range through a sequence of lenses to magnify images of small objects. Though these systems have existed for a long time, arguably invented in the 1500s in their simplest implementation [20], it took many hundred years for the technology to become powerful enough to enable imaging of systems on a molecular scale. This may be understood as follows; the shortest physical distance between two points in an image, also known as the *spatial resolution*, is the single most important feature in optical microscopy [21]. The principal limitations impinging the potential of the spatial resolution of optical microscopy are imposed by an irreducible phenomenon; the physics of *diffraction*. Effectively, even in the absence of any imperfections in lens quality or optical component misalignment, the resolution of a microscope is ultimately proportional to the wavelength of scattered or transmitted light being

measured, and inversely proportional to the so-called numerical aperture (NA) of its objective. This happens since light from an emitting light source in a sample generates a fundamentally limited diffraction pattern known as a point spread function (PSF) when passing to the image plane through an object. By our definition above, the minimal distance between two points in a sample which can be resolved thereby becomes limited by the distance at which two overlapping PSFs can be resolved. This limitation may be transcended for instance through super-resolution imaging techniques, i.e. techniques which transcend the diffraction limit, of which DL-based CV methods already shows great potential as outlined in Paper III and section 2.2.5.

Indeed, optical characterization methods have proven instrumental for fast and accurate measurement of biomolecule properties in- and outside of complex biofluids [22]. Historically, for studies which require single-biomolecule resolution, powerful microscopy techniques such as fluorescence microscopy and iSCAT fill this role [23]. However, fluorescence microscopy requires the molecules to be labelled by a clearly visible fluorophore or nanoparticle, arguably altering their properties and behaviours [12], and iSCAT requires binding them to a surface prior to measurement, thereby losing valuable information regarding their movement behaviour and leaving a large portion of them undetected. To remedy this, we have in Paper I introduced NSM, which relies on the sub-wavelength interference of light scattered from a nanofluidic channel and a nano-object, such as a biomolecule, inside it. Thus, we enable label-free mass and size screening of freely diffusing single-biomolecule trajectories, effectively constituting a leapfrogging step in the field of optical molecular imaging and characterization. The related field of optical microscopy in the context of molecular imaging is irrefutably immensely rich in history and complexity, and the full breadth of its nuances is not recounted in this chapter. Rather, it is given in its proper context for the purposes of the work accomplished in this thesis; in the form of our in-house developed NSM technique.

2.1.1 Nanofluidic Scattering Microscopy

NSM works by a laser irradiating visible light onto a nano-sized channel embedded in a silicon dioxide matrix within a nanofluidic chip containing single biomolecules freely diffusing in solution, as shown in figure 2.1a. In this setup, as shown in figure 2.1b-c, the biomolecule and nanochannel scatter light coherently into the collection optics, resulting in an improved optical contrast of the imaged biomolecules by several orders of magnitude. The total scattering intensity recorded from a region of a nanochannel of length $L = \frac{3\pi}{2k}$ can be approximated as (see section 1 in Supplementary Information of Paper I)

$$I_{\mathrm{t}} \approx I_{\mathrm{c}} + I_{\mathrm{m}} - \sqrt{2 I_{\mathrm{c}} I_{\mathrm{m}}}. \tag{2.1}$$

Here, I_{t} is the total scattered intensity from the nanochannel, I_{c} scattered intensity from the channel and I_{m} scattered intensity from the biomolecule (see [24] for further details). A video is then recorded of the total scattered intensity with a camera

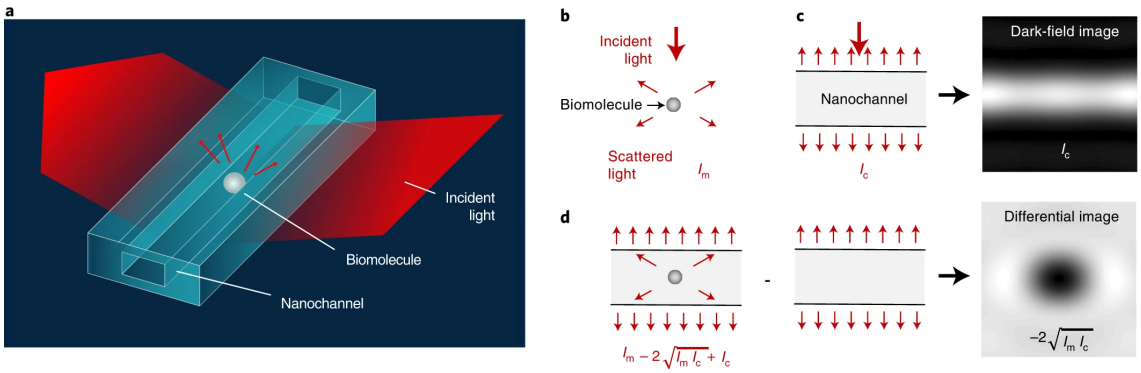

Figure 2.1: **a**, Artist's rendition of the experimental configuration where visible light irradiates a nanochannel with a biomolecule inside, and where the light scattered from the system is collected in dark-field configuration. **b**, Schematic of light scattered by a single biomolecule. **c**, Schematic of 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. Figure reproduced with permission from [24].

with high enough frame-rate to also be able to measure the biomolecules' diffusivity along the nanochannel. By subtracting an image of the empty nanochannel from an image with a biomolecule inside, as depicted in figure 2.1d, we acquire the dark-field differential image containing only the inference term $\sqrt{2I_c I_m}$ in equation 2.1 (I_m is vanishingly small since scattering from sub- wavelength objects scales as volume squared and the nano-channel's volume is much larger than the molecular volume). This inference term can be several orders of magnitude larger than the scattering intensity produced by the biomolecule alone (I_m) outside the nanochannel, which in essence constitutes the key feature of NSM and enables the direct imaging of diffusing biomolecules and other nano-objects inside a nanochannel. From videos of a sequence of differential images acquired in this way, so-called kymographs, the molecular weight (MW) and hydrodynamic radius (R_s) of each individual molecule can be measured.

Determining the MW of a biomolecule from its trajectory is made possible since the integrated optical contrast (iOC) of said trajectory is linearly dependent on the polarizability α_m of the biomolecule [25], where iOC is defined as

$$iOC = \int_{x=0}^{L} \frac{I_t(x) - I_c(x)}{I_c(x)} dx \tag{2.2}$$

for a channel of length L. This polarizability α_m is in turn proportional to $a \times MW$, where $a \approx 0.46$ Å^3/Da [26]. It is also inversely proportional to the cross-sectional area A of the nanochannel and proportional to $\bar{n} = \left(1.5n_{H_2O}^2 + 0.5n_{SiO_2}^2\right) / \left(n_{H_2O}^2 - n_{SiO_2}^2\right)$ [24]. These two factors are constant and easily measured before an experiment, meaning that the MW of individual biomolecules measured in our system can be determined as

$$MW = iOC\frac{A}{\bar{n}a}. \tag{2.3}$$

Determining the R_s of a biomolecule from its trajectory is made possible since its diffusivity is correlated with its size. In particular, by approximating the biomolecule as a solid neutrally charged sphere, its R_s can be estimated using the Stokes-Einstein equation with a correction term K for hindrance effects arising from the diffusion of small objects in restricted volumes [27]. R_s is estimated as

$$R_s = K\frac{k_B T}{6\pi\eta D}, \tag{2.4}$$

where k_B is Boltzmann's constant, T is temperature, and η is the viscosity of the liquid in the nanochannel. By looking at a differential image of the nanochannel over time, even exceedingly minuscule biomolecules can be detected and analyzed, and by utilizing the fact that a biomolecule's optical contrast and diffusivity is correlated with its weight and size through its polarizability and the Stokes-Einstein equation respectively, the largest limiting factor for our analysis of smaller biomolecules quickly becomes the analysis of faint signals in overwhelming noise. To overcome these challenges, we employ DL-based approaches as described in sections 2.2 and 3.1 for the data analysis.

2.1.2 Nanofluidic Chip

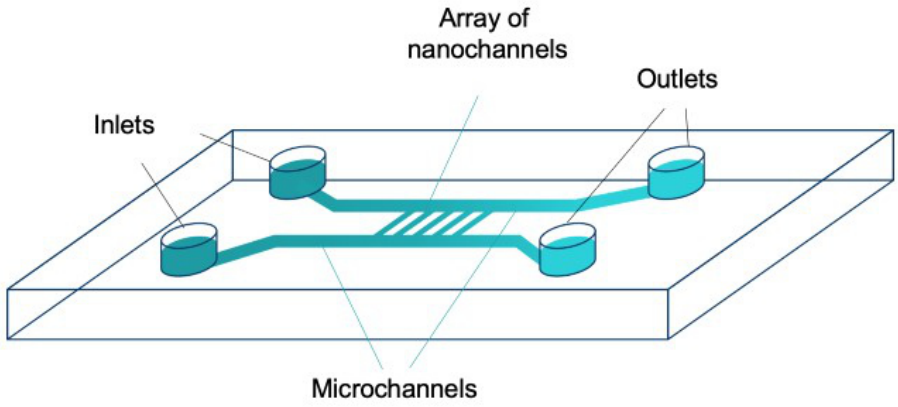

Figure 2.2: Schematic of the nanofluidic chip. Two inlets connect to corresponding outlets through a series of nanochannels with different cross-sectional dimensions. A laser irradiates the nanochannel, wherein the imaged molecule always remains in focus, as a camera measures the light scattered from said nanochannel whilst biomolecules either freely diffuse or are pushed through with an induced flow. Figure reproduced with permission from [24].

The bases of NSM are the nano-fabricated chips containing the nano-channels within which the biomolecules are analyzed. A schematic of the nanofluidic chips used for our experiments is shown in figure 2.2, which consists of a series of nanochannels with specific cross-sectional dimensions chosen to accomodate certain ranges of

particle size. The channels are connected to macroscopic in- and outlets through two microchannels, which is pressurized to 2 bar to facilitate sample transport into the channels. During measurement, this flow is usually turned off to let the molecules freely diffuse inside the channels. In this thesis, three channels were used: channel I with cross-sectional area $A_I = 100 \times 27$ nm^2 to study mainly proteins, channel II with $A_{II} = 110 \times 72$ nm^2 to study mainly DNA sequences, and channel III with $A_{III} = 200 \times 225$ nm^2 to study larger molecules such as extracellular vesicles (EVs) which are further explored in the next section.

2.1.3 Biological Molecules and Nanoparticles

Biological molecules are ubiquitously present in all living organisms, existing at multiple different size scales and regulating essentially all biological functions in both animal and plant bodies. Among the most important molecules are the proteins; a type of molecule comprised of amino acids, which influence cell structure, enzyme activity, molecule transport among many other functions [28]. In Paper I, we see evidence both of monomeric proteins, which consist of single molecules made up of one peptide chain, and polymeric proteins which are composed of either two peptide chains (dimeric proteins) or three peptide chains (trimeric proteins) [29]. Polymeric proteins consist of multiple copies (polymers) of a single protein molecule [29]. Studying these proteins and their polymeric aggregates are crucial for safe drug development and furthering our understanding of biological systems in general [28]. In Paper 1, we characterise the dimeric protein thyroglobulin, an important biomarker for thyroid cancer treatment [30], ferritin important for iron regulation [31], Bovine Serum Albumin (BSA) which regulates many biological processes and is an important biomarker for several pediatric diseases [32], as well as Alcohol Dehydrogenase (ADH) which serves to break down alcohols that otherwise are toxic [33].

DNA strands, similarly, are considered polymeric since they contain multiple strands of a molecule bound together with hydrogen bonds between complementary nucleotides [29]. The study of DNA and the corresponding genes they encode is the source of several separate fields of biological study, playing a key role in everything from molecular [29] to evolutionary biology [34], as well as genomic DNA analysis [35] and genomic sequencing [36].

Extracellular Vesicles (EVs) are small lipid vesicles that are secreted from cells into the extracellular environment and can be found in all body fluids, including blood, saliva, urine and breast milk [37]. They play an important role in cell-to-cell communication and have been implicated in a variety of diseases such as Alzheimer's and Parkinson's and certain types of cancer [37]. EVs contain proteins, lipids and genetic material that can be used to identify the cell of origin. This makes them valuable tools for studying cellular function and disease pathology, i.e. by analyzing the RNA content of EVs it is possible to determine which genes are being expressed by a particular cell type. This information can then be used to develop new therapies for aforementioned diseases and cancers. Therefore, EVs have potential as diagnostic

tools, i.e. if one could detect specific types of EVs in a patient's body fluid, then this would indicate that they're in a particular disease state, which would potentially allow doctors to diagnose several diseases earlier and begin treatment sooner than is currently possible [37]. Thus, EVs effectively act as mediators of physiological intercellular communication and have key functions in the pathobiology of many diseases [38]. However, their specific properties (surface markers, molecular cargo, weight, size, etc.) are difficult to measure precisely with existing metholodogies, such as dynamic light scattering [39] or nanoparticle tracking analysis [40], thereby necessitating new methods for their precise characterization. In Paper I, we show that NSM is one such promising method for characterization of EVs, exosomes and lipid nanoparticles in general.

2.1.4 Growing Need for Advanced Data Science in Life Science

The scale of modern life science has grown equally impressive and imposing in recent decades, as constantly improving automatized digital acquisition devices along with researchers' rapidly increasing knowledge of the molecular and genetic underpinnings of diseases is making the field inordinately complex. This presents a major challenge for traditional bioinformatics methods that rely on hand-crafted algorithms optimized for a specific dataset size or type. Advanced data science techniques are needed to handle this flood of genomic information effectively and extract insights from it. The availability of experimental data is often abundant, which transforms the challenge of analysis to "digging out some key piece of information from that morass of data", to borrow the words of Sir Roger Penrose [41]. To keep pace with this rapid evolution, advanced data science methods grow evermore central and crucial to the modern scientist's toolbox. In particular, there is a growing need for automatized algorithms which can quickly and robustly make sense of large and diverse datasets, as well as faint signals hidden in noise which is difficult or even infeasible to mitigate experimentally.

Fortunately, recent years have seen the advent of powerful data-science-focused languages such as R and Python, as well as AI platforms such as Tensorflow [42] and PyTorch [43] which are advancing at unprecedented speeds thanks to their open-source nature. As these tools continuously grow to be more democratized and easily available, they will effectively enable all scientists to perform sophisticasted statistical analyses and (deep) machine learning on their large and complex datasets. In addition, they provide a variety of libraries that can be used for specific applications, such as gene expression analysis [44] or drug discovery [45].

In addition to handling large datasets efficiently, advanced data science methods can help identify patterns in complex biological systems [46]. For example, network inference algorithms can be used to identify modules within protein networks or transcriptional regulatory networks [47], or as famously featured in "Method of the year" in both Science and Nature; to predict the 3-dimensional structure of proteins using limited information [48]. Such patterns may reveal new insights into how

diseases develop and could lead to the development of new drugs or treatments.

The study of small proteins is a field ready to emerge; its abscence a symptom of lacking available methods of their characterization rather than any lack of scientific interest [14]. Studies have shown that sub-10 kDa molecules are ubiquitously present in many biological materials, particularly in membranes of cells, and may even regulate diverse processes, such as spore formation, cell division, enzymatic activities and signal transduction [14]. Further, there remain many questions regarding the fundamental nature of small protein interaction which are yet to be answered [14].

2.2 Deep Learning in Computer Vision

Neural networks are effectively piecewise linear function approximators, capable in principle of approximating any function through finding hyperplane partitions of the input data during training [49]. The depth of a neural network affects how complex these partitions might grow, in essence improving the representational ability of the whole architecture [49]. Effectively, subsequent layers of neural networks (assuming equal or growing layer size) increase the dimensionality of the latent representation of the input data to the extent where finding some hyperplanar partition to approximate the output values is always possible [49]. This simple yet powerful idea forms the basis of all neural network-based computation, and the associated field of modern deep learning revolves around finding ways of steering these networks to learn more physically relevant, generalizable and interpretable functions during their training process [50]. The simplest example of this steering is the CNN, the backbone of modern computer vision, which connects neurons together in specific ways which bias them toward learning locally relevant correlations (important for, e.g., images).

Such black-box based DL analysis has seen a healthy amount of skepticism in recent years, owing equally to its inherent lack of interpretability and unreliable generalizability beyond its training set due mainly to underspecification [51]. However, many industry veterans argue that this skepticism often reaches unfounded levels in many practical applications; many problems common to deep learning in the past are better understood and can be mitigated through rigorous implementation [52]. There are even accounts of non-mission-critical DL-based algorithms reaching high accuracy never being employed in industry [52], mainly due to a mystical distrust of self-learnt algorithms. Thus, it is often difficult to see whether the error lies in the foundations of DL, or in the faltering diffidence of the end-users who should be benefitting the most. To this dilemma, the efficacy of its remedy lies in its simplicity; the mathematical foundations of DL must be deeper explored and its limitations unequivocally presented. This section is based on presenting these mathematical foundations of (convolutional) neural networks in its proper context; as that of the cornerstone of modern computer vision and the catalyst for the deep learning revolution as a whole.

2.2.1 The Artificial Neuron

Conceptually, deep learning is inspired by the structure and function of the brain. Specifically, the artificial neural network at the heart of all deep learning is explicitly inspired by our own biological neural networks; the conglomeration of neurons, synapses, chemical and electrical signals which make up our brains. In its simplest interpretation, the biological neuron can be approximated as a computational unit which fires a signal of a certain strength if its input signal exceeds a certain threshold. Specifically, the biological neuron "fires" (activates) with strength $w \times x$ if its input signal overcomes threshold b [53]. Every artificial neural network consists of these individual computational units, known as artificial neurons, with a certain weight w and bias b, and activation function a such that each neuron computes the simple linear function

$$z = a(wx + b) \tag{2.5}$$

given an electrical signal of strength x which overcomes a treshold (bias) b, for a given weight w and activation function a [53]. The values of the weights w and biases b are determined during training of the network on some dataset, whilst the activation function is fixed and chosen by the engineer dependent on the nature of the dataset in question [53]. This simple function is the basis of all deep learning-based systems, regardless of their ultimate manifestations' complexity or advanced functionality.

This computational model of artificial neural networks was introduced in 1943 by McCulloch and Pitts [54], and the following decades of developing theory and practice of neural networks is a complex sequence of cycles of growing hype, investment, disappointment and stagnation. This story of the highly nonlinear development of deep learning has been told many times [55]–[57] and will not be reiterated in this thesis. Rather, we note the most relevant contribution to the field of deep learning in the context of computer vision, and arguably, the most significant contribution in modern deep learning overall; the work of LeCun et. al. [7] in introducing convolutional neural networks trained with learnable parameters. This would later form the basis for AlexNet [58], which would go on to convincingly win the long-standing image classification challenge ImageNet and spark the deep learning revolution which is still gaining speed to this day. Following this success, several specially-designed CNN architectures have been introduced for different purposes: most notably for the purposes of this thesis are those developed for biomedical image analysis (see section 3.1.2) [59], object detection [60] (see section 3.1.3) and skip-connection-based networks for enabling deeper networks, i.e. networks with more consecutively connected layers, than previously possible (see section 3.1.4) [61]. To understand the significance of these algorithms in the context of computer vision in broad and microscopy in particular, we need to delve into their precise mathematical formulation.

2.2.2 Convolutional Neural Networks

As alluded to in the introduction of this chapter, the power of deep learning comes from representation (feature) learning. The reason convolutional neural networks, and allegorically all deep learning in the fields in which it is leading a revolution, dominate computer vision is because it learns to represent inherent features of datasets better than handcrafted algorithms. The inductive locality prior, weight sharing and stochastic gradient descent through backpropagation [7] condense this computational power into a tractable and ultimately parallelizable form which enables training neural networks in feasible timeframes with practically relevant data [62].

In their basic manifestation, convolutional neural networks consist of a combination of three types of layers; fully connected neural layers, convolutionally connected neural layers, and pooling layers, the details of which are recounted in the sections below.

Fully Connected Neural Layers

The fully connected neural network layer consists of any number of neurons computing the linear transformation in equation 2.5, where every neuron in the layer is connected to every other neuron in the previous layer (see figure 2.3).

As shown in figure 2.3, the neurons' activation consists of a weighted sum of the inputs to the neuron plus a bias, where the values of the weights and bias are learnt during training. This value, known as the *logit*, is then passed through an activation function (e.g. a sigmoid) before being passed further to the next layer of neurons (or output if it was the last layer). A sequence of such fully connected layers constitutes the simplest version of a deep neural network, and in turn the basis for the simplest realization of deep learning. Even this simple network should not be underestimated, since mathematical theorems [63] prove that such a fully connected network with only a single sufficiently sized layer can approximate any finite-dimensional function to arbitrary accuracy. However, the lack of any inductive bias within the network architecture makes learning a very data-intensive process, and any trained model is highly prone to both overfitting and underspecification (i.e. manifested as unrobust generalizability). One simple way of introducing inductive bias into our models is by, for instance, including memory of past inputs in the neuron activations (recurrent neural layers) or only considering information from inputs which are spatially localized close together (convolutional layers).

Convolutional Layers

Convolutional layers are a special case of the fully connected layer; we remove connections between neurons such that only the inputs which are close together count toward the activation sum. This has shown to be a particularly powerful approach for images, in which information about the position and classification of objects is usually highly correlated within spatially proximal pixels. In practice, this is easiest accomplished through a convolution as shown in figure 2.4.

