



Cross-modality transformations in biological microscopy enabled by deep learning

Downloaded from: <https://research.chalmers.se>, 2025-02-01 02:58 UTC

Citation for the original published paper (version of record):

Hassan, D., Dominguez, J., Midtvedt, B. et al (2024). Cross-modality transformations in biological microscopy enabled by deep learning. *Advanced Photonics*, 6(6).
<http://dx.doi.org/10.1117/1.AP.6.6.064001>

N.B. When citing this work, cite the original published paper.

Cross-modality transformations in biological microscopy enabled by deep learning

Dana Hassan,^{a,b} Jesús Domínguez,^{a,c} Benjamin Midtvedt,^a Henrik Klein Moberg^{©,d}, Jesús Pineda,^a Christoph Langhammer,^d Giovanni Volpe,^a Antoni Homs Corbera,^b and Caroline B. Adiels^{©a,*}

^aUniversity of Gothenburg, Department of Physics, Gothenburg, Sweden

^bCherryBiotech, Research and Development Unit, Montreuil, France

^cElvesys – Microfluidics Innovation Center, Elvessys, Paris, France

^dChalmers University of Technology, Department of Physics, Gothenburg, Sweden

Abstract. Recent advancements in deep learning (DL) have propelled the virtual transformation of microscopy images across optical modalities, enabling unprecedented multimodal imaging analysis hitherto impossible. Despite these strides, the integration of such algorithms into scientists' daily routines and clinical trials remains limited, largely due to a lack of recognition within their respective fields and the plethora of available transformation methods. To address this, we present a structured overview of cross-modality transformations, encompassing applications, data sets, and implementations, aimed at unifying this evolving field. Our review focuses on DL solutions for two key applications: contrast enhancement of targeted features within images and resolution enhancements. We recognize cross-modality transformations as a valuable resource for biologists seeking a deeper understanding of the field, as well as for technology developers aiming to better grasp sample limitations and potential applications. Notably, they enable high-contrast, high-specificity imaging akin to fluorescence microscopy without the need for laborious, costly, and disruptive physical-staining procedures. In addition, they facilitate the realization of imaging with properties that would typically require costly or complex physical modifications, such as achieving superresolution capabilities. By consolidating the current state of research in this review, we aim to catalyze further investigation and development, ultimately bringing the potential of cross-modality transformations into the hands of researchers and clinicians alike.

Keywords: cross-modality transformations; virtual staining; superresolution; deep learning; fluorescence; bright-field; phase contrast.

Received Jun. 18, 2024; revised manuscript received Oct. 9, 2024; accepted for publication Oct. 28, 2024; published online Nov. 27, 2024.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.AP.6.6.064001](https://doi.org/10.1117/1.AP.6.6.064001)]

1 Introduction

Modern optical microscopy methods provide researchers with a window into the microscopic world with visual clarity not possible using traditional bright-field microscopy. While bright-field microscopy relies on light absorption by the sample to generate visual contrast, biological specimens often lack sufficient light absorption for clear, analyzable images.¹ To overcome this challenge, scientists have traditionally employed various staining techniques and specialized microscopy methods tailored to

derive contrast from diverse properties of the sample across different scales, portrayed in Fig. 1. For tissues, researchers use chemical dyes to stain the sample and create contrast.² Similarly, fluorescent dyes are employed to highlight specific cellular structures.^{3,4} At the molecular level, fluorophores are commonly utilized to bind to target molecules, enabling researchers to track and observe individual molecules using fluorescence microscopy.^{5,6}

However, modern microscopy techniques also present challenges. Staining tissue samples is a laborious, invasive, and often irreversible process, resulting in varying staining outcomes for different tissues and limiting their reuse for alternative

*Address all correspondence to Caroline B. Adiels, caroline.adiels@physics.gu.se

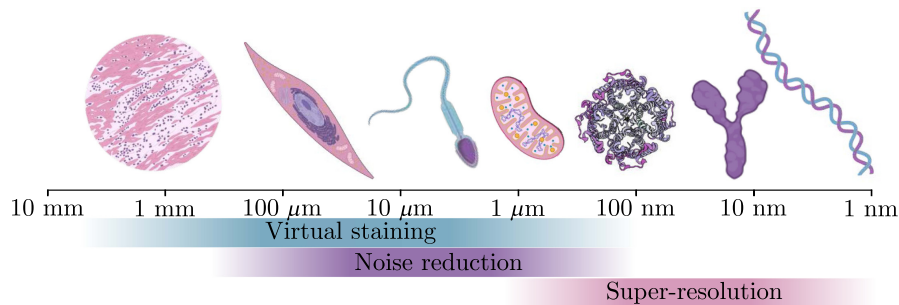


Fig. 1 Applications of cross-modality transformations across biological scales. At the largest scales, virtual staining is used to enhance imaging contrast. At intermediate scales, virtual staining is used in conjunction with noise reduction techniques. At the smallest scales, superresolution is used to study systems far beyond the optical diffraction limit. Image created with the assistance of BioRender.