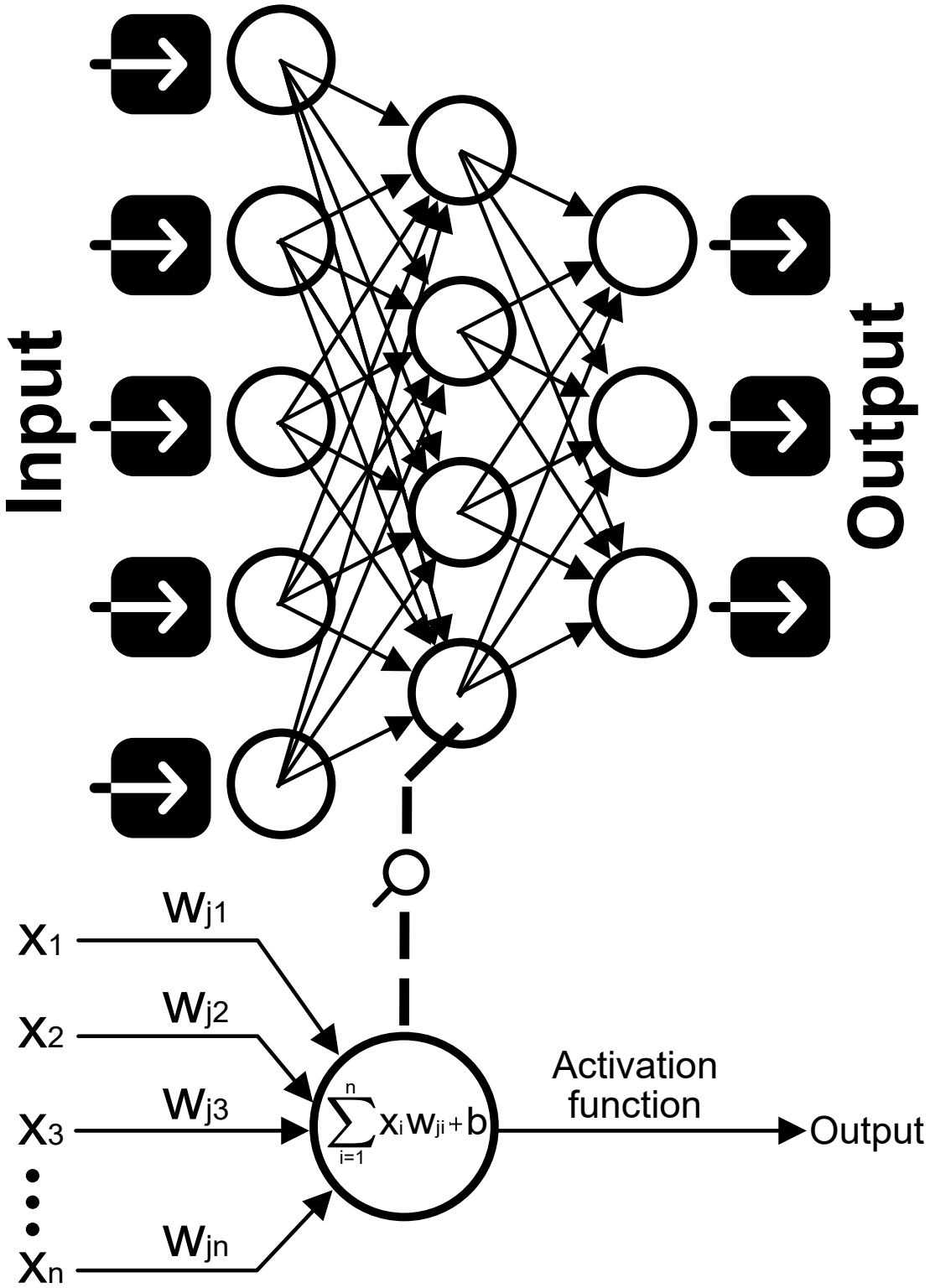

Figure 2.3: Schematic of a fully connected neural network, in which each point x_i in the input tensor is connected through a weight w_{ji} to every neuron n_j in the next layer, as shown in the schematic on the bottom. A sequence of such layers is denoted as a *deep* (and fully connected) neural network.

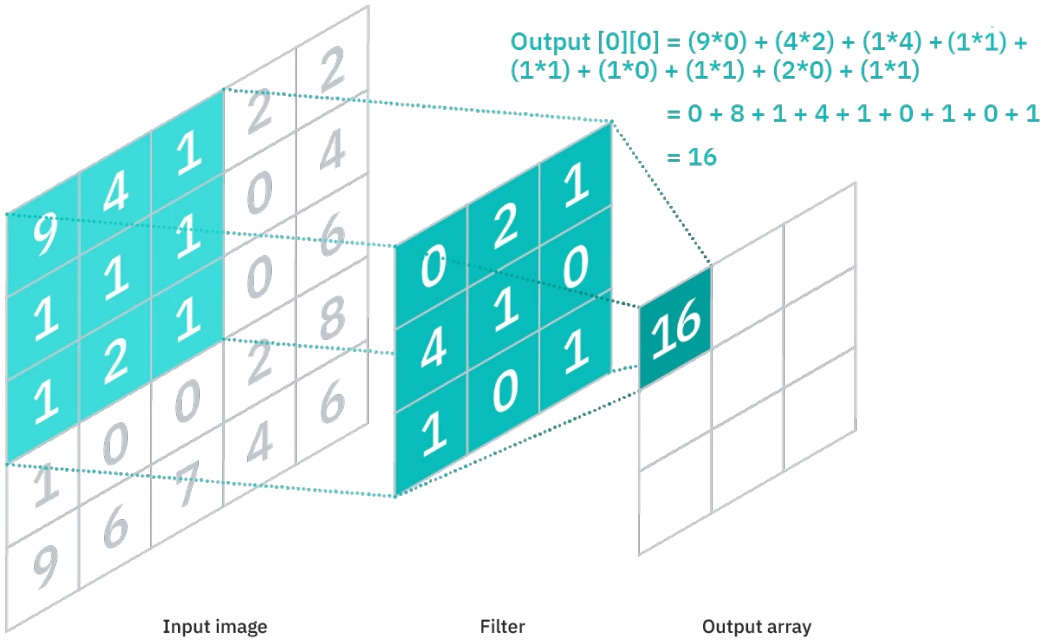

Figure 2.4: Demonstration of the operation of convolution, wherein a segment of size $[ks_1, ks_2]$ of the input image is convolved with a filter of equal size. The operation consists of a dot product between the pixels of the input and filter, whereafter the filter shifts by stride s and repeats the process until the entire image has been convolved. In a neural convolutional layer, the numeric values of the weights of this filter are learnt during training. Figure from [64], reproduced under the Creative Commons license.

The weighting of each pixel in the convolutional layer is learnt during training, effectively applying the neuron's computation in equation 2.5 in a convolution rather than a linear fully connected fashion. When creating the layer, the engineer defines the kernel size, stride (of the sliding window), and number of filters. Filters work essentially as parallel channels all applying convolutions to the same input but with separately learnt weights, effectively developing separate "feature maps" which may represent different relevant features of the input. Thus, using convolutional layers, a full CNN uses various kernels to convolve the whole input into intermediately sampled feature maps, which are in turn convolved in an equivalent manner. Ultimately, this realization of neural networks enables fully rich feature representations of images whilst prospectively fully preserving locality and richly detailed information. Further, it has been shown that early layers of a CNN tend to learn basic features relevant for image analysis, i.e. shapes, edges and simple structures, whereas later layers learn more abstract features such as vehicle outlines and facial details [65].

Pooling Layers

The purpose of pooling layers is to reduce the dimensionality of data between subsequent convolutional layers, reducing the impact of, as coined by Prof. Bellman, *the curse of dimensionality* [66]. In short, as the dimensionality of data increases, the total volume of the information space increases so quickly that any available dataset

becomes sparse, worsening still the problems of underspecification [51]. Hence, the downsampling of dimensionality which pooling layers provide results not only in less computational overhead but also has been shown to improve generalization, latent space featurization and faster training convergence [67].

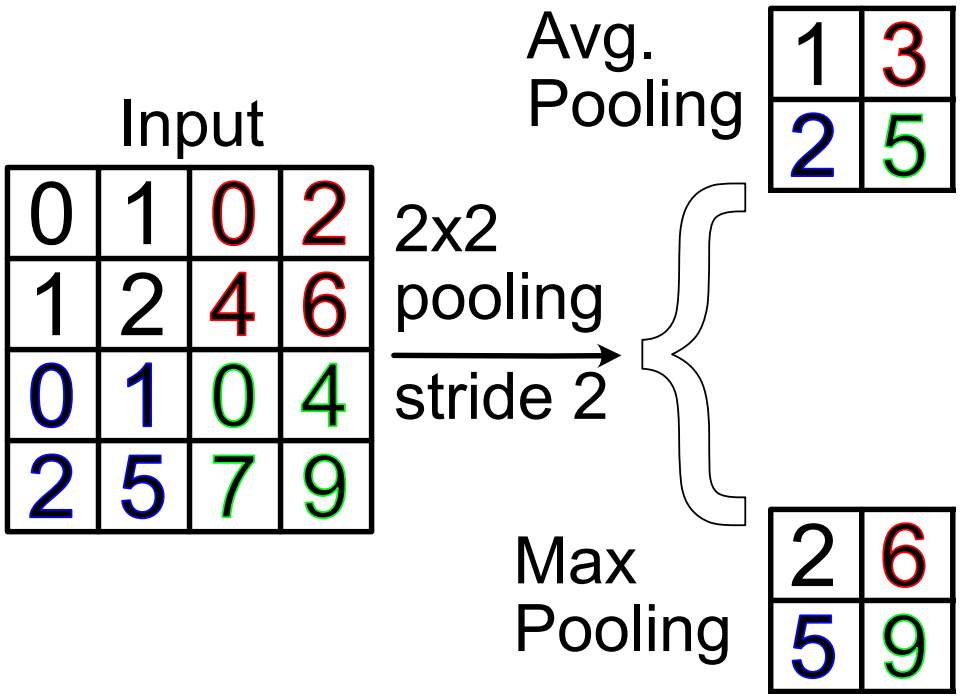

Figure 2.5: Schematic of the pooling convolution, demonstrated here for average pooling on top and max pooling on bottom for a 2x2 convolution passing over the input with a stride of 2. In average pooling, the output of each convolution is the average value within the corresponding pooling window. In max pooling, the output is instead the maximal value within said pooling window.

In figure 2.5, an example of max- and average pooling layers is shown. Note that though these are the most common pooling layers, and used in most standard implementations of CNNs, there are several others [68]. For instance, stochastic pooling for network regularization [69] and spatial pyramid pooling to remedy certain architectures' strict sizing requirements [70].

Convolutional Architecture

Most convolutional neural networks in use today use some combination and configuration of the convolutional, pooling and fully connected layers within their architecture. To reiterate in concrete terms, each CNN architecture uses these layers to accomplish three features: localized receptive fields (convolutional layers), weight sharing, and spatial subsampling. Localized receptive fields and weight sharing enable the

introduction of inductive bias into our models in the form of translational invariance, which promotes the learning of general features relevant for (translationally invariant) imaging data. Each subsequent convolutional layer in a deep convolutional architecture receives inputs from a set of neighbouring units from the previous layer. In practice, this configuration has been shown [8] [65] to effectively allow the correlations learnt in each subsequent layer to grow in complexity; i.e. early layers learn basic abstractions, edges and shapes whereas later layers learn higher-order representational features. Elementary features are often equally important regardless of where they appear in the image; hence local receptive fields with tied weights are appropriate [71].

In a traditional convolutional architecture, each convolutional layer is followed by a pooling layer to reduce the problems related to the curse of dimensionality [66]. Following several sequences of convolutional and pooling layers, the final layer of highest-level reasoning occurring within the network is usually achieved via fully connected layers. The neurons in fully connected layers are connected to every other neuron in its two neighbouring layers, hence their name, and their output is simply computed as an activation function of a matrix multiplication with a bias offset as also described earlier in this section. This generic architecture effectively enables the definition of a function approximation whose input is a data type with locally relevant correlations (i.e. an image) and whose output is a vector of numbers. By connecting this vector of numbers to a regression (i.e. mean absolute error) or classification (i.e. cross-entropy) [58] loss when compared with labelled data, and then *backpropagating* the numeric value of loss through the network's weights and biases in the opposite direction of the gradient, we arrive at a trainable convolution-based network architecture for generic image-to-vector translation. Note that this output does not necessarily have to be the final output of a full model architecture - it could rather be considered a latent feature vector for further processing through another network type or standard methods [72]. Note that it is also possible to control the nature of the output data type by modifying the output layer of the network architecture - i.e. a convolutional layer instead enables learning an image-to-image translation function approximation. Herein lies one of the greatest strengths of neural network architectures in general: changing the architecture to accommodate different data types, i.e. sequences, graphs, images or tabular data, is relatively straightforward with modern powerful deep learning frameworks like Keras [73] and PyTorch [43].

2.2.3 Quantitative Deep Learning for Optical Microscopy

Perception-based tasks in which data are collected under conditions which are consistent across training and inference, are where the greatest strengths of deep learning truly manifest [71]. The difficulties of deep learning; its black-box nature, its oft-brittle performance [74] and fickle unpredictability inferring on datasets orthogonal to its training set due to underspecification [51] and high sensitivity to poorly labelled data [75], are mostly avoided in these applications.

In practice, DL-based algorithms often outperform handcrafted algorithms [10], [76], as the heuristics necessary for complex image analysis are often limited by technical expertise, technological limitations (e.g. processing speed for real-time applications) and ultimately limits of human ingenuity and imagination. Through powerful representation learning, deep learning-based AI systems can learn features more representative and relevant than other approaches [50].

Intuitively, this may be understood as a result of the complexity inherent in nature being difficult to capture in simplified heuristics or theoretical models alone, whereas data-driven DL approaches can learn directly from minimally processed measurements of physical systems only with limited inductive bias. Thus, DL systems trained in a supervised fashion can be interpreted in practice as an opportunity to draw direct statistical correlations between datasets without the potentially limiting intervention of researcher bias, an oft-cited reason for why promising yet unfashionable theories struggle to gain ground in physics [41]. In this section, I describe a few uses of CNNs in practice, and particularly how they have proven useful for quantitative analysis within optical microscopy.

Semantic Image Segmentation

The purpose of (semantic) image segmentation is to convert the pixel-wise information in an image to the class of the object in the pixel with a given probability. In classic image segmentation, this often reduces to outputting the probability in each pixel containing any of a pre-defined set of classes, as exemplified in figure 2.6a.

For purposes of optical microscopy, the benefits of this application quickly becomes clear; given a complex image of several cellular components and different types of biological molecules, we can output a probability in each pixel of that pixel containing a molecule or component of a certain class, as exemplified in figure 2.7. Thus, the complexity of the image is drastically reduced, enabling biologists to analyze the data in a far faster and more intuitive manner. This is especially relevant in the field of biomedical image diagnosis [78], which is a pioneering field of applied DL, and an example of a human-in-the-loop AI system where human experts and AI work in symbiosis to accelerate experts' work. In particular, this approach forms the backbone of cross-modality transformations in optical microscopy, wherein one aims to employ DL-based algorithms to virtually transform images taken with one modality into another [11]. For instance, the virtual staining [11] of cellular components allows one to bypass the manual chemical staining of patient samples, prospectively saving time when diagnosing patients under the effects of anesthesia. This is further enabled by the property of semantic segmentation architectures being able to predict not only the likelihood of existence of objects of interest in an image, but also the likelihood of said images belonging to a certain class along with other properties of interest.

Arguably, the most effective and influential CNN architecture in this field is the U-net [59], whose technical details are elaborated further upon in section 3.1.2.

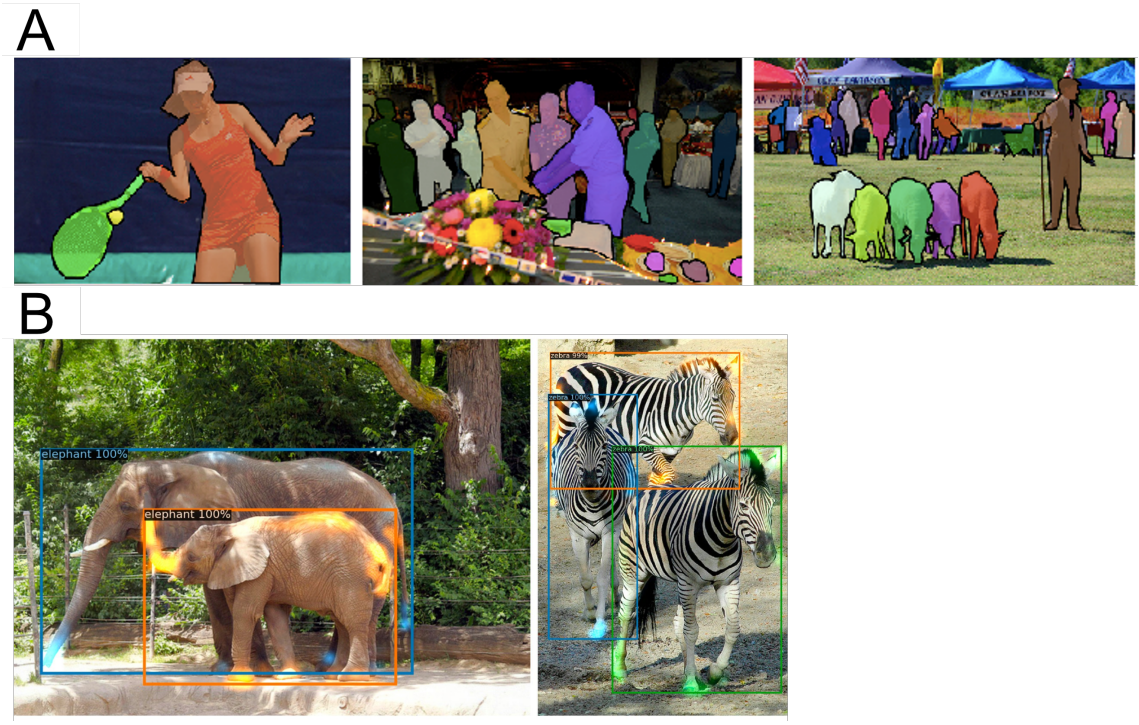

Figure 2.6: **a,b** Examples from Common Objects in Context (COCO) image dataset [77]. **a** shows the result of semantic segmentation, wherein each image is "segmented" using DL such that every pixel of each separate object class is coloured in with a separate colour. **b** shows the result of object detection, wherein each separate object is localized with a so called "bounding box" with a corresponding probability, and classified with a separate corresponding probability.

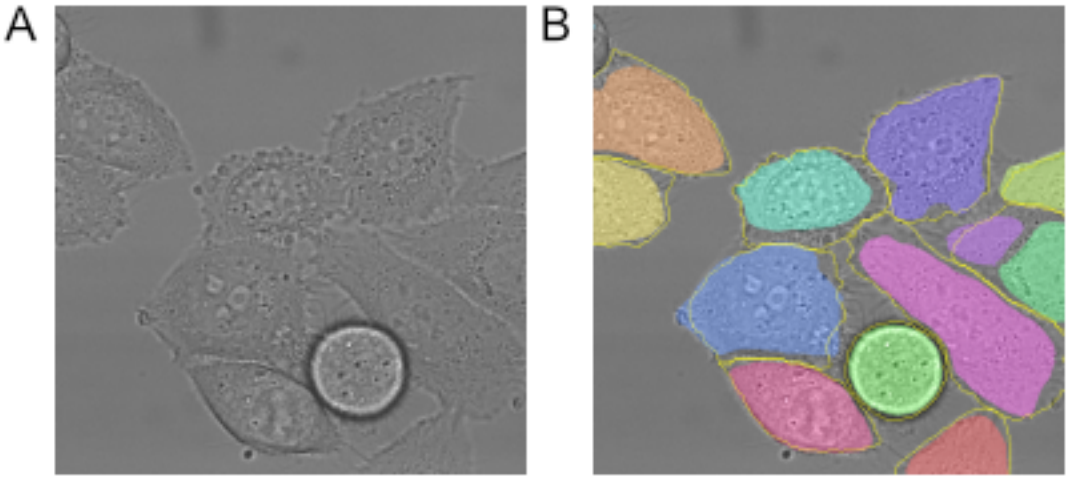

Figure 2.7: A) Examples of HeLa cell microscopy images. B) Equivalent segmented images, where different types or structures of cells may be identified and coloured in different colours, through a U-net architecture. Figure reproduced with permission from [59].

Object Detection

Object detection is a massively popular application within machine learning, whose popularity and influence only seems grow every year through a wide variety of advances within computer vision. For instance, great strides have been made in facial detection [79], pedestrian detection [80] and autonomous driving [81] in the past few years alone. One of the most popular object detection algorithms leading these recent developments is that of You Only Look Once (YOLO), due to its impressive speed and higher generalizability than many other competing algorithms [82]. These traits are a consequence of the method's namesake, where a single feed-forward pass of an image through the CNN backbone is enough to predict (singular) bounding boxes and class probabilities for each object in said image [82], as is further elaborated upon in section 3.1.3.

The purpose of object detection is to classify, localize and separate individual objects within an image. In object detection applications, the exact classification of each pixel is not important. Rather, each image consists of an unknown number n of objects $o_1, o_2, ...o_n$ pertaining to class c_o from a pre-defined set of classes. A classic object detection algorithm trained in a supervised manner, can never output that an object belongs to a class outside its pre-defined list. Its function is exemplified in figure 2.6b, where the benefit of object detection becomes clear; a segmentation model could not identify that the two elephants in the bottom left figure are separate objects. Rather, it would fill in each pixel with a high probability of 'elephant', and contain no quantitative information on the number or position of the separate elephants. Object detection is useful in localizing objects within optical microscopy when pixel-wise segmentation information is unavailable or infeasible to extract [84], as exemplified in figure 2.8.

Quantitative Property Estimation

It should be noted that whilst object detection and semantic segmentation are objectives which have come to dominate computer vision applications, convolutional neural networks can in principle learn to output any property of interest, as long as the property can be properly mathematically defined and the relevant information and correlations exist in the dataset. In general, these property calculating networks work using the same fully convolutional CNN backbone structure as for semantic segmentation and object detection, with the significant difference only manifesting in the data type which the *neck* and *head* layers of the neural network output. In particular, the last layer of the neural network, the head, is typically a dense layer with continuous outputs connected to a regression-based loss.

One of the most effective and influential CNN architectures for this application is the ResNet [61], whose deep skip connections between non-subsequent layers enabled deeper network architectures than previously possible, and whose subtle yet innovative change to calculate residuals rather than direct output value enabled numerically accurate calculations. This architecture inspired that of the biomolecule characterization network described in section 3.1.4 below.

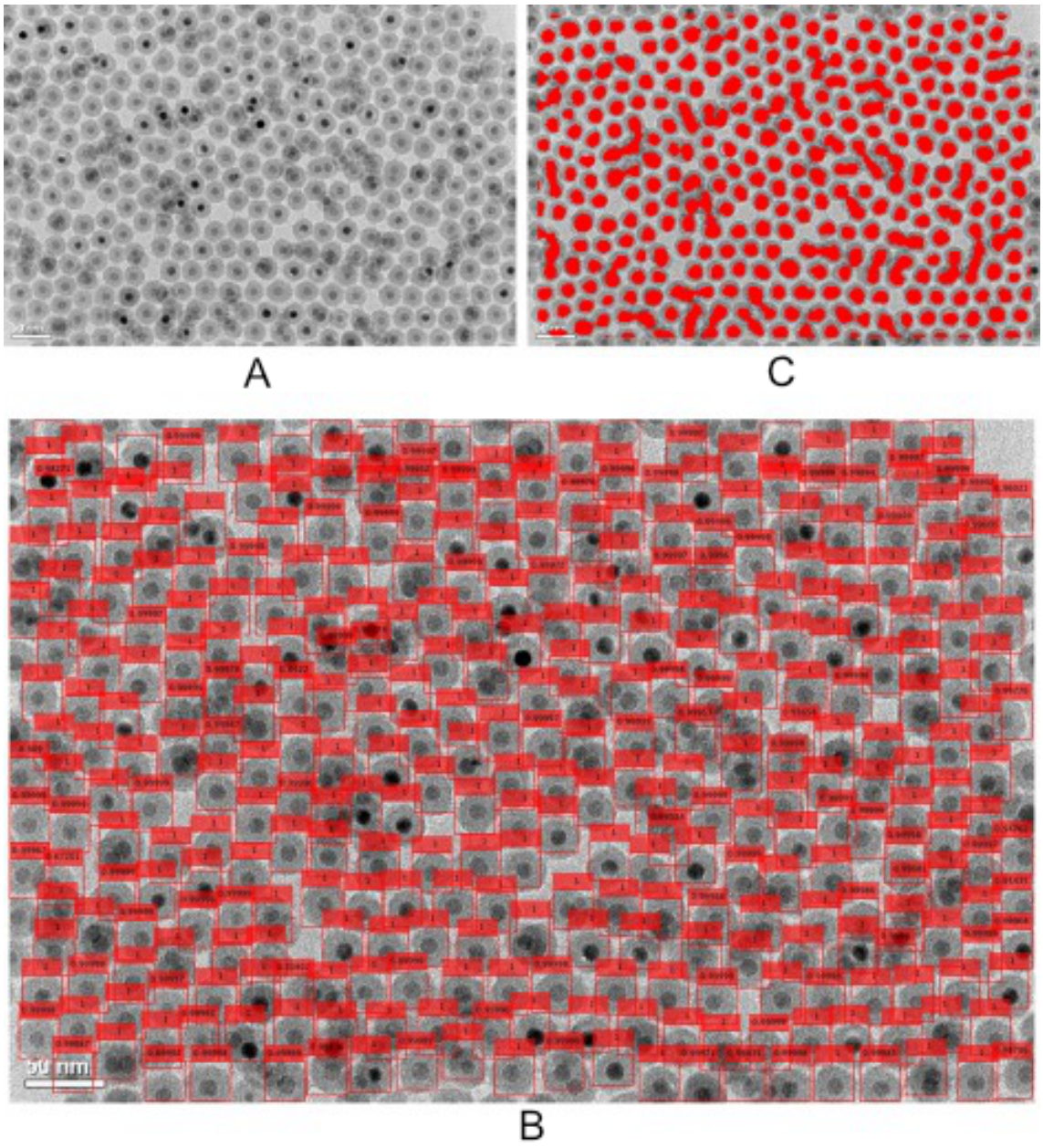

Figure 2.8: A) Raw unprocessed image, consisting of nano-particle cells. B) The boundaries automatically detected by an object detection network, where each cell is identified with a so-called "bounding box". C) Centers of the detected cells, where the colour of each detection corresponds to the prediction accuracy of the corresponding box. Figure reproduced with permission from [83].

The most important takeaway from this section is that despite the differing applications, the underlying structure of the CNN remains the same. Connections between layers vary - between upsampling and downsampling layers in the U-net and between residually connected layers in the ResNet, but the backbone of convolutional and pooling layers in sequence is consistent.

2.2.4 Beyond Supervised Learning

So far, the contents of this thesis have been constrained to discussing the particular paradigm of machine learning known as supervised learning, wherein a neural network is used to approximate a function between input x and output y. However, this is likely not where the most significant future strides in deep learning will manifest. Though it is a powerful paradigm, it is significantly limited in the sense that it cannot automatically learn relationships between objects unless specifically pre-defined by their engineer. In particular, they cannot recognize new types of objects not pre-specified within the model definition, or for instance the relationship between words in different languages as is highly relevant for modern natural language processing [85]. Indeed, many of the current issues associated with DL are in practice rather a consequence of supervised learning [86]. Other paradigms of AI, such as unsupervised, self-supervised, generative and reinforcement learning, show much greater robustness beyond their training set and greater potential with smaller datasets (i.e. single-shot [87] or few-shot learning [88]). Therefore, to serve as a backdrop against the future outlook for further development of the work conducted within this thesis, and to introduce the reader to some of the most promising avenues of current AI research, this section is dedicated to describing paradigms beyond the standard of supervised learning.

Most straightforward to describe is unsupervised learning, wherein AI algorithms which do not require any labelled data (i.e. output y) are of definition. Rather, they automatically discover patterns and correlations only in input data x on their own during training. Classes of algorithms which are encompassed by this umbrella are e.g. clustering algorithms such as K-means clustering [89], which automatically cluster input data in a predefined number of clusters through appropriate parameters chosen by the network during training, and auto-encoders which are often used as anomaly detection models [90]. Most famously, some of the statistical analysis underlying the discovery of gravitational waves [91] were based on this approach: in short, an auto-encoder is trained to perfectly recreate the noise present in the LIGO interferometer device used to measure gravitational wave signals. The portion of LIGO device output which cannot be perfectly recreated by a such trained model is therefore either some entirely new source of noise (which thus can be corrected for with a new model) or quantitatively accurate gravitational wave signal down to a very low SNR [92].

Another promising paradigm is that of self-supervised learning, wherein we exploit the inherent symmetries in certain datasets to effectively teach a neural network-based algorithm to learn certain useful transformations within unlabeled datasets.

For instance, the most influential use of self-supervised learning today is in that of training massive so-called attention-based transformer models, effectively neural networks particularly tuned to learn relationships between input tokens, on natural language processing. In short, one scrapes a large dataset (e.g. the entirety of Wikipedia), randomly removes singular words with some defined frequency, and trains the transformer model on filling in the missing words. Such models, when trained on billions of data-points and containing billions of parameters, have been shown to be able to output human-like text in form of essays, blog posts and even scientific articles [93], [94]. Another example is the Low-shot deep Symmetric Tracking And Regression (LodeSTAR) [87] algorithm, which essentially learns to detect particles in microscopy images by training on a single crop of a single image of a single particle. It achieves this by applying several (randomly modified) affine transformations to said single image, and training a convolutional neural network to always recognize to identify the same particle despite all possible affine transformations and corruption schemes. Thus, the CNN learns robust features which allow it to always track the profile of the particle type in different imaging conditions and self-consistent transformations. Further, the natural continuation of the supervised object detection algorithms presented in this thesis is self-supervised vision transformers such as DINO [95], which can learn to identify and separate objects into an arbitrary number of unseen and undefined classes, thus providing a solution to one of the biggest issues with supervised object detection algorithms.

Reinforcement learning (RL) is another paradigm which has seen significant medial attention in recent years, most notably for its use in developing AI-based agents which can learn to play games to superhuman levels, notably beating grand-master Lee Sedol in Go [96] and eventually going on to beat grandmasters in inordinately complex video games such as Starcraft II [97] and Dota 2 [98]. Its use in real-world problems has of yet been underwhelming, owing to the difficult "sim2real" [99] gap associated with bringing algorithms trained on purely simulated data to work in the real world, as well as the difficulties associated with training RL-agents in the real world [100]. However, significant strides have also been made in this regard, including an agent which develops new floorplan designs for computer chips at orders of magnitude higher speeds than human engineers [101] and a fully autonomous robot which plans, executes and analyzes experiments on its own [102].

Finally, and of particularly important note for the field of optical microscopy and biomedical imaging, generative learning is a paradigm which enables AI to generate new datasets given some original training data set. This may at first stand in great contrast to the aforementioned paradigms where the goal is to find correlations within and between datasets and optimize for a loss(reward) function, but note that the function to optimize may well be defined as that of how to generate unique data indiscriminate from some original dataset. Said generative learning algorithms are able to create quantitatively accurate representations for complex objects and patterns in data, which can then be used for tasks such as image synthesis [103], text generation [104], and cell microscopy image generation [105] in combination with e.g. supervised, unsupervised or self-supervised learning. This essentially solves the

issue of requiring exceptionally large datasets to train certain models, if only one can train a generative model to generate effectively infinite synthetic data.

Generative models are typically trained in a Generative Adversarial Network (GAN) environment [106]; a system composed of two neural networks competing against each other in a zero-sum game. The first network, called the generator, tries to generate fake data that is indistinguishable from real data. The second network, called the discriminator, tries to distinguish between real and fake data through a binary value. The generator gets better at creating realistic data as it learns from feedback given by the discriminator.

2.2.5 Cross-modality Transforms in Biological Microscopy enabled by Deep Learning

As reviewed in Paper III of this thesis, one promising emerging topic in artificial intelligence in the context of microscopy for biological (and non-biological) applications is that of cross-modality transformations enabled by deep learning. In short, this entails the process of training a neural network architecture to transform images taken with one microscope modality to instead contain information as if it were taken by another modality. One example of this is histological staining, wherein we circumvent the time-consuming process of manually chemically staining tissues by transforming e.g. brightfield images of tissues to their chemically stained counterparts through generative learning (see figure 2.9).

Similarly, another virtual staining cross-modality transform is that of using generative learning to fluorescently stain cellular and sub-cellular structures. In pharmaceutical drug screening and digital cytology, quantitative analysis of cell structures is essential: changes in the morphological properties of cell structures are effective readouts into the physiological state of a cell culture and its response under drug exposure [107], [108]. However, chemically staining samples with fluorescent dyes to make them available for study entails invasive and sometimes toxic procedures, potentially affecting cell health and behavior [109]. Additionally, phototoxicity and photobleaching, occurring during the acquisition of fluorescence images, limit the time scales available for live cell imaging [110].

To remedy these challenges, recent works have proposed the use of DL, and particularly generative learning, as an alternative to mitigate the inherent problems associated with conventional chemical staining [111]. In these works, chemical-staining and fluorescence microscopy are replaced with a virtual staining network (generally a variation of the U-Net architecture, see section 3.1.2) that generates virtual fluorescence-stained images from unlabelled samples. The input to the staining network must encode insightful contrast of the various cell structures, providing the network with sufficient information to learn the transformation to the desired high-contrast, high-specificity fluorescently stained samples. Virtual staining of cells and cell structures has been achieved from various imaging modalities, including phase-contrast imaging [112]–[114], quantitative phase imaging [115], and holographic

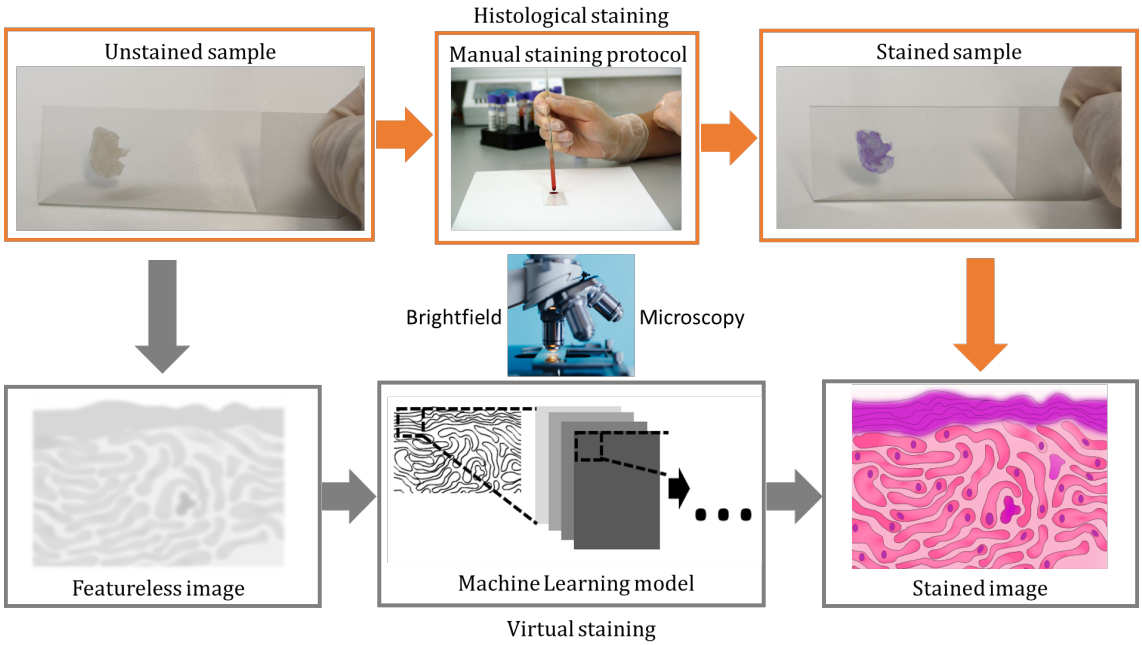

Figure 2.9: Illustration of the process of (conventional) histological staining in orange and DL-based virtual staining in gray. As outlined in Paper III, one can overcome many of the limitations of chemically staining histological tissues by training a generative model to virtually stain (e.g. brightfield) microscopy images of tissues, given a dataset of unstained and correspondingly chemically stained tissues.

microscopy [116]. Moreover, some work [11], [110], [117] suggests that bright-field images contain the information necessary to reproduce different stainings, despite the detail of these images being largely limited by diffraction. Another way of overcoming said diffraction limit is by so-called *super resolution* techniques, which are another example of cross-modality transforms recently enabled by DL, as exemplified in figure 2.10.

Many microscope techniques have been developed to overcome the diffraction limit, including single-molecule localization microscopy (SMLM) methods such as stochastic optical reconstruction microscopy (STORM) [120], photo-activated localization microscopy (PALM) [121] and fluorescence photoactivation localization microscopy [122]. Yet other methods of transcending the standard resolutions of microscopes exist, as mentioned, in the complex numerical estimations of point spread (transfer) functions (PSF) which seek to estimate the diffraction behaviour, in illumination pattern engineering methods which decrease the PSF size [123], as well as specialized flurophores [124]. These approaches come with their own limitations in terms of the complex and multivariate dependency on imaging conditions which make solving the diffraction integrals of the PSF very difficult for practically relevant systems [125], and increased costs associated with aforementioned flurophores [12].

In recent years, another promising paradigm of super resolution has emerged as a consequence of the astounding growth and success of DL-based computer vision

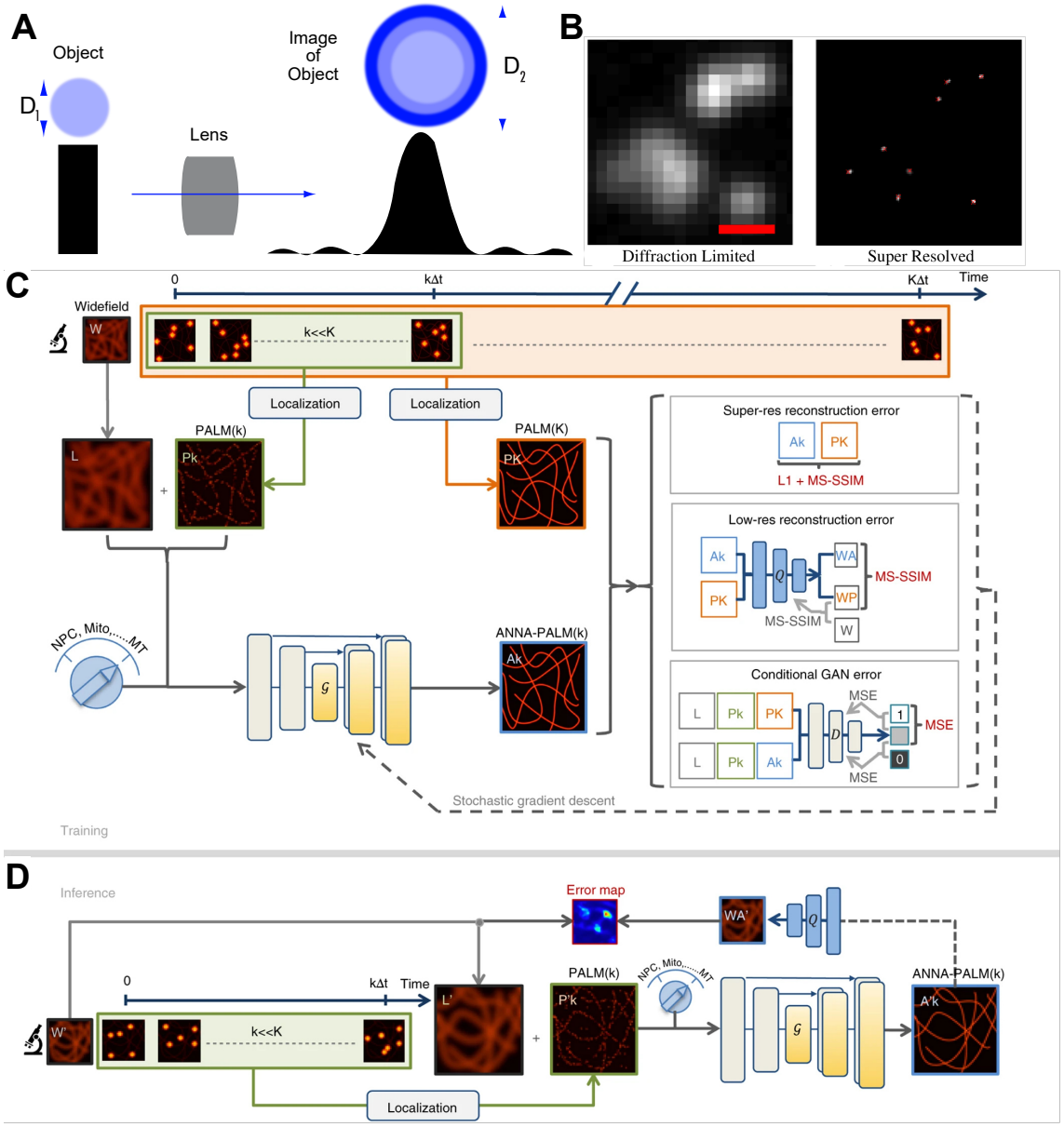

Figure 2.10: A) Illustration of the PSF resultant of imaging objects below the diffraction limit of an optical system. B) Low-resolution image of simulated emitters, scale bar is 0.5 μm, and a CNN's prediction of ground truth emitter locations. Image reproduced with permission from [118]. C,D) ANNA-PALM consists of two stages: firstly, acquisition of training images using standard localization microscopy (PALM) followed by neural network (ANN) training, and secondly reconstruction of super-resolution views and low-resolution error maps from new sparse PALM and/or widefield images at inference time. C) Training images are obtained by acquiring one or a few long sequences of diffraction limited single-molecule images, as in standard PALM experiments. Thereafter, standard localization microscopy algorithms are used to generate super-resolution images. A switch (light blue) can be set to distinguish different types of structures, for example, nuclear pore complexes (NPC), mitochondria (Mito) or microtubules (MT). A GAN generator is trained by using the sparse PALM images Pk as inputs and the corresponding dense PALM image PK as target output with the three error functions in the gray bordered boxes (see [119] for details). D) At inference time, a short sequence of low-resolution single-molecule images are acquired. Standard localization algorithms generate a sparse (undersampled) PALM image (P'k). This sparse image P'k, the upscaled widefield image L' and switch setting are fed as inputs to the trained generator, which outputs a super-resolved ANNA-PALM image (A'k). Images C,D reproduced with permission from [119]

algorithms. Allegorically to the cross-modality transforms above, the DL approach to super-resolution revolves around training neural networks to transform one imaging modality (regular-resolution images) to another (super-resolved images). Some of these approaches are through generative learning, i.e. effectively learning the complex interpolation function between regular- and super-resolved images, or through direct supervised learning e.g. by estimating the positions of underlying diffraction-limited emitters. The details on how these networks are trained in practice vary widely between applications, as elaborated upon further in Paper III. Of course, one would be remiss in this context not to mention that another way of effectively transcending the diffraction limit and studying individual molecules is to use powerful scattering-based microscopy methods like iSCAT or NSM.

Chapter 3

Methods

This chapter is dedicated to introducing and describing the implementation of the deep learning-based characterization methods used in Paper I and II. Specifically, the first section of this chapter first gives an overview of the entire DL-based analysis pipeline as it is implemented in Paper I, and thereafter delves deeper into the technical details of each separate method within. The second section describes some of the methods of training the networks which were crucial to reaching their final level of performance. Finally, the chapter ends with a description of how the biomolecule trajectories used as training data were simulated.

3.1 Deep Learning for Biomolecule Characterization

To track and analyze single biomolecules within nanochannels, which are the core of the NSM method, we implemented a DL analysis pipeline which takes advantage of all the paradigms of CNN-based CV described in 2.2. Image segmentation is used to identify the pixels in an image (kymograph) which contain the scattered light from the biomolecules within the channel. In fact, using a specialized network architecture developed for biomedical image segmentation known as a U-net [59], each output pixel corresponds to the *probability* of said pixel containing scattered light from a particle. Further, object detection is used to identify the separate trajectories and segment the kymographs into sections consisting only of single trajectories. Finally, a fully convolutional neural network is used to calculate the properties of interest - in principle any biomolecule property whose quantitative values can be derived from its trajectory - for each individual trajectory in turn. This pipeline is illustrated in figure 3.1. It is comprised of five parts:

3.1.1 Pre-processing into Kymographs

The raw image data (figure 3.1A) was pre-processed to transform it into kymographs (figure 3.1B), by the following process. First, the intensity of the raw CMOS camera

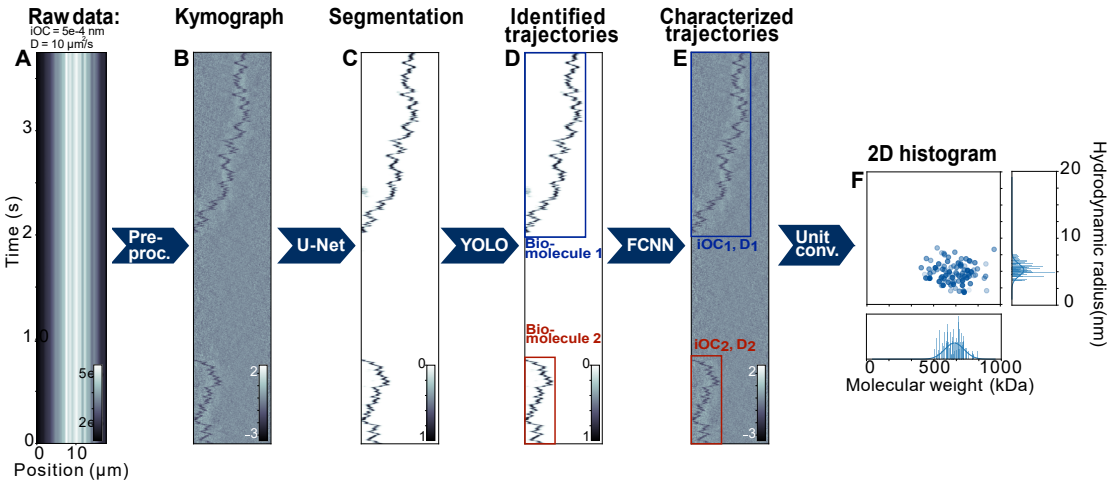

Figure 3.1: Machine learning (ML) analysis workflow. (A) Raw image data simulated for a single biomolecule with optical contrast $iOC = 0.5$ nm and diffusivity $D = 10\ \mu m^2/s$. (B) Preprocessed kymograph with removed background. (C) Segmented image where the particle positions are detected using a U-net. (D) Single biomolecule trajectory identification using the YOLOv3 neural network. (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, for a simulated example molecule. Figure reproduced from [24].

image data was normalized according to

$$\bar{I}(x,t) = \frac{I(x,t) - \langle I(x,t)\rangle}{\langle I(x,t)\rangle} \tag{3.1}$$

where $I(x,t)$ is the intensity at position x and time-frame t, and $\langle I(x,t)\rangle$ represents the time average of said intensity. Second, a low-pass-filtered version of $\bar{I}(x,t)$ was calculated by using two normalized sliding windows of sizes 200×1 and 1×200, and subtracted from $\bar{I}(x,t)$. Finally, to obtain the kymographs used to calculate iOC, the resulting image was down-sampled by a factor of 4 in the length dimension through mean pooling. Instead, to obtain the kymographs used to calculate D, the image was normalized by its standard deviation before being down-sampled by a factor of 4 in the length dimension through mean pooling.