purposes.⁷ Similarly, imaging cellular and subcellular structures poses challenges, such as costly and time-consuming staining procedures that often limit sample utility.⁸ Furthermore, at the molecular scale, light microscopy encounters limitations in image resolution. Researchers must use specialized objectives, sophisticated setups, and complex fluorophore mixes and buffers to observe sufficiently small structures, resulting in expensive and intricate optical arrangements, e.g., interferometric scattering microscopy⁹ and direct stochastic optical reconstruction microscopy (STORM).^{10,11}

Recently, deep learning (DL) has emerged as a potential solution to overcome the aforementioned challenges in microscopy.¹² By employing neural networks to perform numerical transformations of images between different optical modalities,¹³ researchers can capture images using a low-cost modality, such as bright-field microscopy, and convert them to preferred modality, such as fluorescence microscopy, for simplified analysis.¹⁴ This process, known as cross-modality transformation, eliminates the need for costly and invasive staining procedures,¹⁵ allowing for multiple staining techniques to be produced from the same sample with minimal additional expense. Moreover, since the entire process is numerical, results can be easily replicated by independent teams, ensuring reproducible and reliable outcomes.

In this review, we demonstrate the utilization of cross-modality transformations across biological scales. We outline common strategies for training neural networks for cross-modality transformations, while addressing the specific challenges and possibilities associated with each biological scale. Finally, we summarize the most successful techniques as rules-of-thumb and provide guidelines for the development and utilization of cross-modality transformations.

2 Introduction to DL for Multimodal Transformations

DL is a subset of machine learning that uses artificial neural networks to perform specific tasks. Neural networks are complex computational models processing input data to generate output results. The performance of a neural network is determined by its parameters, commonly referred to as weights, which can range from tens of thousands to hundreds of millions, depending on the application. Primarily, the objective of DL is to optimize these weights by a process called training, enabling

the neural network to yield desired outcomes for a given input space.

In cross-modality transformations, the neural networks used are frequently trained using supervised learning.¹⁶ During this process, the network is presented with an image captured from one modality (e.g., bright-field) and trained to generate the corresponding image in another modality (e.g., fluorescence). Typically, these training data are obtained using either a dual-modality microscope,¹⁷ where the sample is imaged using both modalities, or alternatively, the sample can be imaged twice—before and after a specific treatment (e.g., staining)—with subsequent image processing to align the two views.¹⁸ In certain cases, it may be feasible to derive the transformation analytically, necessitating the imaging of the sample using only one modality to train the network to reconstruct the original image. Even if the physical staining or alternative imaging process is conducted at least once to establish a training pool, subsequent experiments benefit from their simplification.

Cross-modality transformation involves translating images from one modality to another, typically employing encoder-decoder-style fully convolutional neural networks (CNNs). A CNN is a regularized network, meaning it adds information unidirectionally and learns features by itself via kernel optimization. These networks utilize convolutional operations to process input images and are commonly exemplified by architectures, such as U-Net, ResNet, and InceptionNet.

An essential aspect in training neural networks for cross-modality transformation is choice of the loss function, a metric minimized during training. Traditionally, minimizing the mean absolute error (L_1) or mean squared error (L_2) distance between the predicted and ground truth images is common.¹⁹ However, this approach often yields low-resolution and nonphysical results. To address this, an auxiliary adversarial loss function is frequently incorporated.²⁰ This involves training a discriminator neural network alongside the main generator network, where the discriminator distinguishes between generated and ground truth images. The generator is trained to deceive the discriminator by producing physically reasonable results. Such networks are referred to as generative adversarial networks (GANs).

Alternatively, diffusion models represent a recent approach generative modeling. These models utilize probabilistic generative techniques in a two-step process: forward diffusion, where noise is iteratively added to images until they become pure Gaussian noise; and reverse diffusion, where images are

iteratively denoised using a neural network. By conditioning the reverse diffusion process, diffusion models effectively handle image-to-image transformation tasks.²¹ While diffusion models can produce higher-quality images compared to GANs, they come with significantly higher computational costs.

3 Tissue Imaging/Histology

Histological staining is a cornerstone of clinical pathology and research, playing a pivotal role in unraveling tissue details at the microscopic level. It enables the visualization of structures crucial for medical diagnosis, scientific study, autopsy, and forensic investigation.²² In recent years, DL advances have revolutionized tissue imaging and histology analysis, offering innovative solutions to overcome the limitations of traditional physical staining methods.^{23,24} In the following sections, we will explore the transformative impact of DL techniques in substituting conventional staining approaches, with the aim of improving the analysis of histological samples.

3.1 Limitations of Chemical Staining

One of the most impactful applications of DL for cross-modality transformation is in histology, where visual tissue analysis often faces challenges associated with traditional physical staining. Tissues, as the largest biological structures routinely observed through optical microscopy, require staining protocols to create visual contrast between features. However, these protocols often rely on chemical dyes that can be hazardous and may adversely affect the samples, especially during critical steps when sample structures are vulnerable.^{25–27} In addition, manual labor and dexterity are necessary, and especially for fluorescence dyes, not all staining procedures are compatible in the same sample, limiting the information obtainable.

The histological process usually comprises several steps, including fixation, embedding, sectioning, staining, and mounting, although the specific steps may vary according to the staining technique and the target tissue. The first step is fixation, where the full tissue sample is preserved using chemicals, such as formaldehyde or glutaraldehyde. Fixation prevents decay and maintains the structural integrity by cross-linking the proteins in the sample. However, the tissue's original chemistry is altered. An alternative approach is to freeze the sample, often using liquid nitrogen, which can preserve the natural state of proteins and lipids without chemicals. The next step is to dehydrate the tissue sample through a series of diluted alcohol solutions and to clear it, using different clearing agents that dissolve remnant lipids and simultaneously homogenize the refractive index. This process renders the tissue transparent as a consequence of even light scattering across the sample,²⁸ and prepares it for infiltration, but can cause tissue shrinkage and morphological alterations. The sample can now be completely encased in an embedding medium, such as paraffin wax, and left to solidify. Once hardened, the sample is cut using a microtome into very thin slices, around 4 to 10 μm thick, in a step known as sectioning. This process requires high skill and precision, as incorrect microtome alignment or use can lead to tearing or crushing of the tissue and obscure important details. Finally, the slices are placed on microscopic glass slides for observation. Handling must avoid stretching or folding, which can distort the samples and hinder analysis. Once mounted, the samples are ready for the next step: the staining. During staining, various stains are applied based on the cellular structures of interest. Among

histological stains, hematoxylin and eosin (H&E) are among the most widely used, with hematoxylin staining cell nuclei purple and eosin coloring the cytoplasm and extracellular matrix pink. Other stains target different structures, such as Picrosirius Red (PSR) for collagen fibers or Alcian Blue for acidic mucins and cartilage. An important technique within this process is immunohistochemistry (IHC), which detects specific proteins in the sample using antibodies. This method requires antigen retrieval to unmask target proteins after fixation, alongside a blocking step to reduce nonspecific antibody binding. The staining process is sensitive to factors such as concentration and timing, and if applied unevenly, it can obscure details. Other limitations of IHC regard unspecific binding or cross-reactions of the added antibodies, which can negatively affect the outcome. Finally, the sample is prepared for detailed observation, and the histological features can be identified via microscopy imaging.

Traditionally a qualitative method, chemical staining can also yield quantitative data through image analysis,²⁹ such as measuring staining intensity, to estimate the presence of biomolecules. However, identifying biological features often requires additional context and expert analysis, and variability in staining intensity, reagent quality, and human interpretation can affect results. Standardized protocols and software tools can mitigate some of this variability, though issues persist compared to virtual staining and inherent contrast techniques.

As an alternative to chemical staining, inherent contrast techniques—such as phase contrast, differential interference contrast (DIC),^{30,31} and quantitative phase imaging (QPI)—provide an alternative to chemical staining by exploiting refractive index variations in biological tissues to enhance contrast.³² These methods require minimal additional equipment but lack the specificity of chemical dyes and are prone to optical artifacts, such as halo effects.³³ To improve specificity, they are often combined with other techniques. The choice between virtual staining and inherent contrast methods depends on factors, such as cost-effectiveness and imaging quality.