3.1.2 Image Segmentation

The purpose of image segmentation, as described in 2.2.3, is to transform the value of each pixel within an image to another property which is more conducive to the purposes of one's analysis. For the purposes of particle tracking, we want to transform our kymographs, consisting of the (differential) intensity of scattered light in each pixel, to "segmented" images which contain the probability in each pixel containing scattered light from a particle. This transformation is accomplished through implementation of a particular CNN architecture, whose function is described in section 2.2.2 and architecture summarized in figure 3.2, trained in a GAN environment as

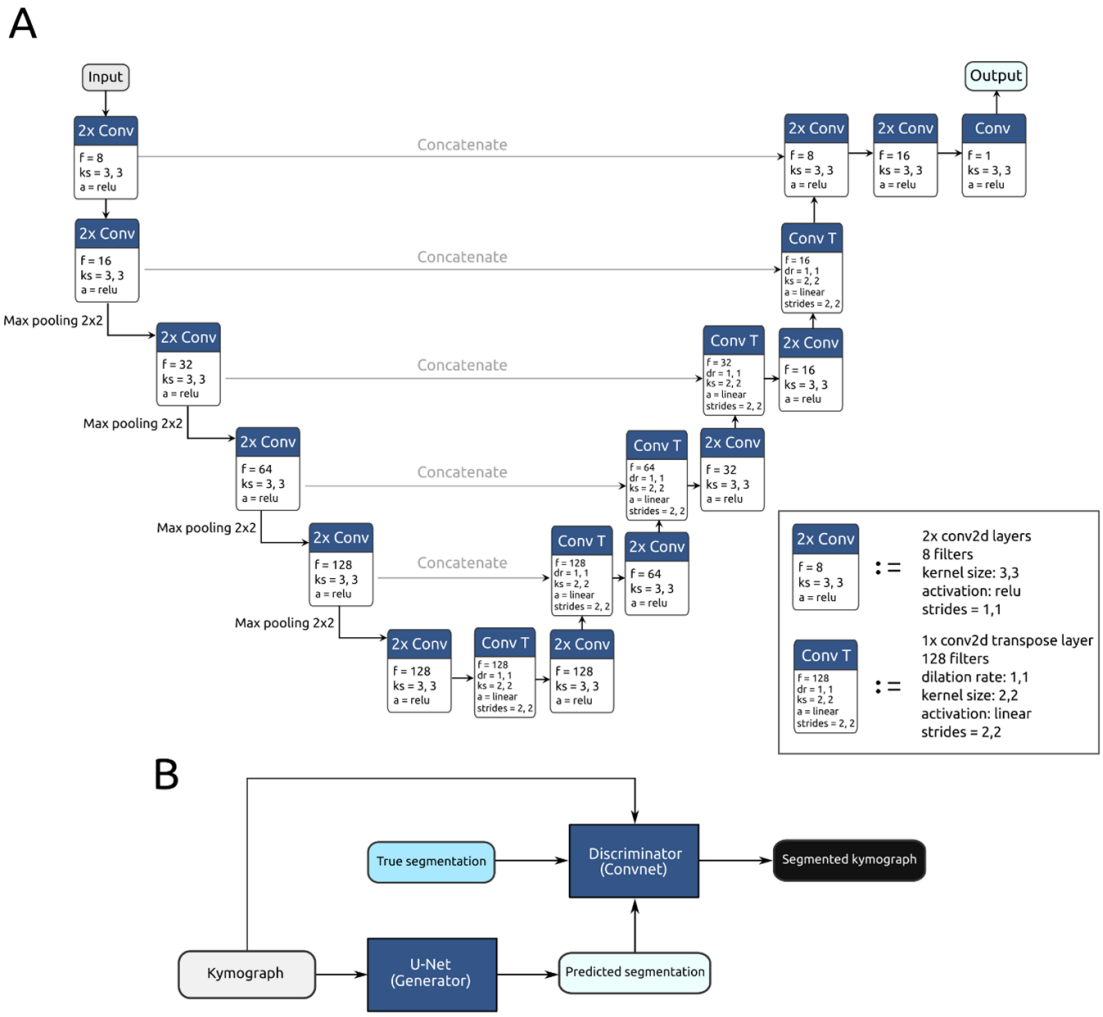

Figure 3.2: The U-net[59] consists of a series of contraction convolutional layers, a bottleneck, and a series of expansion convolutional layers, as well as a series of skip connections between corresponding contraction and expansion convolutional layers to ensure that information learnt during contraction is not lost at the bottleneck. Each 2x Conv box represents a convolutional block corresponding to 2 convolutional layers in sequence, and each Conv T box represents a single convolutional transpose layer as exemplified in the legend. Here, f is the number of filters in each block, ks is the kernel size, a is the activation function, and dr is the dilation rate. The network is visualized with Netron [126]. (B) Block diagram of the GAN training environment, where kymographs are fed both to a U-net generator network, which predicts a corresponding segmentation, and a Convnet (convolutional network) discriminator network, which takes both original kymographs and true segmentations as input to determine whether the predicted segmentation is correct. A basic Convnet consisting of 5 convolutional layers of size 4x4 and stride 1 connected to a single dense layer is used as discriminator network. Figure reproduced with permission from [24].

described in section 2.2.4. Specifically, the GAN consists of two networks: a generator U-net which generates segmented images from input kymographs (pre-processed as described above), and a discriminator ConvNet which outputs a binary value whether the generated segmented image corresponds to the true trajectories or not. By modifying the architecture of this ConvNet, different behaviours of the U-net can be achieved. For instance, a relatively strong architecture which considers multiple scales of the input image, including filters which convolve the entire size of the input kymographs, is useful to force the U-net to always draw fully connected trajectories (broken trajectories are always unphysical, unless they are broken at the borders). Conversely, a relatively weak architecture, whose filters consider only local features of the image, can be used to force the U-net to only output trajectories in regions of the image where the local SNR is high. Generally speaking, the former approach effectively results in "ghost trajectories" in kymographs with heavy noise, as the generator U-net will always draw fully connected trajectories even in regions where the signal is low enough that it should be impossible. In these localized regions, the network tends to "guess" the likely trajectory path. The latter approach usually results in broken trajectories, since the generator will prefer to draw trajectories only where the signal is sufficiently high. The ConvNet used in this thesis was comprised of 5 convolutional layers of filter size 4, effectively corresponding to a compromise between the two aforementioned approaches.

The U-net was trained on ∼300,000 simulated kymographs (generated as described in section 3.2 and pre-processed as described in section 3.1.1) for which the corresponding ground-truth single biomolecule trajectory is known (using the ADAM optimizer [127]) with particle trajectories in the ranges $10^{-5} \leq iOC \leq 3{\cdot}10^{-3}$ μm, $1 \leq D \leq 100$ μm^2/s. The input during training are simulated images (kymographs) of size 128×2048 with a random number of trajectories and output segmented images of equivalent size. The model is also train-validated every 120 epochs ($\cong 30,000$ simulated kymographs) against 150 simulated kymographs (of size 128×512, 128×1024, 128×2048) with experimentally measured channel noise, using an 80-20 train-validation split.

3.1.3 Object Detection to Identify Trajectories

The next stage of analysis is object detection, to accomplish which we implement the YOLOv3 [60] (figure 3.3) algorithm in PyTorch [129]. The purpose of this stage of analysis is to identify and separate individual biomolecule trajectories in a segmented kymograph output from the U-net described in section 3.1.3, which only contains information of the probability in each pixel that it contains the scattered light from a biomolecule. If a kymograph cannot be separated in this manner (for instance as a result of two biomolecule trajectories strongly overlapping each other), the YOLO algorithm instead counts the number of inseparable single biomolecule trajectories in the given kymograph, outputs this value, and disregards the inseparable trajectories as the errors in scattering resultant of particle interactions is not taken into account in any calculations. Concretely, the YOLOv3 algorithm fits a minimally-dimensioned rectangular bounding box around each separate single biomolecule trajectory, and if the minimal bounding boxes of two single molecule trajectories overlap more than

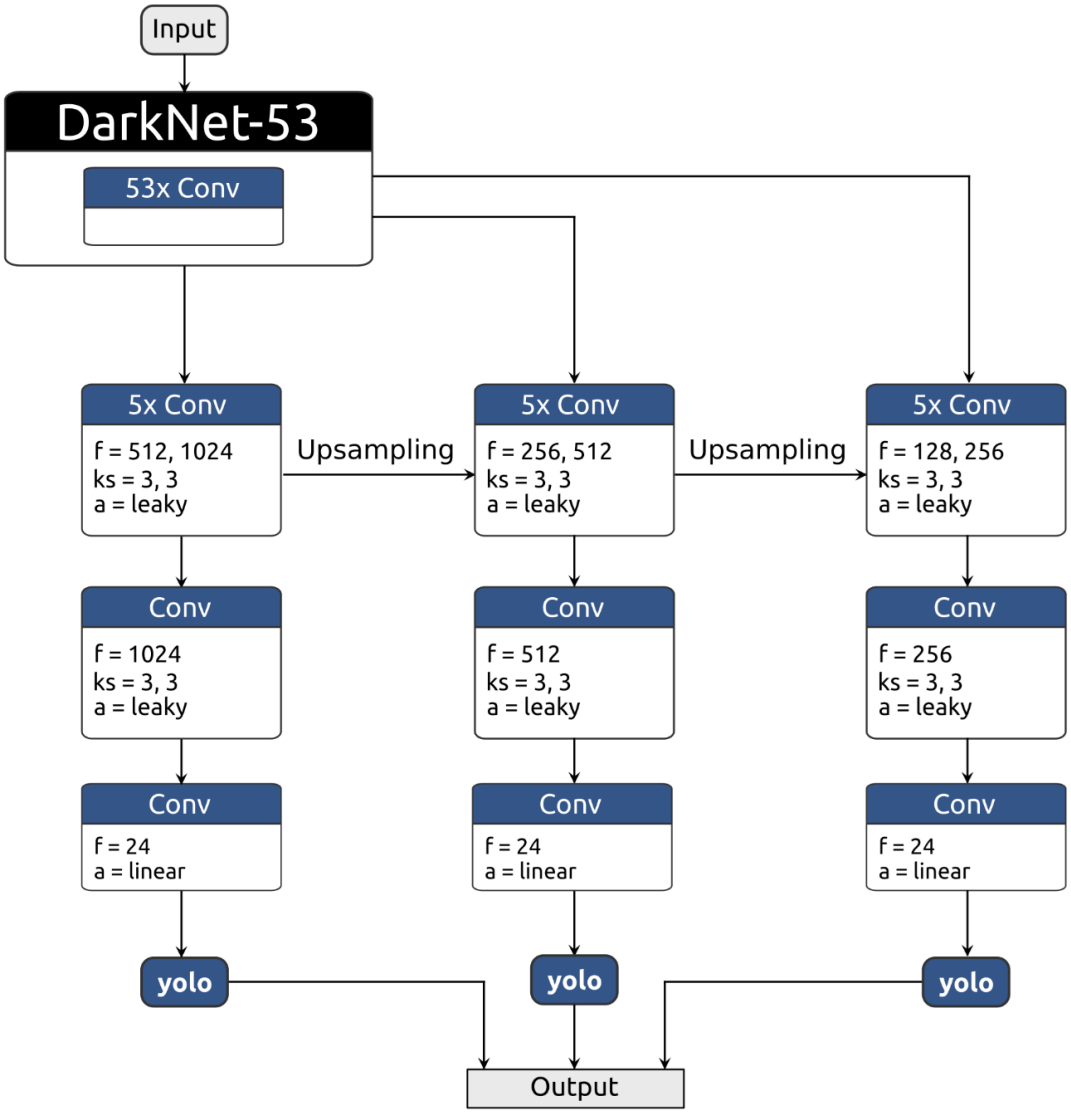

Figure 3.3: The YOLO algorithm is built upon a DarkNet-53 backbone [128], which consists of a neural network architecture comprised of 53 convolutional layers. This backbone feeds into three necks, which consist of a sequence of 5 convolutional layers, which in turn feed into corresponding heads, which output bounding boxes at three different scales through yolo detection layers [60]. Each neck is up-sampled and fed into its neighbouring neck, to make sure the last yolo detection layer benefits from information collected at every previous stage. Here, Conv represents a single convolutional layer, f is the number of filters in each block, ks is the kernel size, and a is the activation function. The network is visualized with Netron [126]. Figure reproduced with permission from [24].

the set threshold $b_t = 60\%$ (roughly corresponding to the point where they are more entwined than not), their bounding boxes are merged to a minimal rectangular bounding box around both of them that is labelled as containing two biomolecules. This process runs recursively until all boxes are either combined or separated.

The YOLO algorithm is usually built upon a pretrained backbone, such as a ResNET [130], GoogleNET [131] or in the case of this implementation, a DarkNET [132]. In particular, we use the implementation known as DarkNet-53, its namesake arising as a result of its architecture consisting of 53 convolutional layers. This backbone is then attached to its *neck*, whose purpose is to form features from the base convolutional neural network backbone. The YOLOv3 algorithm uses something akin to a Feature Pyramid Network (FPN) as neck, adding a few convolutional layers in sequence which connect into an eponymous yolo (detection) layer, which outputs a tensor containing bounding boxes, the probability of an object being inside each bounding box and the probability of the object being of a certain class. The feature map from the second prior layer (i.e. two layers back from the detection layer) is upsampled and subsequently merged with a much earlier layer in the network through concatenation, allowing the network to learn semantic information from the upsampled feature map and still retain information of finer details from the early feature map. At this step, a few more convolutional layers are added and another yolo layer makes new predictions, but at a different scale (proportional to the upsampling factor). Through yet another concatenation with an early feature map and upsampling, this process is repeated a third and final time into the third and last yolo layer, which benefits from computations from the backbone DarkNet network and all previous yolo layer computations. Thus, the YOLO algorithm makes detections at three different scales, where each scale of detections make use of information from prior layers detecting at larger scales, effectively resulting in quick and highly accurate detections of generic objects in images.

The YOLOv3 algorithm was trained on simulated kymographs, generated in the same way as for the U-net (described in section 3.2), but where the input is a perfectly segmented kymograph and the outputs are minimally sized bounding boxes around each separate trajectory. Trajectories are considered separate in the training data if they are more than $T/16$ frames apart, where T is the total amount of frames in the kymograph. Images are continuously generated during training, such that each subsequent image is entirely unique and the risk of overtraining is eliminated. The algorithm was trained using approximately 2 million such unique segmented images, where each image was generated with a diffusion in the range $1 \leq D \leq 100$ $\mu\text{m}^2/\text{s}$ (the optical contrast is not relevant here because we employ perfectly segmented simulated images) with the standard YOLO error function [128] using the ADAM optimizer [127] with a learning rate of 0.001. The input during training are simulated images (segmented kymographs) of size 128×8192, down-sampled to 128×128 to improve performance, and the output is a list of YOLO-labels containing class, position, probability of occurrence and probability of class of each trajectory in the input image.

3.1.4 Characterizing Trajectories

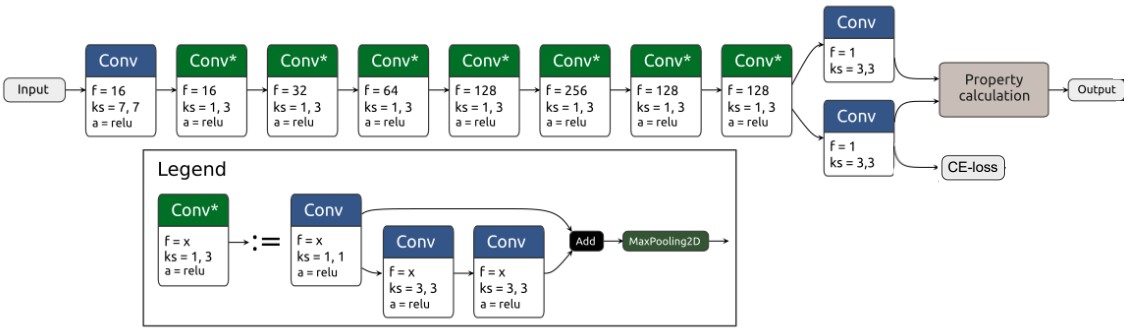

Figure 3.4: The custom FCNN architecture is comprised of Conv blocks, which are singular convolutional layers, and a sequence of Conv* blocks, which in turn consist of a sequence of convolutional layers followed by a max pooling layer and a skip connection, as exemplified in the legend. This sequence culminates into a custom property layer that calculates iOC and D by the methods described in the text below. Here, f is the number of filters in each block, ks is the kernel size, and a is the activation function. The network is visualized with Netron[126]. Figure reproduced with permission from [24].

Finally, to calculate intensity iOC and diffusivity D for each single biomolecule trajectory, we use a custom neural network architecture, which consists of a FCNN connected with residual blocks. The architecture itself is a sequence of convolutional layers followed by a max pooling layer, including short skip connections to preserve information and mitigate the problems of vanishing gradient as outlined in section 2.2.2. The full *body* of the architecture consists of seven of these connected blocks, culminating in a computational *head* module, as exemplified in figure 3.4. The *body* blocks are the neural networks which bear the brunt of the computation, whereas the *head* controls the nature of the data output by the network. In concrete terms, the *head* of the neural network returns three outputs: (1) The raw output O_{matrix} from the last convolutional layer of the neural network. (2) A mask, which is a down-sampled representation of the original kymograph normalized by its pixel sum and passed through a sigmoid activation function. This effectively gives us a weighted matrix, where each matrix element is the probability that the corresponding down-sampled region in the kymograph contains a biomolecule trajectory. (3) The calculated property for each kymograph, obtained by multiplying O_{matrix} by the mask to produce a value of the property for each pixel in the down-sampled kymograph weighted by the probability of said pixel containing a trajectory, and finally summing over the entire down-sampled image to obtain a single mean value of the property for the entire kymograph. Note here that when calculating D or iOC, the network architecture is identical - the only difference is which property is included in the labelled data and therefore included in the loss which the network's parameters are trained to minimize.

The FCNN networks were trained using the ADAM optimizer [127] with a learning rate of 0.0001 on approximately $300,000$ simulated kymographs in the same range as the U-net and train-validated equivalently to the U-net. The input during training of

the FCNNs are simulated images (kymographs) of size $128x2048$ with a single particle trajectory, and the output is a single value of either iOC or D of said trajectory, as well as the mask which as mentioned is a learnt downsampled representation of the original kymograph.

3.1.5 Training Methods

Now that we have a better idea of the underlying algorithms used in this thesis, it is natural to delve deeper into the specific training methods which are vital to reaching the high accuracy and precision and low LoD of the models used in paper I & II. In particular, I describe transfer learning; the process of transferring learnt knowledge from one network to another, ensemble modelling; the process of including the predictions of several neural networks in a given task, and curriculum learning; the process of successively increasing the difficulty of the task during training to be learnt to facilitate learning in very challenging conditions.

Transfer Learning

Transfer learning is a technique pioneered in the fields of AI and DL that involves transferring knowledge gained while solving one task to another (similar but not identical) task. The underlying idea is that some of the knowledge and skills learned while solving the first problem can be applied directly to the second problem, resulting in improved performance and shorter training time. It has been shown that a neural network trained on a particular type of dataset is much better at quickly recognizing new instances of similar data, when compared to a completely untrained network [133]. For instance, though a neural network may need thousands of examples of cars to start to accurately recognize cars, such a network can learn much quicker to recognize other vehicles when introduced into its training set compared to an untrained network [134]. This is a consequence of, as mentioned previously, neural networks learning increasingly complex abstractions in deeper layers of the network. Thus, a majority of the learning a network has to do to solve any given task is simply to find a good representation of the data, and transfer learning can thus be achieved by ”freezing” the training of the first layers of a neural network and only training the last few layers on the new data.

There are several benefits of this approach, most evidently that it can help reduce training time and data requirements for new tasks. Another benefit is that it can improve generalization performance, meaning that models trained using transfer learning are more likely to achieve good results on unseen data than those trained from scratch specifically for the new task at hand. Additionally, because transfer learning typically relies on pre-trained models which have been already tuned for good performance on a wide range of tasks, it can help avoid overfitting problems which can plague neural networks when they are retrained from scratch for a specific purpose.

Despite its many advantages, there are also some potential drawbacks to consider with transfer learning strategies. One issue may be poor adaptation if there are

significant differences between how tasks were originally learned and how they need to be applied later on, and special care should be taken when applying pre-trained models so as to not inadvertently introduce bias into results due to inaccuracies or irregularities in how the old network was transfer-learned into the new.

Ensemble Modeling

Ensemble modelling is a generic term meaning incorporating several models' predictions to improve accuracy, precision and overall generalizability of a final (ensemble) model. This technique has been used in traditional statistical methods for many years [135], but is equally appropriate in (deep) machine learning owing to the typically fast inference time of individual models. In general, models in an ensemble can be trained on different datasets or different permutations of the same datasets, be defined as identical architectures or variations of the same underlying architectures, or even be entirely based on completely separate architectures or algorithms. In principle, an ensemble model can even be any combination of the above alternatives.