3.2 DL for Tissue Imaging

DL models can be trained to virtually stain samples, whether they are unstained³⁴ or have been stained using a different method.³⁵ By bypassing the physical staining process, multiple readouts equivalent to different dyes can be obtained from the same image. This not only maximizes information output³⁶ for analysis and diagnostics³⁷ but also simplifies the experimental setup requirements, as shown in Fig. 2.

In histology, DL models are trained to virtually stain tissue using collections of stain/unstained image pairs as, a reference,^{36,37,39–47} as shown in Fig. 3(a). This method constitutes a pure virtual approach to staining samples, although the starting point is not always an unstained sample. Some models have the capability to transform one type of staining into another.³⁶ Hong et al., as shown in Fig. 3(b),⁴⁸ washed and restained the samples to obtain equivalent pairs for a stain-to-stain translation model. Still, there are cases of cross-modality transformation of tissues based on unpaired samples.⁴⁹ These cases are notably more intricate as they typically do not rely on training pools consisting of paired samples,⁵⁰ but rather on disparate samples obtained through different techniques or dyes. In alternative scenarios, a multistain model,⁵¹ such as the one shown in Fig. 3(c), is developed, where the model is trained to virtually apply various

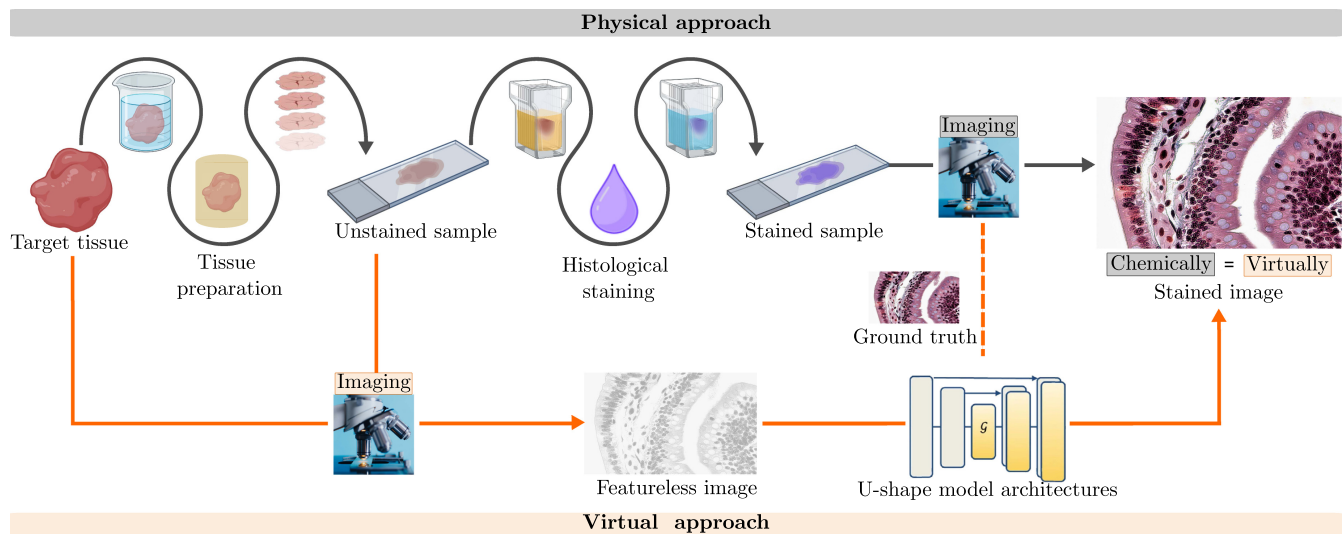


Fig. 2 Contrast between physical and virtual approaches to obtain a stained image. In the physical approach, the sample undergoes a series of complex procedures, including preparation, staining, and imaging. Tissue preparation may involve fixing, embedding, and sectioning, among other steps. Similarly, histological staining of an unstained sample requires permeabilization, chemical dye application, washing, counterstaining, and protocol optimization before imaging. In contrast, virtual staining offers a simplified alternative to these protocols, eliminating the need for physical processing³⁷ or staining of the sample.³⁸ In the virtual approach, an unaltered or unstained sample is processed through a virtual staining network to generate a stained image, with results equivalent to physical staining. Physically stained images serve as training data, or input, for the model, especially when transforming between different stains is the objective. Created with the assistance of BioRender. (Tissue image adapted from Berkshire Community College Bioscience Image Library.)

dyes to an unstained sample, enabling it to transform a stained image into each of the other dye types.³⁶ An illustrative example is provided by Rivenson et al.,⁵² who used a CNN to transform wide-field autofluorescence images of unlabeled tissue sections⁵³ into images equivalent to the bright-field images of histologically stained versions of identical samples. Their study demonstrates the feasibility of this approach to generate multiple types of stains on different tissue types through the autofluorescence signal.

Therefore, virtual and histological stainings are not mutually exclusive and can be complementary. While virtual staining offers convenience and reproducibility, it still relies on histological staining to provide the ground truth for generating large training data sets. The main trends are summed up in Table 1 in Sec. 6 guidelines. Some models benefit from existing databases of images of stained tissues for their training, reducing the manual effort required to obtain a sufficient training set.^{34,61,62} Virtual staining also offers advantages over traditional histology, including the potential for real-time staining of tissue samples⁶³ and three-dimensional (3D) reconstructions of full tissues.⁶³ In the latter case, Wang et al.⁶³ virtually stained light-field microscopy—not to be confused with bright-field microscopy—images of volumetric samples, merging two typically incompatible techniques. Furthermore, certain virtual staining models extend their capabilities by integrating segmentation of the highlighted region of interest.^{48,64–66} Segmentation, and potentially staining protocols, can be applied in imaging techniques where usually contrast-enhancing staining is not available due to cross-modality transformations, as shown in Fig. 3(d). In this example from Dou et al.,⁴⁹ segmentation, obtained through artificial

intelligence in magnetic resonance imaging (MRI) measurements, is transferred to CT images. This is particularly relevant for the utilization of machine-learning models for diagnostic purposes, facilitating the differentiation of healthy tissue samples from those associated with conditions, such as infections or tumors.^{61,62,67,68} Moreover, in certain cases,³⁷ not only diagnosis is accomplished through virtual staining, but also the direct acquisition of suitable input images from patients via noninvasive methods, as represented in Fig. 3(e), underscoring the potential of this powerful technique. While most cross-modality transformations in tissues are based on the same imaging modality with different nonfluorescent dyes, some examples involve transformation into fluorescent dyes³⁵ or other imaging techniques. Nevertheless, fluorescence is of greater relevance at the cellular level, where emphasis is placed on specific structures and organelles.

Virtual staining models must undergo rigorous validation against chemically stained samples to ensure accurate representation of biological features, as variations in color, texture, and detail may occur. A comprehensive comparison between traditional and virtual staining methods is therefore crucial for assessing reliability,⁶⁹ which requires either manual or automated ground-truth data annotation. Objective metrics, such as index measure (SSIM)^{51,70–72} and peak signal-to-noise ratio (PSNR),^{73–75} are used to measure accuracy, while expert visual evaluations are essential to confirm that virtual stains accurately reflect key histological features for research and diagnostic purposes.

In general, a single dye is insufficient to provide comprehensive information about a particular tissue sample. Instead,