There are several advantages of this approach [73], dependent fully on its specific implementation, but in general it's usually used to improve accuracy and generalizability by mitigating the consequences of exploding state spaces [136] and underspecification [51] and bias. In short, models trained on different permutations (or sometimes even the same) of the same dataset tend to learn different biases in different directions, such that their individual biases all cancel out with a large enough ensemble of models. Another related benefit is that of stability, wherein the variability of an ensemble model across different datasets or with changed parameters is decreased compared to using a single model. Additionally, ensemble modelling has been shown to be able to identify patters than may not be visible to any individual module [137]. In essence, combing information from several different models may reveal otherwise hidden patterns, making them especially useful for complex problems where traditional methods struggle to find accurate solutions.

Curriculum Learning

Another technique of significant impact to this project is curriculum learning [138], wherein we take full advantage of the benefits of transfer learning. Human beings often do not learn anything when approached with an insurmountable challenge, learn only the minimally required information when approached with a trivial problem, and learn optimally only when approached with a problem perfectly tuned to be slightly challenging given their particular skillset [139]. Neural networks, interestingly, have been shown to learn in analogous ways [65]. To take advantage of this, we may implement active learning [140] or curriculum learning schemes, wherein we vary which portions of the datasets are fed into the neural network during training to tune the challenges of learning, for instance by automatically feeding in more examples of data-points in regimes where the network retains low accuracy. As described in section 3.1.5, neural networks need to train on a large amount of data only to learn optimal representations of the dataset in its initial layers, but once

these representations are learnt, transfer-learning it on new data is usually only a question of tuning the specifics of the last few layers' learnt abstractions. To take advantage of this, we might employ a curriculum learning scheme wherein we train the networks with the easiest examples or the ones we have the most data-points of first, and successively increase the difficulty of the dataset as the network's loss decreases.

To train the intensity- and diffusivity-calculating FCNN models, a curriculum learning scheme with intermittent checkpoints to be used for later ensemble modelling prediction was employed. Specifically, the intensity-calculating model was initially trained only on a narrow range of high iOC trajectories, representing the highest SNR and in principle easiest case for the model to begin learning correlations, and then slowly curriculum-learned down to the lowest range of relevant iOC values. In each narrow range of iOC values, checkpoint models which are more accurate in that particular narrow range of values are saved separately. Upon model inference, an initial prediction is made with a model trained on the entire range of iOC values with the scheme described above, and then a second model trained on the narrower range of values makes a second prediction on the same trajectory to achieve higher accuracy. The process is equivalent for diffusivity, with the difference being that the range of D values being trained on increases rather than decreases during curriculum learning.

We note here that this ensemble modelling technique adds yet another barrier against false signal detection, since if the epistemic uncertainty inherent in the data is high, which may be the case for very short trajectories or regions of unusually high noise, the two predictions will differ significantly and hence the prediction should be discarded. Of course, as is common in ensemble modelling [141], one can implement an even larger and finer-grained ensemble for even higher prediction accuracy and accurate uncertainty estimation.

3.2 Simulations of Biomolecule Trajectories

As mentioned previously, all models are trained using supervised learning, i.e. by training each network to approximate a function correlating input x to label (output) y. To accomplish this, not only do we need to show the networks examples of kymographs containing particle trajectories, but also provide it with reliable ground truth regarding the position and properties of each individual trajectory. This can be done through manual labelling, but we will quickly run into issues regarding the limited datasets which can be procured by manual labour and problems related to human-induced biases. It is also true that such a trained network can never transcend the abilities of the human which provides the ground truth labels. Thus, we turn instead to the possibilities regarding generation of synthetic data, understanding the prospects of which necessitates a deep review of the underlying physical nature within the NSM system.

Firstly, let us remind ourselves that NSM works by imaging nanofluidic channels

within an optically transparent matrix by dark-field light-scattering microscopy as shown in figure 2.1. The channels exceed the field of the view lengthwise, and extend between tens to hundreds of nanometres in cross-sectional dimensions depending on the species to be investigated. As a result of this arrangement, the imaged biomolecules are localized within the microscope focal plane during the entire measurement and the optical contrast of the imaged biomolecules is enhanced by several orders of magnitude.

For small biomolecules, such as the proteins and DNA strands imaged in Paper I, their scattering in our system is effectively that of a diffraction-limited spot. To simulate their generated optical response in our system, we simulate trajectories with four main varying properties; integrated optical contrast iOC, diffusivity D, gaussian width s and velocity v. At time zero, a position x_0 is randomly chosen along the nanochannel, and the position x_i of the molecule at frame i is generated as Brownian motion with $x_i = x_{i-1} + v\Delta t + \mathcal{N}(0,1)\sqrt{2D\Delta t}$, where a random value $\mathcal{N}(0,1)$ is drawn from a normal distribution with mean 0 and standard deviation 1. The optical response of the biomolecule was then simulated along this path as a gaussian of width s, positions $x_0 ... x_k$ for length of trajectory k and magnitude iOC.

The generated response was then combined with simulated background noise (figure 3.5B) as $I = I^0 \cdot I^r$, where I^0 is the response of the empty channel and I^r the response of the biomolecule, and kymographs were created according to the procedure described in section 3.1.1. This simulated noise is generated as following. We begin by assuming that the noise in the system is dominated by two sources; shot noise from the channel and imperfections & dirt in the channel, and the vibration of the channels. The shot noise is generated as an inevitable consequence of quantum-level fluctuations in the number of photons interacting with the sensors of the camera over short periods of time. Shot noise, as a sequence of independently occurring events of photon interaction with a constant mean rate, can be modelled through a Poisson distribution, which in turn can be modelled by a Gaussian distribution in the limit of large N. The second source of noise is the vibrations of the channel, modelled as a harmonic vibration in the centroid of the Gaussian distribution. Thus, for each frame t, the background is first initialized as

$$b_0 = e^{(-(x-(x_0+x_\lambda))^2/\lambda^2)} \cdot (1 + \mathcal{N}(0,\,1) * b) \cdot (1 + A\mathcal{N}(0,\,1)) \tag{3.2}$$

where x denotes the position along the channel, x_0, λ, A are numerical parameters randomly drawn from a normal and uniform distribution, respectively, $*$ denotes the convolution operation, and where

$$x_\lambda = (2\mathcal{N}(0,\,1) + sin((\pi - 0.05)t)) \cdot d_x,$$
$$b = \frac{e^{(-(x)^2/C^2)} + x_\lambda}{\sum(e^{(-(x)^2/C^2)} + x_\lambda)},$$

where d_x, C are numerical parameters randomly drawn from uniform distributions. Here, in equation 3.2, the first factor is the aforementioned (partially sinusoidally

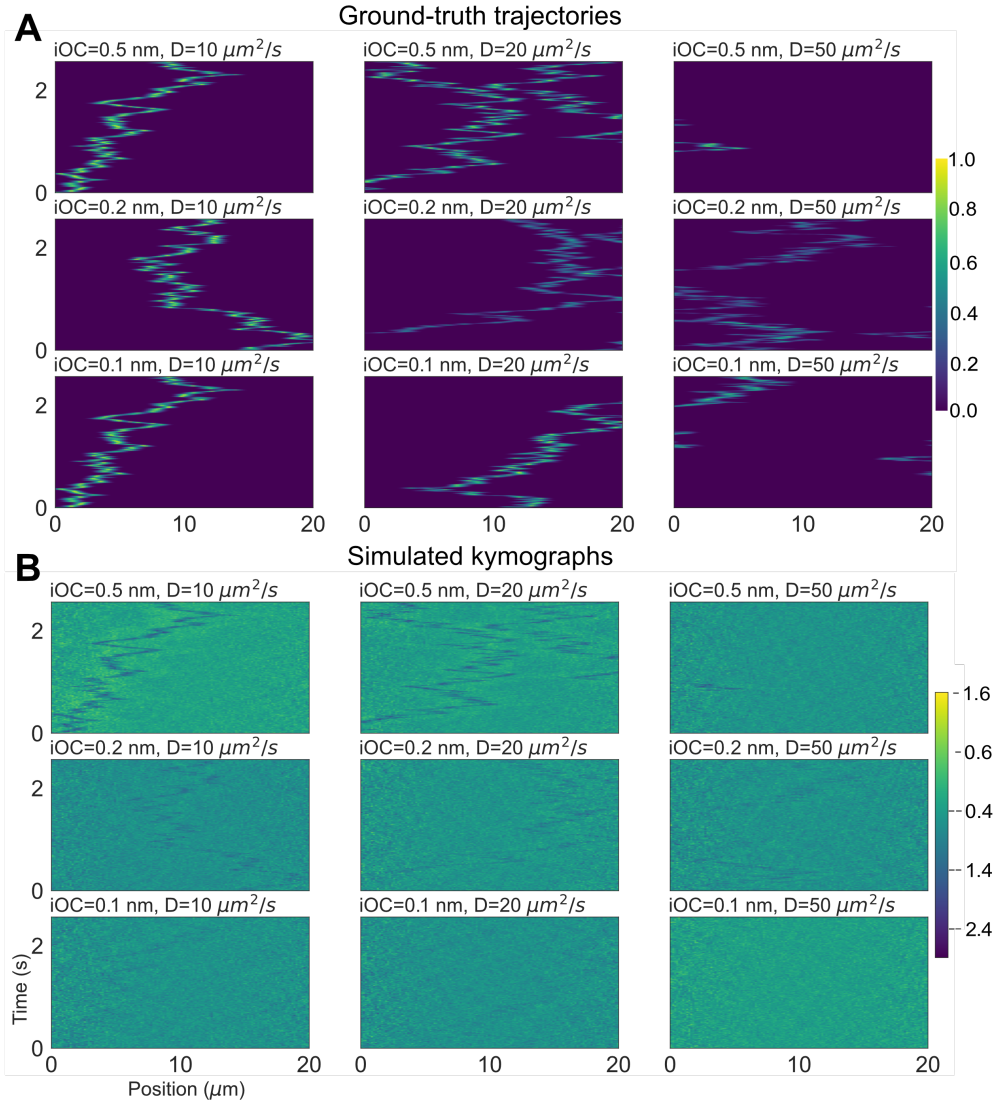

Figure 3.5: **A:** Examples of the simulated response of biomolecules defined by different combinations of optical contrast iOC and diffusivity D. **B:** Examples of simulated kymographs, acquired by combining the responses in **A** with simulated noise.

variant and partially randomly sampled) Gaussian. The second factor corresponds to noise resultant of a perfect channel (1) plus noise resultant of dirt and imperfections in the channel ($\mathcal{N}(0,\,1) * b$), which can be modelled by Gaussian noise with a large correlation length. We achieve this by a convolution with a generic (randomly sampled) point spread function (PSF) of the microscope and the (randomly sampled) position and size of said dirt. The third factor corresponds to the aforementioned shot noise, and is effectively a randomly sampled scalar.

Finally, the total background noise b_f is calculated as

$$b_f = b_0 \cdot (1 + b_A\mathcal{N}(0,\,1)) + 0.4 b_A\mathcal{N}(0,\,1) \tag{3.3}$$

where a new normal variable is drawn for each position x along the channel, and b_A is a numerical parameter drawn from a uniform distribution correlated with the total quantitative noise level in the system.

The value of the numerical parameters are chosen such that they roughly correspond to the quantitative levels of noise seen in experimentally measured noise, within a numerical span large enough such that any quantitative variations of the noise is still appropriately accounted for. In principle, so long as the true values of the experimentally measured noise are encapsulated within the range of the simulated noise, we can expect the neural network to properly interpolate and correct for any such variation of noise. However, note that this approach is sensitive to sources of noise which are not included in the simulations. Specifically, if a qualitatively different source of noise excluded in the simulations is present in a measurement at inference time (i.e. spurious vibrations as a result of human movement, or device malfunction), it will be ambiguous to the neural network whether this portion of the data should be considered signal or noise. This problem can, to an extent, be mitigated by transfer-learning the trained network on a few examples of simulated trajectories within experimentally measured noise before each new set of measurements, effectively constituting a re-calibration step of the network. Alternatively, it is always possible to include new sources of noise into the simulations. However, note that this increases the state-space of possibilities which the network needs to learn and thereby increases the overall training time and increases the requirements for representational ability within the choice of network architecture.

Regardless, this method demonstrates a powerful benefit of using deep learning systems: we do not have to perfectly describe or understand the noise in order to correct for it. In general, so long as the noise is consistent (i.e., no significantly different qualitatively new sources of noise are expected to exist at inference time), and we know the generic representation of the dominating sources of noise, we can expect deep learning systems to properly interpolate learnt correlations to experimentally relevant data.

Chapter 4

Summary of Paper I and II

In this chapter, I present the results of the DL pipeline introduced in figure 3.1. First, I present the results of the pipeline on a synthetically generated dataset consisting of simulated biomolecule trajectories in experimentally measured noise, to serve as benchmark for the data analysis as a whole. Secondly, I present results on a biomolecule library, consisting of a selection of a proteins and biomolecules relevant for biological study. Thirdly, I present results of measuring a complex biofluid consisting of cell-secreted EVs, whereby I demonstrate the method's versatility and overall biological relevance.

4.1 Simulated Biomolecules in Measured Noise

To estimate the accuracy and precision of the algorithm, we test it on a combination of simulated biomolecules of $iOC = [0.1, 0.2, 0.5, 1, 2]$ nm and $D = [10, 20, 50]$ μm^2/s on top of experimentally measured noise. The trajectories of biomolecules were found using the segmentation & object detection algorithm described in sections 3.1.2 and 3.1.3, from which iOC and D were calculated using the FCNN described in 3.1.4. Such determined values of iOC and D of all detected trajectories are shown as a scatter plot in figure 4.1. The mean values of iOC and D in figure 4.1 correspond to the mean calculated value of iOC and D for each separate simulated trajectory. Thus, for a perfect algorithm, we would expect the circular points (mean calculated value) to perfectly overlap the crosses (ground truth), with error bars of minimal length (see [24] for theoretical limits in precision). The values are compared with that found by a standard heuristic-based algorithm (SA) in figure 4.1B. Clearly, the values identified by the pipeline correspond to the ground truth very well even within experimentally measured conditions, and particularly outperform the standard analysis at the lowest iOC, thereby corroborating the used data analysis pipeline as a whole.

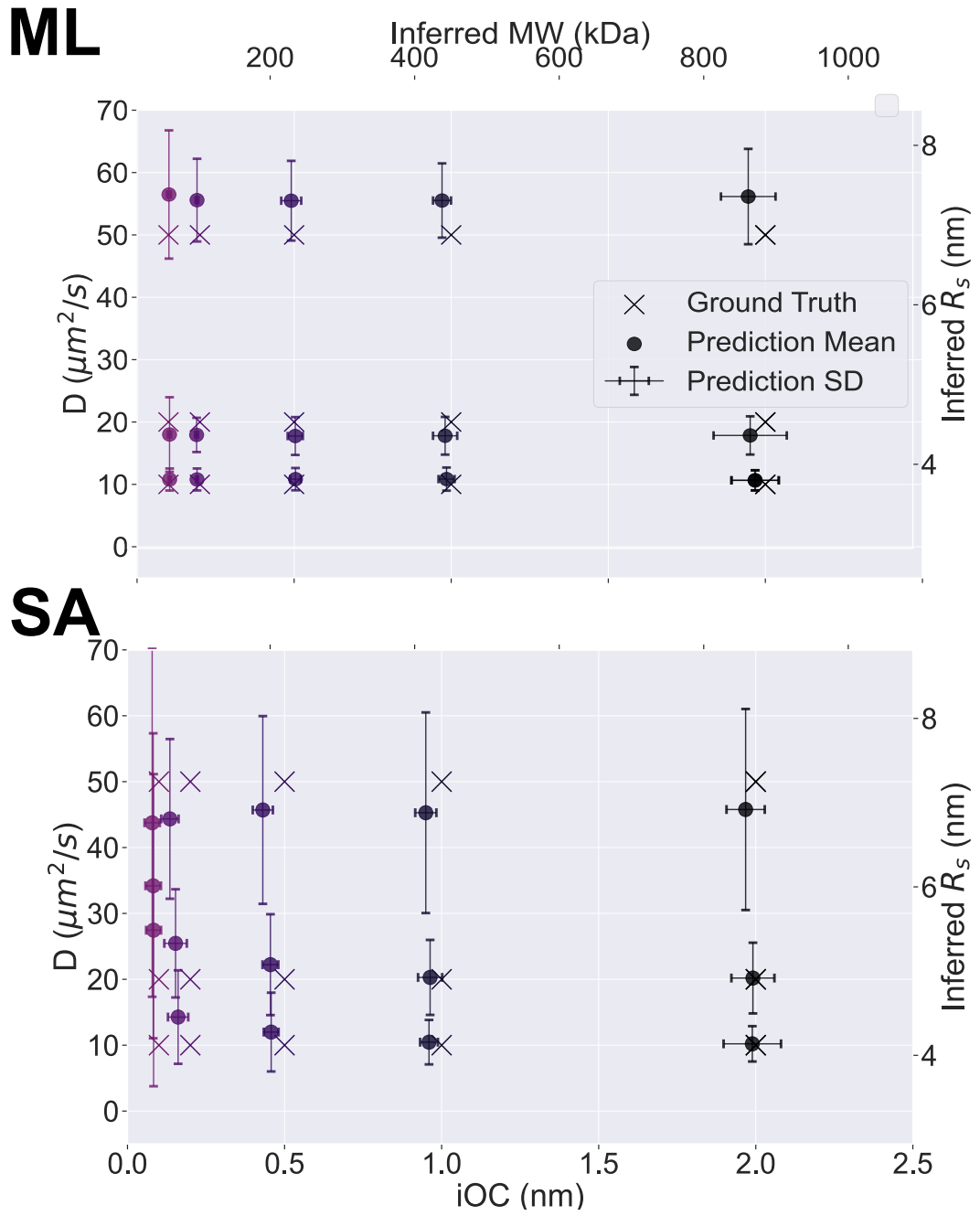

Figure 4.1: Analysis of the artificially generated data set comprised of experimentally recorded background signal and simulated response of biomolecules defined by a combination of values for $iOC = [0.1, 0.2, 0.5, 1, 2]$ nm, and $D = [10, 20, 50]\mu m^2/s$. **DL:** Scatter plot of mean value of iOC and D. Error bars correspond to the resolution in iOC and D defined as the standard deviation of the calculated values for iOC and D for each separate D, iOC permutation. **SA**: Same as for DL, but for the standard analysis algorithm as described in [24].

4.2 Summary of Paper I

4.2.1 A Biomolecule Library

In figure 4.2, the results of applying the ML pipeline to a collection of proteins and DNAs with well-known molecular weights and hydrodynamic radii are presented, and the overall correspondence fits very well. In figures 4.2A and 4.2B, we have used two different nano-channels for the measurements, whose cross sections affect the conversion between iOC/D and MW/R_s as described in section 2.1.3. Here, each data point corresponds to the iOC and D determined for a single trajectory from a single biomolecule within the nanochannel, with its transparency value correlated to the length of each respective biomolecule trajectory. Further, in figures 4.2C and 4.2D, we show one-dimensional histograms of the MW (converted from iOC) of all trajectories, where the number of counts within each peak corresponds to the sum of the lengths of all detected trajectories corresponding to the MW at the respective peak. Thus, the number of counts in each separate (gaussian) histogram distribution correspond to the relative concentration of that species within the studied sample. The guiding line in black is the expected curve for globular protein, whose hydrodynamic radius approximately scales as $R_s \cong b \cdot MW^{\frac{1}{3}}$, where $b = 0.88\text{nm/Da}$ [142].

For elongated molecules, such as DNA, their diffusivity is expected to be considerably lower compared to globular molecules at the same corresponding MW [144]. This is indeed consistent with what is found by the DL analysis. Note also the presence of molecular populations with higher MW and R_s than expected from its main monomer peak, specifically for thyroglobulin and BSA, wherein the indications of a second peak at double the MW and c.a. a factor 1.3 greater R_s suggest that these are protein dimers (as elaborated upon in section 2.1.3) [145].

The peaks corresponding to molecular monomers are shown by ellipses in figure 4.2A, B, where centers of said ellipses correspond to the mean value of iOC ($\overline{iOC}$) and D ($\overline{D}$) and their diameters to the resolution in iOC and D defined as the full-width-half-maximum (FWHM) of the peaks in figure 4.2C,D. Further, in figures 4.2E and 4.2F, we compare the performance of ML analysis to that of the SA [146] for MW and R_s approximation, respectively. Here, we find that the ML leads to better accuracy and precision for all included protein monomers and DNA strands. The black lines show the theoretical prediction of iOC/MW and D/R_s conversion, in channel I as a dotted line and channel II as a striped line, and it is clear that the theoretical predictions are reproduced very well in the experimental results.

It should be noted that there are significant molecular populations revealed by the AI results which correspond to higher MW and R_s than expected from protein monomers. In particular, results for thyroglobulin, ferritin and BSA show a second population with approximately $2\times$ higher MW and $\approx 1.3\times$ higher R_s than expected, which likely corresponds to a dimer population [147] [145] as mentioned. Further, for ferritin, we find multiple subpopulations characterized by molecules with different numbers of coordinated iron [148].