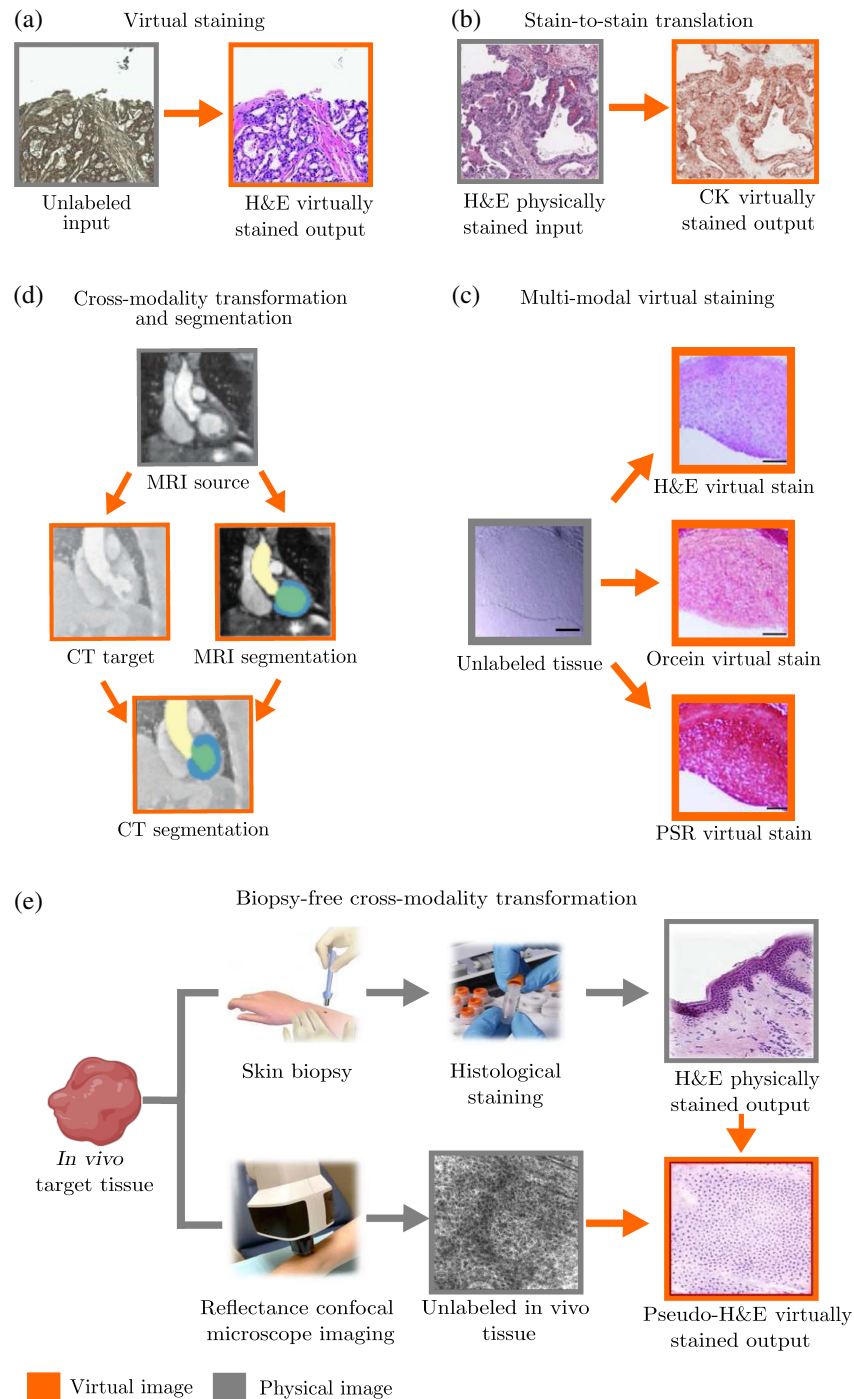


Fig. 3 Representative applications of cross-modality transformations for tissue imaging using DL. (a) Virtual staining of an unlabeled sample image to obtain the equivalent H&E stained image. Adapted from Rana et al.³⁹ (b) Stain-to-stain translation where the input and output are images from two different staining procedures, in this case H&E to IHC staining for cytokeratin (CK). Adapted from Hong et al.⁴⁸ (c) Multi-stain model that is able to transform unlabeled tissue images into different staining options simultaneously: H&E, orcein, and PSR. Adapted from Li et al.⁴⁰ (d) Cross-modality transform to apply a segmentation method, or potentially a stain, in a previously incompatible modality. In this case, an AI segmentation for MRI images is transcribed to CT images. Adapted from Dou et al.⁴⁹ (e) Biopsy-free cross modality transformation, where not only the staining procedure but also the sample preparation is avoided. Using CRM as a noninvasive technique for *in vivo* measurements, the resulting images incorporate features comparable to H&E, despite being incompatible with traditional staining techniques in such conditions. Adapted from Li et al.³⁷

various dyes can be applied on different samples of the same original tissue, such as different slices of a single specimen. In the study by Li et al.,⁴⁰ three distinct dyes—H&E, PSR, and orcein—the last used to demonstrate elastic fibers—were utilized on carotid tissue. Each dye targets different components of artery tissue, aiding in the identification of coronary artery disease and vascular injuries. First, an independent model was trained for each dye to virtually stain an unstained sample. Then, a complete model was trained to simultaneously produce all three modalities from the original sample. This was accomplished by applying one of the three corresponding staining protocols to the originally unstained samples, generating pairs of stained and unstained images for each type. A total of 60 (whole slide images) of each stain, along with their unstained equivalents, were utilized, yielding 1500 to 1800 divided images for training and 150 to 200 images for validation for each staining protocol. Following standard practice, a conditional generative adversarial network (cGAN) was implemented to learn the generation of stained images from the acquired data set. The generator architecture was based on U-Net, while the discriminator comprised a PatchGAN architecture. To accomplish virtual staining according to three distinct protocols, the StarGAN⁷⁶ architecture was implemented. This framework enables image-to-image translations across multiple domains using only a single model, offering practically unlimited potential for utilizing unstained samples, as they can potentially be transformed into any other protocol with an appropriately trained network.^{36,77}