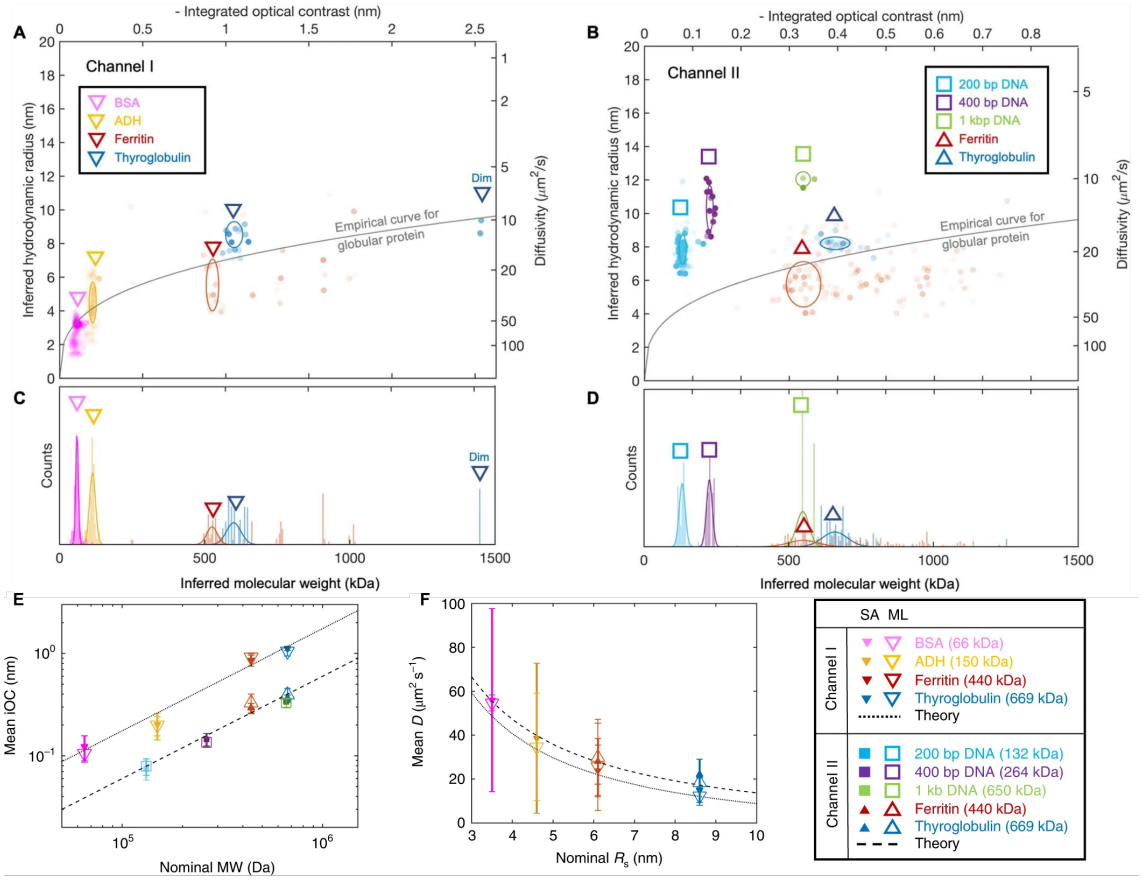

Figure 4.2: DL analysis results. (A, B) Scatter plots of *iOC* translated into *MW* using equation 2.3 for in the main text, and D translated into R_s using equation 2.4 for individual biomolecules of different types, measured in (A) Channel I and (B) Channel II using ML. Each dot is extracted from a single biomolecule trajectory. Color intensity scales linearly with frame number of the trajectory (N). The highest intensity corresponds to $N = 2130$ frames. The grey line corresponds to an empirical relation between MW and D for globular protein [143] (C, D) *iOC* histograms of the biomolecules in (A, B) translated into *MW* using equation 2.3. Number of counts is normalized by the total number of frames in which biomolecules of each type were identified. (E,F) Dependency of iOC of protein monomers on nominal MW compared with the theoretical model (equation 2.3), and protein monomer D dependency on nominal R_s compared with the confinement-corrected Stokes–Einstein equation (equation 2.4). E and F also show the agreement between the independent results of SA and ML; data are presented as mean values and error bars correspond to the resolution in *iOC* and D, respectively. Here, the solutions containing ADH, ferritin and thyroglobulin had a concentration of 28 nM of the protein solution in PBS, and the DNA fragments and BSA had a concentration of 28 nM in 0.05× TBE. Figure reproduced with permission from [24].

Moreover, note that the values determined by ML are in agreement with the SA, further corroborating that the data analysis procedure is valid. As the ML has higher speed, accuracy and precision (especially for smaller molecules), we use it exclusively for all further analysis.

4.2.2 Extracellular Vesicles

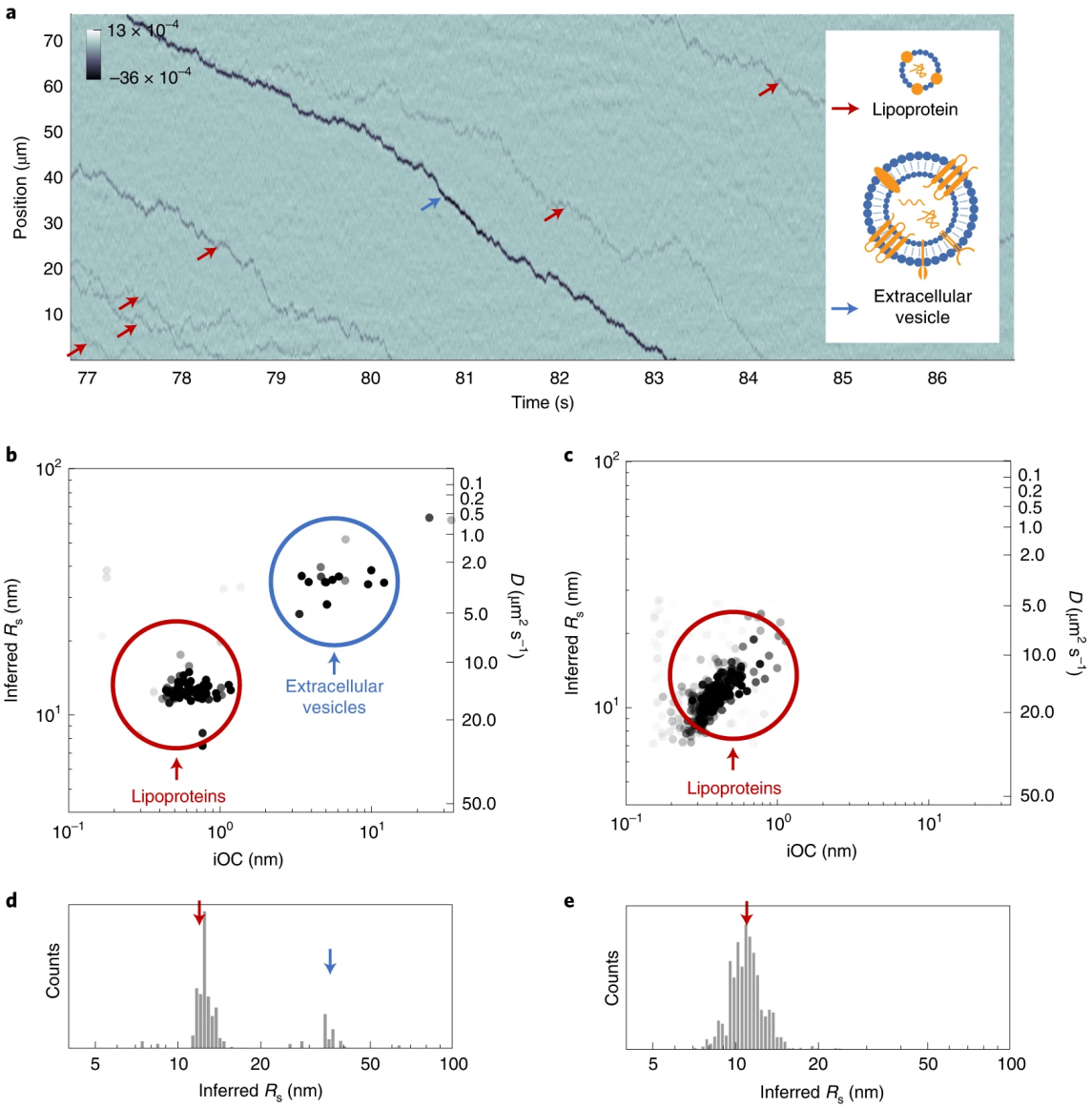

Figure 4.3: Analysis of BNPs in conditioned cell culture medium containing serum. (A) Kymograph of multiple lipoprotein particles (marked by red arrows) and one larger EV (marked by the blue arrow) moving through the nanochannel. (inset) Schematics of an EV and a lipoprotein (depicted to scale). (B,C) Scatter plots of *iOC* and D translated into R_s. (D,E) Histograms of R_s, analyzed using ML. Data in (B,D) corresponds to the conditioned SH-SY5Y human cell medium and (C,E) the control where the medium had not been in contact with the cells. All data were acquired in Channel V that had been passivated by an SLB prior to the measurement. Figure reproduced with permission from [24].

In order to demonstrate the technique's biological relevance, we chose to investigate a complex biological sample containing a mixture of serum proteins and secreted EVs (see section 2.1.3). In practice, this requires only a few minor adjustments compared to the procedure used to measure the biomolecules shown in figure 4.2. First of all, to accomodate for the significantly increased size of EVs, we use the larger channel III described in section 2.1.2. Secondly, we coated the walls of the layers with a supported lipid bilayer (SLB) [149] to prevent nonspecific binding of EVs to the nanochannel. Thirdly, we applied a slow flow by introducing a 0.25 Pa pressure drop along the channel to increase overall measurement throughput. Finally, the data analysis was modified only in two ways; firstly, new iOC- and D-calculating FCNN models (with identical architecture, described in section 3.1.4) are trained in higher ranges of iOC and with velocity-induced flow. Secondly, since the stability of the microscope and power of the laser had degraded prior to these measurements, vibrations and shot noise were overall higher than for the measurements of the previous subsection. To solve this, the denoising U-net described in section 3.1.2 was trained on quantitatively higher levels of noise. These last two points show the robustness of the data analysis procedure; accounting for these new effects with the current AI pipeline is simply a matter of simulating said effects and re-training new models. Indeed, accounting for higher noise is in effect simply transfer-learning existing models on said higher noise, thereby still retaining all trained knowledge inside the models. Due to the principles of transfer learning as described in section 3.1.5, re-training an existing model on new simulations requires significantly less data and thus computational power compared to a model trained fully from scratch.

Using this slightly modified procedure, with everything else in the analysis pipeline consistent, we collected a significant number of trajectories of these biological nanoparticles (BNPs) and derive their iOC and D as described previously (see section 3.1.4). This analysis revealed two distinct populations of nanoparticles, one in the regime of smaller iOC (R_s) and one in the regime of higher iOC (R_s). Through investigating a control experiment consisting only of the cell culture medium within which the EVs were produced, we found only the first population corresponding to smaller iOC (R_s) reproduced. Thus, it is likely that the second population does indeed correspond to EVs secreted by the cells. The identified values of $R_s \approx 10 - 13$ nm correspond to low-density lipoproteins [150] and $R_s \approx 20 - 70$ nm correspond to EVs [151]. Note here, though, that the translation of iOC into MW is considerably more complex for EVs than for single molecules due to the varying cargo within the BNPs adding spatial dimensions to the BNPs' refractive indices and changing the effective refractive index of the BNP as a whole. The particles' larger sizes also introduce uncertainty resultant of interactions with the walls of the nano-channel as discussed in Supplementary Information 4 of Paper I.

4.3 Limits Beyond State-of-the-art

This last section of the thesis regards the aspects of biomolecule detection and characterization of which the introduced AI pipeline exceeds the established state-

of-the-art. In particular, we investigate how the accuracy and precision varies as function of trajectory length for different molecular species below state-of-the-art for limit of detection, and draw conclusions on which hitherto elusive species may become available for study using this AI approach.

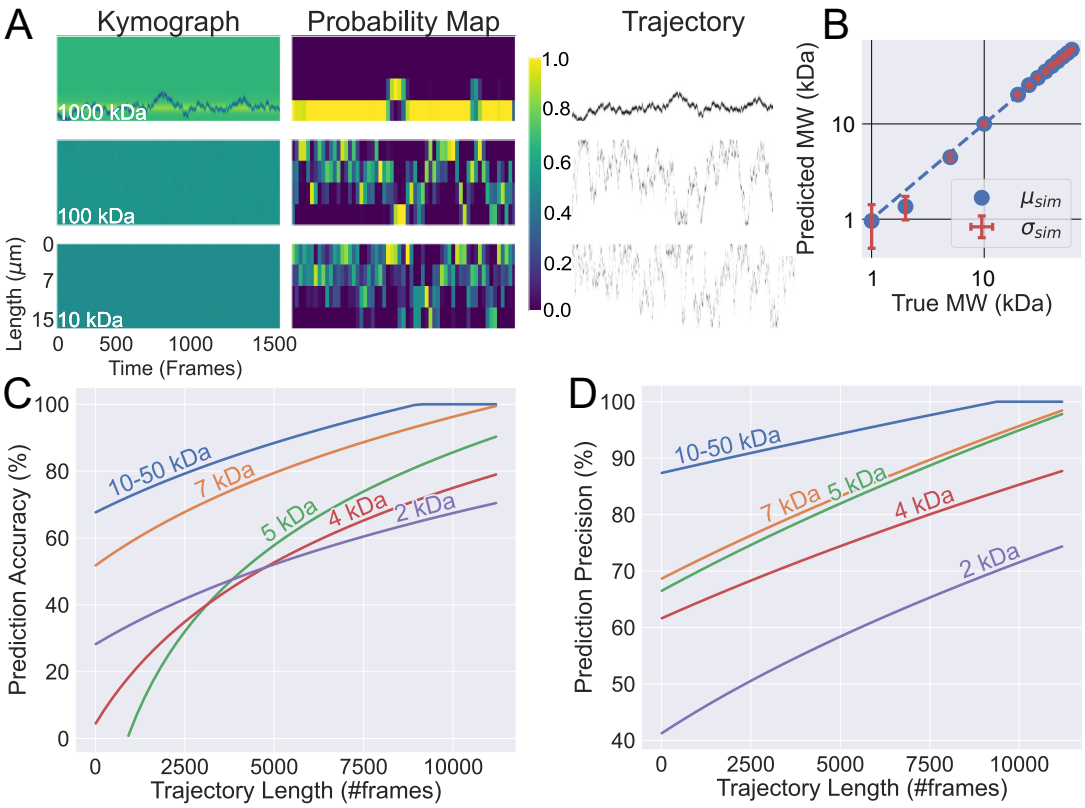

Figure 4.4: Limits of Machine Learning Analysis. (A) Kymographs in column 1 showing simulated optical response of a 1000, 100 and 10 kDa molecule, respectively, column 2 the FCNN's predicted probability map of their trajectories, and column 3 their ground truth trajectories. (B) Log-scale scatter plot showing mean predicted MW μ_{sim} versus true MW in blue dots for select masses in the range $1-60$ kDa, with corresponding standard deviation σ_{sim} in red bars and ideal prediction shown as a blue line. (C,D) Accuracy (Precision) of prediction as function of trajectory length for select masses in the range $2-50$ kDa. All masses assume channel cross-sectional area equivalent to channel I.

The three images within the first column of figure 4.4A show the kymographs of the simulated optical response of a simulated 1000, 100, and 10 kDa particle over time, generated as described in section 3.2 and processed as described in section 3.1.1. Note that whilst the trajectory of the heaviest biomolecule is clearly visible even to the eye, the optical response of the lightest molecule is almost indistinguishable from the background. The three images in the third column of figure 4.4A show the ground truth trajectories of the same particles. Here, each trajectory is a sequence in time of the x-position of the particle along the nano-channel, where its position is modelled as a gaussian centered around x with maximal value 1 and gaussian width correlated with the size of the particle (further details in section 3.2). The three images in the middle column of figure 4.4A show the neural network's predicted probability

map of the particle, i.e. the probability of each region of the kymograph containing a particle trajectory. As described in section 3.1.4, not pictured here is a separate parallel map output by the network, corresponding to the calculated iOC within each rectangular region of the probability map assuming that it contains a single particle trajectory. The true value of iOC of the single trajectory contained within the kymograph is acquired by multiplying the two aforementioned maps together and summing them, effectively employing a sum of iOC along the trajectory weighted by its probability of existence. This scheme effectively enables the quantification of trajectory properties without inherently necessitating pixel-wise particle tracking, in turn enabling robust analysis of trajectories below even the standard deviation of noise (i.e. where the probability of each pixel containing a trajectory is always less than one).

In figure 4.4B, we plot the resulting predicted MW as a function of true MW in logarithmic scale, for a set of MW between 1 and 60 (roughly corresponding to the limit of current state-of-the-art), with corresponding standard deviation as error bars. Each point corresponds to the prediction of 25 samples, where each sample is a set MW and trajectory length (11200 frames) within a randomly generated noise profile which quantitatively and qualitatively matches the power spectrum of the experimentally measured noise at the time of measuring the results in figure 4.2 (see section 3.2 for details).

In figure 4.4C, we show the accuracy and precision of predicting MW as a function of trajectory length, for a selected set of MW between 2 and 50 kDa. Both accuracy and precision are estimated from 25 samples per point, same as for figure 4.4B. Accuracy is defined here as (one minus) the average error normalized by the correct MW, i.e. $1 - |\mu_{pred} - \mu_{real}|/\mu_{real}$ and precision as the maximum of 0, and the mean predicted MW minus the standard deviation of the prediction normalized by the mean predicted MW, i.e. $max(0, (\mu_{pred} - \sigma_{pred})/\mu_{pred})$. The network is evaluated at 6 linearly separated discrete trajectory lengths between $N = 10$ and $N = 11200$ frames, and an interpolating logarithmic and inverse quadratic function is then fitted for accuracy and precision, respectively. Note that accuracy is generally very high for particles above 10 kDa in MW, and it only begins to drop off below this regime. Here, the accuracy drops significantly for short trajectories; there simply isn't a high likelihood that the trajectory's signal overcomes the noise in any significant number of frames. However, the accuracy is still very high ($\sim 80\%$) for long trajectories of small molecules, thus principally enabling their accurate quantification if only one can measure them for longer periods of time. Note also the peculiar nature of a seemingly improving accuracy for the smallest molecules (< 5 kDa) and shortest trajectories. This is simply a consequence of the signal being almost fully obscured by the noise in these regimes, and the algorithm therefore tending to output values close to 0 in these cases, leading to a technical (but uninteresting) increase in accuracy for molecules with MW closer to 0. The precision of predictions shows similarly precise results for $MW \geq 10$ kDa, and quickly decreasing precision for smaller molecules.

By these simulated results, it is clear that the state-of-the-art of molecule detection and characterization may be surpassed significantly, including a LoD below 10 kDa

assuming average trajectory lengths of ≈ 1500 frames for DNA molecules and 900 frames for proteins of this size range (see [24] for details) and below even 5 kDa assuming long trajectory lengths. Such long trajectory lengths may be achievable through statistical bootstrapping [152], simply by chance given long measurements, or potentially by electrostatic entrapment of molecules [153].

Chapter 5

Conclusions & Outlook

5.1 Conclusions

NSM has recently been introduced as a technique that enables the label-free real-time characterization of freely diffusing single biomolecules down to the tens-of-kDa regime inside nano-sized channels. In principle, there is no fundamental limit to the lower limit of detection of this system in the nano-scale; it is only technically limited by e.g. the recorded scattering intensity (i.e. maximum available laser power and frame-rate of the camera), cross-sectional dimensions of the nanofabricated channels (i.e. strength of interference between light scattered from the channel and the biomolecule confined therein), and the sophistication of the statistical analysis used to resolve properties of biomolecules from the scattering profile of the resulting kymographs. The focus of this thesis has been to optimize the statistical analysis; leveraging the astounding progress of DL-based CV algorithms to push the possibilities of statistical analysis beyond conventional limits and enable biomolecule screening below the 10 kDa regime.

The two main conclusions of the thesis are; firstly, that it is possible to accurately characterize biomolecules from statistical variations in a sequence of consistent measurements without explicitly needing to recreate the biomolecule's frame-by-frame trajectory. Secondly, we have found that in studying a system where simulating its properties is considerably simpler than calculating them directly, deep learning can be an ideal general solution. Thus, characterizing a biomolecule below the 10 kDa-regime in the NSM system with conventional heuristic-based methods is fundamentally more difficult than characterizing a biomolecule in the 1000 kDa-regime. However, once the performance of a deep learning system has been validated on simulated molecules in the 1000 kDa-regime, simulating and training on molecules below the 10 kDa-regime is straightforward and can be enabled by well-established training schemes such as ensemble modelling and curriculum learning.

In terms of the long-term future outlook of NSM and the role of deep learning therein, the possibilities for innovative and impactful research in a broad range of fields is both promising and abundant, as outlined in section 5.3. This transforms

the challenge of determining the short-term future outlook of this thesis from one of finding interesting further applications, to instead narrowing the focus on the most immediately promising prospects of study.

5.2 Short-term Future Outlook

The focus of said near-term future projects can be split in two; firstly, applying the deep learning methods developed in this thesis to analyze single molecule trajectories to a high degree of accuracy, precision and detection limit, and secondly to develop new deep learning methods to enable analysis beyond the single trajectory regime. All the examples brought up in the sections below are already being worked on or are in early planning stages, through collaboration with a broad range of research groups and institutions.

5.2.1 Single Molecule Analysis

All projects within this sub-section share one thing in common; they can in principle be solved by the deep learning tools introduced in this thesis as-is, including the high precision, accuracy and detection limits said tools entail. The only required modification is to train the models in the corresponding parameter range of noise and generated optical response that each inividual project requires.

Insulin

Following the theoretical limits of detection and quantification defined in Paper II and outlined in section 4 of this thesis, accurate characterization of single molecule trajectories below the 10 kDa regime is possible in channel I as defined in section 2.1.2. Since the iOC of molecules scales linearly with decreasing nano-channel cross section, smaller channels than channel I may also be constructed to improve the accuracy of said characterization. Thus, we should expect to be able to accurately characterize even the protein-hormone insulin with $MW \approx 6$ kDa, particularly important for diabetes research. This would consitute a world-first in the label-free characterization of freely diffusing molecules in this size range in general, and for insulin in particular.

λ-DNA

Over the past two decades, nanofluidic structures have become a potent platform for in-depth investigation of DNA on the kilo-bp length scale [154]. The combination of excluded volume and DNA stiffness causes DNA to be stretched to almost its full length when contained in nano-sized channels [154]. NSM is therefore an ideal complementary method to study these long DNA structures, and preliminary results on comparing fluorescently stained λ-DNA to unstained λ-DNA show that the staining process indeed affects the length and optical contrast of said molecules. Therefore, to study their properties without significantly altering them, label-free methods such as NSM are crucial. In this project, all deep learning models were

used as described in this thesis, with the only difference being that the models were instead trained on the generated simulated response of $\lambda - DNA$ in our nano-sized channels. Further measurements will corroborate these results, and lead to the natural continuation of the project in studying DNA-protein binding events as outlined in section 5.2.2.

Stress Granules

Eukaryotic cells respond to stress by forming stress granules, which are effectively transient membraneless organelles [155]. They have a core-shell substructure, are made up of messenger ribonucleoprotein complexes, and are crucial for controlling the rate of protein translation and signaling under stress [155]. This has significant implications for cancer and neurological illnesses [155]. The molecular structure of stress granules and their biochemical function in plant growth and fitness are the subjects of active research, yet no methods have so far been able to measure each granule individually in a satisfactory manner [155]. Stress granules, as high-density protein and RNA aggregates formed by organisms under stress, may be analyzed in an effectively equivalent fashion to EVs in Paper I through single molecule trajectory analysis in large channels.

Extracellular Vesicles

In Paper I, we introduce NSM as a powerful characterization technique for EVs, and we expect it to remain a key method for label-free EV analysis in the future. In the short term, we will use NSM not only to characterize EVs, but also their secreted contents by use of nanovesicle traps built directly into the nano-sized channels. Further, as noted in Paper I, precise conversion of iOC into MW is more complicated for BNPs such as EVs than for single molecules due to a large variety of molecular constituents with different optical properties whose representation and spatial distribution might be different for each BNP. Therefore, we did not further explore the content of the EVs via the measured iOC. Further theoretical work suggests that correcting for such effects is possible through modelling the electromagnetic properties of EVs, which would enable more informative BNP characterization in general.

5.2.2 Beyond Single Molecule Analysis

Going beyond the analysis of isolated trajectories requires the development of new deep learning models. Unpublished results on simulated data show that node-regression-based graph neural networks (GNNs) are ideal for high-density samples which require high-fidelity output temporally and spatially through the channel [156]. In principle, GNNs retain the inherent benefit of a CNN approach in being able to consider an entire trajectory simultaneously in its prediction of its properties, without the disadvantage of having to consider an entire kymograph (or portion of a kymograph). Thus, each frame of a trajectory can be given a separate prediction of its properties, whilst still considering the entire connected trajectory in the prediction. This approach is ideal in cases where we expect molecule trajectories to change

over time, as is the case for enzyme-molecule interactions, DNA-protein binding events and investigate conformational changes in trajectories over time, among other applications.

Protein-Protein Interactions

When performing their tasks *in vivo*, proteins hardly ever behave as separate species [157]. In fact, it is estimated that over 80% of proteins have been found to function in complexes rather than alone [157]. Studying how protein-protein interactions impact their function is crucial for determining how proteins work in their natural environment, i.e. in the inside of cells. Furhter, to understand the cell's biochemistry, it would be incredibly helpful to characterize the interactions between the proteins in said cell [157]. There are numerous approaches to determine the outcome of an interaction between two or more proteins that has a clear functional goal.

Cell-to-cell contacts, metabolic regulation, and control of developmental processes are only a few of the biological activities that are handled by said protein-protein interactions [158]. These interactions can be categorized in a number of ways based on their differing structural and functional traits [158]. For instance, they can be homo- or heterooligomeric depending on their interaction surface, obligate or non-obligate depending on their stability, and transient or permanent depending on their persistence [158]. Any combination of these three distinct pairs may be used to characterize a certain protein-protein interaction. For instance, while temporary contacts likely create signaling pathways, long-term interactions are more likely to result in a stable protein complex. NSM is a promising method to study such interactions, and the aforementioned GNN approach is an ideal analysis tool for such an application. Another project along the same focus is to study enzyme-molecular interactions, i.e. how an enzyme (catalyst) affects the properties of a biomolecule when bound to it, as is the case with CRISPR/Cas-9 binding events.

λ-DNA - CRISPR/Cas-9 Protein Binding Events

Proteins are lengthy chains of amino acids connected by peptide bonds and, as established in section 2.1.3, are versatile and serve a number of purposes as building blocks of the cell. Transporter proteins, receptor proteins, structural proteins, and enzymes are a few of the various protein subgroups [159]. The latter work by speeding up chemical reactions and highly selectively controlling a wide range of functions within living cells [159]. They are also crucial to the replication, repair, and alteration of DNA [159]. The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated protein-9 nuclease (Cas9) is one example of an enzyme involved in DNA modification [160]. Cas9, effectivelly a gene-editing tool, can catalyze extremely sequence-specific double stranded breaks by cleaving links in the DNA backbone [160]. The enzyme has many scientific uses, such as gene targeting, control, and involvement in the adaptive immune system of bacteria and archaea [160]. The Cas9 enzyme uses an RNA molecule, known as Guide RNA (gRNA), with a 20 bp motif corresponding to the site of interest to identify the precise DNA

sequence at which it will cause a double-stranded break with very high precision. Direct detection and characterization of these binding events, as principally enabled by NSM, is a highly relevant and sought-after feature in the fields of gene editing and genomic mapping [159]. To study how active each binding site is over time, GNNs are again an ideal solution.

Temperature Influence on Oligomeric State of Protease

Protein quality control is a crucial biological task primarily carried out by a wide variety of distinct proteases [161]. One of these, the *DegP* protease, is essential to the Escherichia coli protein quality control system because it eliminates unfolded proteins or stops them from aggregating [161]. Upon temperature activation, this protease is expected to go through extensive oligomeric state modifications, including massive structural transitions from its hexameric resting form via a trimeric intermediate state toward the higher oligomeric states [161]. However, despite being thoroughly studied and having a number of high-resolution crystal structures of DegP in its many states, there are still many unanswered concerns regarding how its structural transitions are actually accomplished [161]. Prior studies using NMR diffusion and Cryo-EM have been unable to elucidate these questions [161], but temperature-controlled NSM is in contrast quite promising. To study the structural transitions over time (with varying temperature), the aforementioned GNNs are ideal.

5.3 Long-term Future Outlook

In terms of DL advancement, recent progress in geometric DL for particle tracking and characterization in terms of GNNs represents promising complementary methods for use in high-fidelity high-density studies [162], where the CV pipeline introduced in this thesis will struggle. It has recently been independently shown [156] that reconstructing the trajectories of diffusive particles is not a necessary requisite to accurately calculate the properties of the underlying stochastic process which gave rise to the trajectory. Machine-learned models, like the ones trained in this thesis, have the potential to exploit this hidden information to push the boundaries of particle characterization methods in the regime of low SNR where heuristic-based methods often struggle significantly. Since the models are trained entirely on simulated data, their performance is also technically limited by available computational power and are thereby expected to improve in tandem with the breakneck rates at which newer generations of parallelizable (GPU) and tensor-processing-optimized (TPU) chips operate and are developed. Therefore, we expect that beyond enabling the analysis of biomolecules in the single-digit kDa regime, these DL tools will become ever-more crucial in the complex and versatile analysis that NSM and other modern methods of optical characterization enable.

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. Advances in terms of instrumentation and DL tools is

expected to push the performance even further, and particularly we expect that high-throughput and resolution in single-digit kDa regime will find numerous bioanalytical applications requiring analysis of highly heterogenous samples, such as exosome characterization and direct detection of small-molecule binding events [163]. Further, 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. Looking further forward, we expect as mentioned that NSM will find applications outside the field of bioanalysis, such as characterization of inorganic nanoparticles, particle counting, and single particle catalysis [18].

Acknowledgments

I would like to acknowledge Barbora Špačková, whose insightful descriptions of the underlying biochemistry and physics was absolutely vital for the successful completion of this project, and for providing several useful figures and graphs that were featured in this thesis.

I would like to acknowledge Daniel Midtvedt for his technical expertise within machine learning, without which the successful implementation of these tools would have been impossible.

I would like to acknowledge Gustaf Sjösten, who worked as my de facto partner throughout the early process, together with whom I co-developed the early implementations of the machine learning algorithms in close collaboration.

I would like to acknowledge my supervisor Christoph Langhammer and co-supervisor Giovanni Volpe, without the initiative of whom this project would never have existed.

I would like to acknowledge Erik Olsén, whose valuable feedback on this thesis was significantly helpful.

I would like to acknowledge the computer cluster Alvis, through which some of the computations in this thesis were made possible. These computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreement no. 2018-05973

I would like to acknowledge Knut and Alice Wallenbergs Foundation and Stiftelsen för Strategisk Forskning for funding my project.

Finally, I would like to acknowledge my unnamed yet invaluable friends and family, without the persistent support of whom this thesis would never have been possible.

Henrik Klein Moberg

Gothenburg, December 2022

Bibliography

[1] Y. Huang and Y. Chen, *Autonomous driving with deep learning: A survey of state-of-art technologies*, 2020. DOI: 10.48550/ARXIV.2006.06091. [Online]. Available: https://arxiv.org/abs/2006.06091 (cit. on p. 1).

[2] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, and Q. Sun, "Deep learning for image-based cancer detection and diagnosisa survey," *Pattern Recognition*, vol. 83, pp. 134–149, 2018, ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2018.05.014 (cit. on p. 1).

[3] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, no. 1, p. 83, Aug. 2019, ISSN: 2057-3960. DOI: 10.1038/s41524-019-0221-0 (cit. on p. 1).

[4] D. Baron, *Machine learning in astronomy: A practical overview*, 2019. DOI: 10.48550/ARXIV.1904.07248. [Online]. Available: https://arxiv.org/abs/1904.07248 (cit. on p. 1).

[5] J. A. Weyn, D. R. Durran, and R. Caruana, "Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 9, e2020MS002109, 2020, e2020MS002109 10.1029/2020MS002109. DOI: https://doi.org/10.1029/2020MS002109. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002109 (cit. on p. 1).

[6] O. Garcia-Garin, T. Monleón-Getino, P. López-Brosa, A. Borrell, A. Aguilar, R. Borja-Robalino, L. Cardona, and M. Vighi, "Automatic detection and quantification of floating marine macro-litter in aerial images: Introducing a novel deep learning approach connected to a web application in r," *Environmental Pollution*, vol. 273, p. 116 490, 2021, ISSN: 0269-7491. DOI: https://doi.org/10.1016/j.envpol.2021.116490 (cit. on p. 1).

[7] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989. DOI: 10.1162/neco.1989.1.4.541 (cit. on pp. 1, 10, 11).

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, ISSN: 1476-4687. DOI: 10.1038/nature14539 (cit. on pp. 1, 15).

[9] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution

in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, Jan. 2019, ISSN: 1548-7105. DOI: 10.1038/s41592-018-0239-0 (cit. on p. 1).

[10] B. Midtvedt, E. Olsén, F. Eklund, F. Höök, C. B. Adiels, G. Volpe, and D. Midtvedt, "Fast and accurate nanoparticle characterization using deep-learning-enhanced off-axis holography," *ACS Nano*, vol. 15, no. 2, pp. 2240–2250, Feb. 2021, ISSN: 1936-0851. DOI: 10.1021/acsnano.0c06902 (cit. on pp. 1, 16).

[11] S. Helgadottir, B. Midtvedt, J. Pineda, A. Sabirsh, C. B. Adiels, S. Romeo, D. Midtvedt, and G. Volpe, "Extracting quantitative biological information from bright-field cell images using deep learning," *Biophysics Reviews*, vol. 2, no. 3, p. 031 401, 2021. DOI: 10.1063/5.0044782. eprint: https://aip.scitation.org/doi/pdf/10.1063/5.0044782 (cit. on pp. 1, 16, 23).

[12] M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, "Single-molecule localization microscopy," *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 39, Jun. 2021, ISSN: 2662-8449. DOI: 10.1038/s43586-021-00038-x (cit. on pp. 2, 4, 23).

[13] D. Hassabis, *The promise of artificial intelligence with demis hassabis - deepmind: The podcast (s2, ep9)*, 2022. [Online]. Available: https://www.youtube.com/watch?v=GdeY-MrXD74 (cit. on p. 2).

[14] G. Storz, Y. I. Wolf, and K. S. Ramamurthi, "Small proteins can no longer be ignored," eng, *Annual review of biochemistry*, vol. 83, pp. 753–777, 2014, PMC4166647[pmcid], ISSN: 1545-4509. DOI: 10.1146/annurev-biochem-070611-102400 (cit. on pp. 2, 9).

[15] N. M. Luscombe, S. E. Austin, H. M. Berman, and J. M. Thornton, "An overview of the structures of protein-dna complexes," *Genome Biology*, vol. 1, no. 1, reviews001.1, Jun. 2000, ISSN: 1474-760X. DOI: 10.1186/gb-2000-1-1-reviews001 (cit. on p. 2).

[16] C. Corbo, R. Molinaro, A. Parodi, N. E. Toledano Furman, F. Salvatore, and E. Tasciotti, "The impact of nanoparticle protein corona on cytotoxicity, immunotoxicity and target drug delivery," en, *Nanomedicine (Lond)*, vol. 11, no. 1, pp. 81–100, Dec. 2015 (cit. on p. 2).

[17] A. Vianello, R. L. Jensen, L. Liu, and J. Vollertsen, "Simulating human exposure to indoor airborne microplastics using a breathing thermal manikin," *Scientific Reports*, vol. 9, no. 1, p. 8670, Jun. 2019, ISSN: 2045-2322. DOI: 10.1038/s41598-019-45054-w (cit. on p. 2).

[18] J. Fritzsche, D. Albinsson, M. Fritzsche, T. J. Antosiewicz, F. Westerlund, and C. Langhammer, "Single particle nanoplasmonic sensing in individual nanofluidic channels," *Nano Letters*, vol. 16, no. 12, pp. 7857–7864, Dec. 2016, ISSN: 1530-6984. DOI: 10.1021/acs.nanolett.6b04124 (cit. on pp. 2, 56).

[19] R. Weissleder and U. Mahmood, "Molecular imaging," *Radiology*, vol. 219, no. 2, pp. 316–333, 2001, PMID: 11323453. DOI: 10.1148/radiology.219.2.r01ma19316. eprint: https://doi.org/10.1148/radiology.219.2.r01ma19316 (cit. on p. 3).

[20] E. Britannica, *History of the Optical Microscope*. Encyclopedia Britannica, Inc. (cit. on p. 3).

[21] X. Zhuang, "Nano-imaging with storm," *Nature Photonics*, vol. 3, no. 7, pp. 365–367, Jul. 2009, ISSN: 1749-4893. DOI: 10.1038/nphoton.2009.101 (cit. on p. 3).

[22] K. Thorn, "A quick guide to light microscopy in cell biology," eng, *Molecular biology of the cell*, vol. 27, no. 2, pp. 219–222, Jan. 2016, 27/2/219[PII], ISSN: 1939-4586. DOI: 10.1091/mbc.E15-02-0088 (cit. on p. 4).

[23] S. Shashkova and M. C. Leake, "Single-molecule fluorescence microscopy review: Shedding new light on old problems," eng, *Bioscience reports*, vol. 37, no. 4, BSR20170031, Jul. 2017, BSR20170031[PII], ISSN: 1573-4935. DOI: 10.1042/BSR20170031 (cit. on p. 4).

[24] B. Špačková, H. Klein Moberg, J. Fritzsche, *et al.*, "Label-free nanofluidic scattering microscopy of size and mass of single diffusing molecules and nanoparticles," *Nature Methods*, vol. 19, no. 6, pp. 751–758, Jun. 2022, ISSN: 1548-7105. DOI: 10.1038/s41592-022-01491-6 (cit. on pp. 4–6, 28, 29, 31, 33, 41, 42, 44, 45, 49).

[25] M. Piliarik and V. Sandoghdar, "Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites," *Nature Communications*, vol. 5, no. 1, p. 4495, Jul. 2014, ISSN: 2041-1723. DOI: 10.1038/ncomms5495 (cit. on p. 5).

[26] M. Piliarik and V. Sandoghdar, "Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites," *Nature communications*, vol. 5, no. 1, pp. 1–8, 2014 (cit. on p. 5).

[27] P. Dechadilok and W. M. Deen, "Hindrance factors for diffusion and convection in pores," *Industrial & engineering chemistry research*, vol. 45, no. 21, pp. 6953–6959, 2006 (cit. on p. 6).

[28] J. A. Marsh and S. A. Teichmann, "Structure, dynamics, assembly, and evolution of protein complexes," *Annual Review of Biochemistry*, vol. 84, no. 1, pp. 551–575, 2015, PMID: 25494300. DOI: 10.1146/annurev-biochem-060614-034142. eprint: https://doi.org/10.1146/annurev-biochem-060614-034142 (cit. on p. 7).

[29] A. Bruce, J. Alexander, L. Julian, *et al.*, *Molecular Biology of the Cell*. 4th edition, 2002 (cit. on p. 7).

[30] M. Prpić, M. Franceschi, M. Romić, T. Jukić, and Z. Kusić, "Thyroglobulin as a tumor marker in differentiated thyroid cancer - clinical considerations," en, *Acta Clin Croat*, vol. 57, no. 3, pp. 518–527, Sep. 2018 (cit. on p. 7).

[31] R. E. Casiday and R. F. Frey, "Iron use and storage in the body : Ferritin and molecular representations iron in biology : Study of the iron content in ferritin , the iron-storage protein .," 2004 (cit. on p. 7).

[32] C. B. Chen, B. Hammo, J. Barry, and K. Radhakrishnan, "Overview of albumin physiology and its role in pediatric diseases," *Current Gastroenterology Reports*, vol. 23, no. 8, p. 11, Jul. 2021, ISSN: 1534-312X. DOI: 10.1007/s11894-021-00813-6 (cit. on p. 7).

[33] H. J. Edenberg and J. N. McClintick, "Alcohol dehydrogenases, aldehyde dehydrogenases, and alcohol use disorders: A critical review," en, *Alcohol Clin Exp Res*, vol. 42, no. 12, pp. 2281–2297, Nov. 2018 (cit. on p. 7).

[34] S. T. Kalinowski, M. J. Leonard, and T. M. Andrews, "Nothing in evolution makes sense except in the light of DNA," en, *CBE Life Sci Educ*, vol. 9, no. 2, pp. 87–97, 2010 (cit. on p. 7).

[35] V. Müller and F. Westerlund, "Optical DNA mapping in nanofluidic devices: Principles and applications," *Lab on a Chip*, vol. 17, no. 4, pp. 579–590, 2017 (cit. on p. 7).

[36] B. TA., "Genomes," *2nd edition. Oxford: Wiley-Liss; 2002. Chapter*, vol. 1, 2002 (cit. on p. 7).

[37] R. J. Simpson, J. W. Lim, R. L. Moritz, and S. Mathivanan, "Exosomes: Proteomic insights and diagnostic potential," *Expert Review of Proteomics*, vol. 6, no. 3, pp. 267–283, 2009, PMID: 19489699. DOI: 10.1586/epr.09.17. eprint: https://doi.org/10.1586/epr.09.17 (cit. on pp. 7, 8).

[38] E. van der Pol, A. N. Böing, P. Harrison, A. Sturk, and R. Nieuwland, "Classification, functions, and clinical relevance of extracellular vesicles," *Pharmacological Reviews*, vol. 64, no. 3, M. P. Mattson, Ed., pp. 676–705, 2012, ISSN: 0031-6997. DOI: 10.1124/pr.112.005983. eprint: https://pharmrev.aspetjournals.org/content/64/3/676.full.pdf (cit. on p. 8).

[39] B. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*, ser. Dover Books on Physics. Dover Publications, 2013, ISBN: 9780486320243 (cit. on p. 8).

[40] V. Filipe, A. Hawe, and W. Jiskoot, "Critical evaluation of nanoparticle tracking analysis (nta) by nanosight for the measurement of nanoparticles and protein aggregates," *Pharmaceutical Research*, vol. 27, no. 5, pp. 796–810, May 2010, ISSN: 1573-904X. DOI: 10.1007/s11095-010-0073-2 (cit. on p. 8).

[41] R. Penrose, *Fashion, Faith, and Fantasy in the New Physics of the Universe*. Princeton University Press, Jun. 2016, ISBN: 9780691178530. DOI: 10.2307/j.ctvc775bn (cit. on pp. 8, 16).

[42] Martín Abadi, Ashish Agarwal, Paul Barham, *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/ (cit. on p. 8).

[43] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017 (cit. on pp. 8, 15).

[44] F. Rapaport, R. Khanin, Y. Liang, M. Pirun, A. Krek, P. Zumbo, C. E. Mason, N. D. Socci, and D. Betel, "Comprehensive evaluation of differential gene expression analysis methods for rna-seq data," *Genome Biology*, vol. 14, no. 9, p. 3158, Sep. 2013, ISSN: 1474-760X. DOI: 10.1186/gb-2013-14-9-r95 (cit. on p. 8).

[45] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery," en, *Br J Pharmacol*, vol. 162, no. 6, pp. 1239–1249, Mar. 2011 (cit. on p. 8).

[46] S. Leonelli, "What difference does quantity make? on the epistemology of big data in biology," *Big Data & Society*, vol. 1, no. 1, p. 2 053 951 714 534 395, 2014, PMID: 25729586. DOI: 10.1177/2053951714534395. eprint: https://doi.org/10.1177/2053951714534395 (cit. on p. 8).

[47] J. Zrimec, C. S. Börlin, F. Buric, A. S. Muhammad, R. Chen, V. Siewers, V. Verendel, J. Nielsen, M. Töpel, and A. Zelezniak, "Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure," *Nature Communications*, vol. 11, no. 1, p. 6141, Dec. 2020, ISSN: 2041-1723. DOI: 10.1038/s41467-020-19921-4 (cit. on p. 8).

[48] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2 (cit. on p. 8).

[49] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory*. Cambridge University Press, May 2022. DOI: 10.1017/9781009023405 (cit. on p. 9).

[50] P. Baldi, *Deep Learning in Science*. Cambridge University Press, 2021. DOI: 10.1017/9781108955652 (cit. on pp. 9, 16).

[51] A. D'Amour, K. Heller, D. Moldovan, *et al.*, *Underspecification presents challenges for credibility in modern machine learning*, 2020. DOI: 10.48550/ARXIV.2011.03395. [Online]. Available: https://arxiv.org/abs/2011.03395 (cit. on pp. 9, 14, 15, 35).

[52] C. Dwork and M. Louise Minow, "Distrust of artificial intelligence: Sources responses from computer science law," in 2022 (cit. on p. 9).

[53] B. Mehlig, *Machine Learning with Neural Networks*. Cambridge University Press, Oct. 2021. DOI: 10.1017/9781108860604 (cit. on p. 10).

[54] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943, ISSN: 1522-9602. DOI: 10.1007/BF02478259 (cit. on p. 10).

[55] C. C. Tappert, "Who is the father of deep learning?" In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2019, pp. 343–348. DOI: 10.1109/CSCI49370.2019.00067 (cit. on p. 10).

[56] H. Wang, B. Raj, and E. P. Xing, "On the origin of deep learning," *CoRR*, vol. abs/1702.07800, 2017. arXiv: 1702.07800 (cit. on p. 10).

[57] J. Kelleher, *Deep Learning*, ser. The MIT Press Essential Knowledge series. MIT Press, 2019, ISBN: 9780262537551 (cit. on p. 10).

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012 (cit. on pp. 10, 15).

[59] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-*

Assisted Intervention (MICCAI), ser. LNCS, (available on arXiv:1505.04597 [cs.CV]), vol. 9351, Springer, 2015, pp. 234–241 (cit. on pp. 10, 16, 17, 27, 29).

[60] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018. arXiv: 1804.02767 (cit. on pp. 10, 30, 31).

[61] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. arXiv: 1512.03385 (cit. on pp. 10, 18).