These studies suggest that DL holds significant potential for histological staining, yet its widespread adoption remains limited. Though DL has emerged as a leading choice for analyzing and interpreting histology images with the potential to enhance medical diagnostics,⁷⁸ very few algorithms have transitioned to clinical implementation.⁷⁹ Several challenges persist, notably the need for accurate labeling and addressing variations in slide colors,^{80–82} as addressed in Sec. 6. Looking ahead, the availability of high-throughput experimental devices is crucial for optimizing the performance of DL-based digital histopathology methods. Fanous et al.⁸³ showed the potential of DL to speed up data acquisition in scanning microscopes, using a GAN-based image restoration approach to reconstruct motion-blurred scanned images.⁸⁴ This suggests that DL-based approaches could significantly improve experimentation efficiency and speed, thereby improving algorithm performance. In addition, integrating multiple modalities, such as molecular profiling information,^{85,86} could further enhance the accuracy of disease classification and prognosis accuracy in digital pathology. Furthermore, the development of transfer-learning techniques that profit pretrained models on large data sets could mitigate the challenge of limited training data in certain biological applications, expanding DL's utility in digital pathology. Promising advancements are expected in the use of diffusion models for data augmentations and synthetic image generation, as demonstrated by Moghadam et al.,⁸⁷ where the generated histopathology images were indistinguishable by experienced pathologists.

Diffusion models have emerged as a powerful tool in histology for virtual staining, addressing some of the limitations seen in traditional models, such as GANs and encoder-decoder networks. For instance, StainDiff is a diffusion probabilistic model designed to improve the stain-to-stain transformations by overcoming issues, such as mode collapse, where the generator of a GAN produces images based on a limited range of the training samples and posterior mismatching found in other networks.⁸⁸

These models are also applied to generate virtual IHC images from H&E stained slides, as seen in PST-Diff, which ensures structural and pathological consistency through mechanisms such as asymmetric attention and latent transfer.⁸⁹ Despite their potential, diffusion models generally require large data sets, making them less feasible for histological applications with limited data. To address this, multitask architectures such as StainDiffuser have been developed to simultaneously generate cell-specific stains and segment cells, optimizing the performance even with constrained data sets.⁹⁰ In addition, advanced methods, such as virtual IHC multiplex staining, utilize large vision-language diffusion models to generate multiple IHC stains from a single H&E image, addressing tissue preservation challenges often faced in biopsies.⁹¹ However, challenges remain. Diffusion models have shown limitations in unpaired image translation tasks, such as slide-free microscopy virtual staining, where the sample preparation process is bypassed along with the staining process. In such cases, they underperform compared to models such as CycleGAN without additional regularization, highlighting the need for further refinement in certain applications.⁹² Despite these challenges, diffusion models show great promise in virtual staining, with ongoing research focused on enhancing their reliability and applicability in histology.

4 Cellular and Subcellular Structure Imaging

Biologists and clinical laboratories routinely employ optical microscopy to examine cell cultures, enabling the study of cellular and subcellular morphologies and physiology. This examination helps in understanding intercellular communication networks, dynamic cell behaviors, and pathophysiological mechanisms.⁹³ For instance, changes in the morphological characteristics of cellular structures serve as effective indicators of a cell culture's physiological status and its response under drug exposure.^{94,95} In the subsequent sections of this review, we delve into the limitations of fluorescence staining techniques, shedding light on the challenges associated with both fixed and live staining methods. In addition, we explore how DL approaches are revolutionizing cellular imaging analysis, offering innovative solutions to overcome these limitations and ushering in a new era of advanced and automated cell culture investigations.⁹⁶

4.1 Limitations of Fluorescence Staining

Standard cell imaging workflows typically rely on fluorescence microscopy, employing either fixed or live fluorescent staining techniques to highlight specific cell structures. Despite their widespread use, both fixed and live fluorescent staining methods have limitations. These procedures can be invasive and toxic, potentially impacting cell health and behavior.⁹⁷ In fixed staining, as for tissue, the fixation process itself can introduce artifacts by altering the native state of cellular components. In addition, the use of permeabilization methods compromises cell membrane integrity.⁹⁸ Furthermore, fixed staining provides only a static view of cellular processes, limiting the ability to study dynamic processes. Conversely, live staining, while theoretically preserving the native state of cells, often alters their biological activity and can be toxic.⁹⁹ The availability of specific and effective live-staining dyes can also be limiting, restricting the visualization of certain cellular components. In addition, real-time observation of cellular processes in live staining

may be challenging due to phototoxicity and photobleaching over extended imaging periods.¹⁰⁰ Lastly, the use of multiple fluorophores can lead to spectral cross talk between fluorescence channels, potentially resulting in misleading results and complicating image analysis. These challenges hinder the acquisition of reliable longitudinal data, which is often crucial for studying the effects of drug exposure over time.¹⁰¹