[62] J. Hawkins and R. Dawkins, *A Thousand Brains: A New Theory of Intelligence*. Basic Books, 2021, ISBN: 9781541675803 (cit. on p. 11).

[63] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989, ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8 (cit. on p. 11).

[64] I. C. Education, *Convolutional neural networkss*, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/convolutional-neural-networks1 (cit. on p. 13).

[65] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 818–833, ISBN: 978-3-319-10590-1 (cit. on pp. 13, 15, 35).

[66] R. Bellman, R. Corporation, and K. M. R. Collection, *Dynamic Programming*, ser. Rand Corporation research study. Princeton University Press, 1957, ISBN: 9780691079516 (cit. on pp. 13, 15).

[67] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965 (cit. on p. 14).

[68] H. Gholamalinezhad and H. Khosravi, "Pooling methods in deep neural networks, a review," *CoRR*, vol. abs/2009.07485, 2020. arXiv: 2009.07485 (cit. on p. 14).

[69] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2013 (cit. on p. 14).

[70] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *CoRR*, vol. abs/1406.4729, 2014. arXiv: 1406.4729 (cit. on p. 14).

[71] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, p. 7 068 349, Feb. 2018, ISSN: 1687-5265. DOI: 10.1155/2018/7068349 (cit. on p. 15).

[72] R. Girshick, J. Donahue, T. Darrell, and J. Malik, *Rich feature hierarchies for accurate object detection and semantic segmentation*, 2013. DOI: 10.48550/

ARXIV.1311.2524. [Online]. Available: https://arxiv.org/abs/1311.2524 (cit. on p. 15).

[73] F. Chollet *et al.*, *Keras*, https://keras.io, 2015 (cit. on pp. 15, 35).

[74] D. Heaven, "Why deep-learning AIs are so easy to fool," *Nature*, vol. 574, no. 7777, pp. 163–166, Oct. 2019. DOI: 10.1038/d41586-019-03013-5 (cit. on p. 15).

[75] S. Kaplan, D. Handelman, and A. Handelman, "Sensitivity of neural networks to corruption of image classification," *AI and Ethics*, vol. 1, no. 4, pp. 425–434, Nov. 2021, ISSN: 2730-5961. DOI: 10.1007/s43681-021-00049-0 (cit. on p. 15).

[76] J. Grauer, F. Schmidt, J. Pineda, B. Midtvedt, H. Löwen, G. Volpe, and B. Liebchen, "Active droploids," *Nature Communications*, vol. 12, no. 1, p. 6005, Oct. 2021, ISSN: 2041-1723. DOI: 10.1038/s41467-021-26319-3 (cit. on p. 16).

[77] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. arXiv: 1405.0312 (cit. on p. 17).

[78] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017, PMID: 28301734. DOI: 10.1146/annurev-bioeng-071516-044442. eprint: https://doi.org/10.1146/annurev-bioeng-071516-044442 (cit. on p. 16).

[79] K. Sung and T. Poggio, "Example based learning for view-based human face detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 39–51, Feb. 1998. DOI: 10.1109/34.655648 (cit. on p. 18).

[80] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012. DOI: 10.1109/TPAMI.2011.155 (cit. on p. 18).

[81] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2722–2730. DOI: 10.1109/ICCV.2015.312 (cit. on p. 18).

[82] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You only look once: Unified, real-time object detection*, 2016. arXiv: 1506.02640 [cs.CV] (cit. on p. 18).

[83] A. B. Oktay and A. Gurses, "Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images," *Micron*, vol. 120, pp. 113–119, 2019, ISSN: 0968-4328. DOI: https://doi.org/10.1016/j.micron.2019.02.009 (cit. on p. 19).

[84] X. wang, T. Xu, J. Zhang, S. Chen, and Y. Zhang, "So-yolo based wbc detection with fourier ptychographic microscopy," *IEEE Access*, vol. PP, pp. 1–1, Aug. 2018. DOI: 10.1109/ACCESS.2018.2865541 (cit. on p. 18).

[85] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," in *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, M. Cord and P. Cunningham, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 21–49, ISBN: 978-3-540-75171-7. DOI: 10.1007/978-3-540-75171-7_2 (cit. on p. 20).

[86] D. A. Roberts, S. Yaida, and B. Hanin, "The principles of deep learning theory," *CoRR*, vol. abs/2106.10165, 2021. arXiv: 2106.10165 (cit. on p. 20).

[87] B. Midtvedt, J. Pineda, F. Skärberg, *et al.*, "Single-shot self-supervised object detection in microscopy," *Nature Communications*, vol. 13, no. 1, p. 7492, Dec. 2022, ISSN: 2041-1723. DOI: 10.1038/s41467-022-35004-y (cit. on pp. 20, 21).

[88] J. Chen, Q. Ye, X. Yang, S. K. Zhou, X. Hong, and L. Zhang, "Few-shot learning for multi-modality tasks," in *Proceedings of the 29th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 5673–5674, ISBN: 9781450386517 (cit. on p. 20).

[89] H.-H. Bock, "Clustering methods: A history of k-means algorithms," in *Selected Contributions in Data Analysis and Classification*, P. Brito, G. Cucumel, P. Bertrand, and F. de Carvalho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 161–172, ISBN: 978-3-540-73560-1. DOI: 10.1007/978-3-540-73560-1_15 (cit. on p. 20).

[90] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," 2015 (cit. on p. 20).

[91] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf, "Convolutional neural networks: A magic bullet for gravitational-wave detection?" *Physical Review D*, vol. 100, no. 6, Sep. 2019. DOI: 10.1103/physrevd.100.063015 (cit. on p. 20).

[92] H. Shen, D. George, E. A. Huerta, and Z. Zhao, "Denoising gravitational waves with enhanced deep recurrent denoising auto-encoders," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2019. DOI: 10.1109/icassp.2019.8683061 (cit. on p. 20).

[93] T. B. Brown, B. Mann, N. Ryder, *et al.*, *Language models are few-shot learners*, 2020. DOI: 10.48550/ARXIV.2005.14165. [Online]. Available: https://arxiv.org/abs/2005.14165 (cit. on p. 21).

[94] A. D. Cohen, A. Roberts, A. Molina, *et al.*, "Lamda: Language models for dialog applications," in *arXiv*, 2022 (cit. on p. 21).

[95] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9630–9640. DOI: 10.1109/ICCV48922.2021.00951 (cit. on p. 21).

[96] D. Silver, A. Huang, C. J. Maddison, *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–503, 2016 (cit. on p. 21).

[97] O. Vinyals, I. Babuschkin, W. M. Czarnecki, *et al.*, "Grandmaster level in starcraft ii using multi-agent reinforcement learning," *Nature*, vol. 575,

no. 7782, pp. 350–354, Nov. 2019, ISSN: 1476-4687. DOI: 10.1038/s41586-019-1724-z (cit. on p. 21).

[98] OpenAI, : C. Berner, *et al.*, *Dota 2 with large scale deep reinforcement learning*, 2019. DOI: 10.48550/ARXIV.1912.06680. [Online]. Available: https://arxiv.org/abs/1912.06680 (cit. on p. 21).

[99] A. Kadian, J. Truong, A. Gokaslan, A. Clegg, E. Wijmans, S. Lee, M. Savva, S. Chernova, and D. Batra, "Sim2real predictivity: Does evaluation in simulation predict real-world performance?" *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6670–6677, 2020. DOI: 10.1109/LRA.2020.3013848 (cit. on p. 21).

[100] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, Nov. 2017. DOI: 10.1109/msp.2017.2743240 (cit. on p. 21).

[101] A. Mirhoseini, A. Goldie, M. Yazgan, *et al.*, "A graph placement methodology for fast chip design," *Nature*, vol. 594, no. 7862, pp. 207–212, Jun. 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03544-w (cit. on p. 21).

[102] R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, and S. G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, no. 6971, pp. 247–252, Jan. 2004, ISSN: 1476-4687. DOI: 10.1038/nature02236 (cit. on p. 21).

[103] L. Wang, W. Chen, W. Yang, F. Bi, and F. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," *IEEE Access*, vol. PP, pp. 1–1, Mar. 2020. DOI: 10.1109/ACCESS.2020.2982224 (cit. on p. 21).

[104] M. Hossam, T. Le, M. Papasimeon, V. Huynh, and D. Phung, *Text generation with deep variational gan*, 2021. DOI: 10.48550/ARXIV.2104.13488. [Online]. Available: https://arxiv.org/abs/2104.13488 (cit. on p. 21).

[105] P. Baniukiewicz, E. J. Lutton, S. Collier, and T. Bretschneider, "Generative adversarial networks for augmenting training data of microscopic cell images," *Frontiers in Computer Science*, vol. 1, 2019, ISSN: 2624-9898. DOI: 10.3389/fcomp.2019.00010 (cit. on p. 21).

[106] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial networks*, 2014. DOI: 10.48550/ARXIV.1406.2661. [Online]. Available: https://arxiv.org/abs/1406.2661 (cit. on p. 22).

[107] R. Rizzuto, M. Brini, P. Pizzo, M. Murgia, and T. Pozzan, "Chimeric green fluorescent protein as a tool for visualizing subcellular organelles in living cells," *Current biology*, vol. 5, no. 6, pp. 635–642, 1995 (cit. on p. 22).

[108] O. Kepp, L. Galluzzi, M. Lipinski, J. Yuan, and G. Kroemer, "Cell death assays for drug discovery," *Nature reviews Drug discovery*, vol. 10, no. 3, pp. 221–237, 2011 (cit. on p. 22).

[109] V. Lulevich, Y.-P. Shih, S. H. Lo, and G.-y. Liu, "Cell tracing dyes significantly change single cell mechanics," *The Journal of Physical Chemistry B*, vol. 113, no. 18, pp. 6511–6519, 2009 (cit. on p. 22).

[110] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nature methods*, vol. 15, no. 11, pp. 917–920, 2018 (cit. on pp. 22, 23).

[111] N. Pillar and A. Ozcan, "Virtual tissue staining in pathology using machine learning," *Expert Review of Molecular Diagnostics*, vol. 22, no. 11, pp. 987–989, 2022, PMID: 36440487. DOI: 10.1080/14737159.2022.2153040. eprint: https://doi.org/10.1080/14737159.2022.2153040 (cit. on p. 22).

[112] X. Li, G. Zhang, H. Qiao, F. Bao, Y. Deng, J. Wu, Y. He, J. Yun, X. Lin, H. Xie, *et al.*, "Unsupervised content-preserving transformation for optical microscopy," *Light: Science & Applications*, vol. 10, no. 1, pp. 1–11, 2021 (cit. on p. 22).

[113] T. C. Nguyen, V. Bui, A. Thai, V. K. Lam, C. B. Raub, L.-C. Chang, and G. Nehmetallah, "Virtual organelle self-coding for fluorescence imaging via adversarial learning," *Journal of biomedical optics*, vol. 25, no. 9, p. 096 009, 2020 (cit. on p. 22).

[114] Z. Korczak, J. Pineda, S. Helgadottir, B. Midtvedt, M. Goksör, G. Volpe, and C. B. Adiels, "Dynamic live/apoptotic cell assay using phase-contrast imaging and deep learning," *bioRxiv*, 2022 (cit. on p. 22).

[115] M. E. Kandel, E. Kim, Y. J. Lee, G. Tracy, H. J. Chung, and G. Popescu, "Multiscale assay of unlabeled neurite dynamics using phase imaging with computational specificity," *ACS sensors*, vol. 6, no. 5, pp. 1864–1874, 2021 (cit. on p. 22).

[116] Y. N. Nygate, M. Levi, S. K. Mirsky, N. A. Turko, M. Rubin, I. Barnea, G. Dardikman-Yoffe, M. Haifler, A. Shalev, and N. T. Shaked, "Holographic virtual staining of individual biological cells," *Proceedings of the National Academy of Sciences*, vol. 117, no. 17, pp. 9223–9231, 2020 (cit. on p. 23).

[117] Y. Liu, H. Yuan, Z. Wang, and S. Ji, "Global pixel transformers for virtual staining of microscopy images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2256–2266, 2020 (cit. on p. 23).

[118] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-storm: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, Apr. 2018. DOI: 10.1364/OPTICA.5.000458 (cit. on p. 24).

[119] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnology*, vol. 36, no. 5, pp. 460–468, May 2018, ISSN: 1546-1696. DOI: 10.1038/nbt.4106 (cit. on p. 24).

[120] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," en, *Nat Methods*, vol. 3, no. 10, pp. 793–795, Aug. 2006 (cit. on p. 23).

[121] H. Shroff, H. White, and E. Betzig, "Photoactivated localization microscopy (PALM) of adhesion complexes," en, *Curr Protoc Cell Biol*, vol. Chapter 4, Unit4.21, Mar. 2013 (cit. on p. 23).

[122] S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophysical Journal*,

vol. 91, no. 11, pp. 4258–4272, 2006, ISSN: 0006-3495. DOI: https://doi.org/10.1529/biophysj.106.091116 (cit. on p. 23).

[123] M. Jung, D. Kim, and J. Y. Mun, "Direct visualization of actin filaments and Actin-Binding proteins in neuronal cells," en, *Front Cell Dev Biol*, vol. 8, p. 588 556, Nov. 2020 (cit. on p. 23).

[124] V. Mannam, Y. Zhang, X. Yuan, and S. Howard, "Deep learning-based super-resolution fluorescence microscopy on small datasets," in *Single Molecule Spectroscopy and Superresolution Imaging XIV*, I. Gregor, R. Erdmann, and F. Koberling, Eds., SPIE, Mar. 2021. DOI: 10.1117/12.2578519 (cit. on p. 23).

[125] E. Wolf. Elsevier, 2008, p. 355, ISBN: 978-0-08-055768-7 (cit. on p. 23).

[126] L. Roeder, *Netron*, https://github.com/lutzroeder/netron, 2020 (cit. on pp. 29, 31, 33).

[127] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2014. DOI: 10.48550/ARXIV.1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980 (cit. on pp. 30, 32, 33).

[128] J. Redmon, *Darknet: Open Source Neural Networks in C*. http://pjreddie.com/darknet/ (20132016) (cit. on pp. 31, 32).

[129] E. Noren, *Pytorch-yolov3*, https://github.com/eriklindernoren/PyTorch-YOLOv3, 2019 (cit. on p. 30).

[130] Z. Lu, J. Lu, Q. Ge, and T. Zhan, "Multi-object detection method based on yolo and resnet hybrid networks," in *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2019, pp. 827–832 (cit. on p. 32).

[131] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, *Object detection with deep learning: A review*, 2019. arXiv: 1807.05511 [cs.CV] (cit. on p. 32).

[132] E. Linder Noren, *Pytorch-yolov3*, https://github.com/eriklindernoren/PyTorch-YOLOv3, 2019 (cit. on p. 32).

[133] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, May 2016, ISSN: 2196-1115. DOI: 10.1186/s40537-016-0043-6 (cit. on p. 34).

[134] M. D. Zeiler and R. Fergus, *Visualizing and understanding convolutional networks*, 2013. DOI: 10.48550/ARXIV.1311.2901. [Online]. Available: https://arxiv.org/abs/1311.2901 (cit. on p. 34).

[135] L. Rokach, "Ensemble methods for classifiers," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA: Springer US, 2005, pp. 957–980, ISBN: 978-0-387-25465-4. DOI: 10.1007/0-387-25465-X_45 (cit. on p. 35).

[136] O. Sagi and L. Rokach, "Ensemble learning: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1249, 2018. DOI: https://doi.org/10.1002/widm.1249. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1249 (cit. on p. 35).

[137] F. Chollet, *Deep Learning with Python*. Manning, Nov. 2017, ISBN: 9781617294433 (cit. on p. 35).

[138] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09, Montreal, Quebec, Canada: Association for Computing Machinery, 2009, pp. 41–48, ISBN: 9781605585161. DOI: 10.1145/1553374.1553380 (cit. on p. 35).

[139] M. Csikszentmihalyi, "Flow: The psychology of optimal experience," in Jan. 1990 (cit. on p. 35).

[140] B. Settles, "Active learning literature survey," 2009 (cit. on p. 35).

[141] A. Ghosh, B. G. Sumpter, O. Dyck, S. V. Kalinin, and M. Ziatdinov, "Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy," *npj Computational Materials*, vol. 7, no. 1, p. 100, Jul. 2021, ISSN: 2057-3960. DOI: 10.1038/s41524-021-00569-7 (cit. on p. 36).

[142] V. La Verde, P. Dominici, and A. Astegno, "Determination of hydrodynamic radius of proteins by size exclusion chromatography," eng, *Bio-protocol*, vol. 7, no. 8, e2230–e2230, Apr. 2017, e2230[PII], ISSN: 2331-8325 (cit. on p. 43).

[143] M. T. Tyn and T. W. Gusek, "Prediction of diffusion coefficients of proteins," *Biotechnology and bioengineering*, vol. 35, no. 4, pp. 327–338, 1990 (cit. on p. 44).

[144] S. Doose, H. Barsch, and M. Sauer, "Polymer properties of polythymine as revealed by translational diffusion," *Biophys J*, vol. 93, no. 4, pp. 1224–1234, Aug. 2007 (cit. on p. 43).

[145] D. N. Petsev, B. R. Thomas, S.-T. Yau, and P. G. Vekilov, "Interactions and aggregation of apoferritin molecules in solution: Effects of added electrolytes," *Biophysical Journal*, vol. 78, no. 4, pp. 2060–2069, 2000, ISSN: 0006-3495. DOI: https://doi.org/10.1016/S0006-3495(00)76753-1 (cit. on p. 43).

[146] I. Sbalzarini and P. Koumoutsakos, "Feature point tracking and trajectory analysis for video imaging in cell biology," *Journal of Structural Biology*, vol. 151, no. 2, pp. 182–195, 2005, ISSN: 1047-8477. DOI: https://doi.org/10.1016/j.jsb.2005.06.002 (cit. on p. 43).

[147] T. Ha, T. Enderle, D. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss, "Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor.," *Proceedings of the National Academy of Sciences*, vol. 93, no. 13, pp. 6264–6268, 1996 (cit. on p. 43).

[148] P. M. Harrison and P. Arosio, "The ferritins: Molecular properties, iron storage function and cellular regulation," en, *Biochim Biophys Acta*, vol. 1275, no. 3, pp. 161–203, Jul. 1996 (cit. on p. 43).

[149] F. Persson, J. Fritzsche, K. U. Mir, M. Modesti, F. Westerlund, and J. O. Tegenfeldt, "Lipid-based passivation in nanofluidics," *Nano Letters*, vol. 12, no. 5, pp. 2260–2265, May 2012, ISSN: 1530-6984. DOI: 10.1021/nl204535h (cit. on p. 46).

[150] U. Satyanarayana, *Biochemistry (2nd ed.)* (Kolkata, India: Books and Allied, 2002 (cit. on p. 46).

[151] S. Gurunathan, M.-H. Kang, M. Jeyaraj, M. Qasim, and J.-H. Kim, "Review of the isolation, characterization, biological function, and multifarious therapeutic approaches of exosomes," *Cells*, vol. 8, no. 4, p. 307, 2019 (cit. on p. 46).

[152] D. Curran-Everett, "Explorations in statistics: The bootstrap," *Advances in Physiology Education*, vol. 33, no. 4, pp. 286–292, 2009, PMID: 19948676. DOI: 10.1152/advan.00062.2009. eprint: https://doi.org/10.1152/advan.00062.2009 (cit. on p. 49).

[153] H. L. Bethlem, G. Berden, F. M. H. Crompvoets, R. T. Jongma, A. J. A. van Roij, and G. Meijer, "Electrostatic trapping of ammonia molecules," *Nature*, vol. 406, no. 6795, pp. 491–494, Aug. 2000, ISSN: 1476-4687. DOI: 10.1038/35020030 (cit. on p. 49).

[154] K. Frykholm, V. Müller, S. KK, K. D. Dorfman, and F. Westerlund, "Dna in nanochannels: Theory and applications," *Quarterly Reviews of Biophysics*, vol. 55, e12, 2022. DOI: 10.1017/S0033583522000117 (cit. on p. 52).

[155] E. Gutierrez-Beltran, P. N. Moschou, A. P. Smertenko, and P. V. Bozhkov, "Tudor staphylococcal nuclease links formation of stress granules and processing bodies with mRNA catabolism in arabidopsis," en, *Plant Cell*, vol. 27, no. 3, pp. 926–943, Mar. 2015 (cit. on p. 53).

[156] J. Pineda, B. Midtvedt, H. Bachimanchi, S. Noé, D. Midtvedt, G. Volpe, and C. Manzo, *Geometric deep learning reveals the spatiotemporal fingerprint of microscopic motion*, 2022. DOI: 10.48550/ARXIV.2202.06355. [Online]. Available: https://arxiv.org/abs/2202.06355 (cit. on pp. 53, 55).

[157] M. Yanagida, "Functional proteomics; current achievements," *Journal of Chromatography B*, vol. 771, no. 1-2, pp. 89–106, 2002 (cit. on p. 54).

[158] V. S. Rao, K. Srinivas, G. N. Sujini, and G. N. S. Kumar, "Protein-protein interaction detection: Methods and analysis," en, *Int J Proteomics*, vol. 2014, p. 147 648, Feb. 2014 (cit. on p. 54).

[159] V. Müller, F. Rajer, K. Frykholm, L. K. Nyberg, S. Quaderi, J. Fritzsche, E. Kristiansson, T. Ambjörnsson, L. Sandegren, and F. Westerlund, "Direct identification of antibiotic resistance genes on single plasmid molecules using CRISPR/Cas9 in combination with optical DNA mapping," *Scientific Reports*, vol. 6, no. 1, p. 37 938, Dec. 2016 (cit. on pp. 54, 55).

[160] P. Horvath and R. Barrangou, "CRISPR/Cas, the immune system of bacteria and archaea," en, *Science*, vol. 327, no. 5962, pp. 167–170, Jan. 2010 (cit. on p. 54).

[161] D. Šulskis, J. Thoma, and B. M. Burmann, "Structural basis of degp protease temperature-dependent activation," *Science Advances*, vol. 7, no. 50, eabj1816, 2021. DOI: 10.1126/sciadv.abj1816. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abj1816 (cit. on p. 55).

[162] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, Jul. 2017. DOI: 10.1109/msp.2017.2693418 (cit. on p. 55).

[163] G. Young, N. Hundt, D. Cole, *et al.*, "Quantitative mass imaging of single biological macromolecules," *Science*, vol. 360, no. 6387, pp. 423–427, 2018. DOI: 10.1126/science.aar5839. eprint: https://www.science.org/doi/pdf/10.1126/science.aar5839 (cit. on p. 56).