4.2 DL for Cellular and Subcellular Imaging

Recently, research has proposed the use of DL as an alternative to conventional physical staining methods to mitigate inherent problems. These works suggest replacing physical staining and fluorescence microscopy with a neural network that generates virtual fluorescence-stained images from unlabeled samples.

Virtual cell staining has been achieved from various imaging modalities, including phase contrast,^{55,102} QPI,¹⁰³ and holographic microscopy.¹⁰⁴ Moreover, recent studies have shown that bright-field images, despite their limited detail, contain sufficient information for a CNN to reproduce different types of staining.

For example, Ounkomol et al.¹⁰⁵ introduced a CNN-based framework to map the relationship between paired 3D bright-field and fluorescence live-cell images for various key subcellular structures (e.g., DNA, cell membrane, nuclear envelope, and mitochondria). Each cellular component is modeled separately, with a U-Net trained independently for each one. The training process minimizes the mean-squared error between the ground-truth fluorescence image and the predicted image. Once trained, these models can be combined, allowing a single 3D bright-field input to generate multichannel, integrated fluorescence images across multiple subcellular structures. Particularly advantageous, the training data require relatively few paired examples (solely 30 pairs per structure), lowering the machine-learning entry barriers.

The work by Helgadottir et al.⁵⁴ offers yet another compelling example of the potential of virtual staining of cellular structures from bright-field images. Similar to other approaches, this method relies on a modified version of the U-Net to learn the cross-modality mapping. However, it enhances the reconstruction accuracy by incorporating GAN-based training.

GANs have become a widely adopted framework in virtual cell staining due to their capacity to generate high-quality, realistic images.^{54,55,102–104} Particularly, Helgadottir et al. employed a conditional GAN to virtually stain lipid droplets, cytoplasm, and nuclei from bright-field images of human stem-cell-derived adipocytes. The generator network, a U-Net, processes a stack of bright-field images captured at multiple z positions and generates virtually stained fluorescence images. The architecture features three independent decoding paths, each dedicated to one cellular component, to effectively decorrelate the features associated with the predicted fluorescence images. The discriminator, a CNN, is trained to distinguish between the synthetically generated virtual stains and actual fluorescently stained samples (conditioned on the input bright-field image). These two neural networks are trained simultaneously until the generator can fool the discriminator by producing images that closely mimic real fluorescence [Fig. 4(a)]. An interesting feature of Helgadottir's method is its robustness and rapid convergence; the neural

network requires relatively few epochs to quantitatively reproduce the corresponding cell structures [Fig. 4(b)].

While GANs have proven effective in enhancing the performance of virtual staining networks for various applications, they rely on co-registered input and ground-truth images. Nevertheless, obtaining perfectly co-registered training pairs is often challenging due to the rapid dynamics of biological processes or the incompatibility of different imaging modalities. To address this limitation, Li et al.⁵⁵ introduced unsupervised content-preserving transformation for optical microscopy (UTOM). This approach utilizes a CycleGAN to transform images between domains without requiring paired data. Unlike traditional GAN models, CycleGANs employ two generator-discriminator pairs, one for each domain, to learn bidirectional mappings between imaging modalities [Fig. 4(c)]. UTOM has been applied, among other examples, to the virtual staining of phase-contrast images of differentiated human motor neurons, notably delivering competitive performance compared to a CNN architecture trained on paired samples under supervision despite the lack of paired training data [Fig. 4(d)].

Importantly, although the architecture and training of the neural network play a decisive role in the performance of virtual staining models, the input imaging modality must capture sufficient contrast of the different cell structures, providing the network with enough information to learn the transformation to the desired high-contrast, high-specificity fluorescently stained samples.

Recent research has centered on the development of optical systems that capture the rich structural details of cells and embed inductive bias within the network to enhance its performance. For instance, Cheng et al.¹⁰⁶ profited from the rich structural information and high sensitivity in reflectance microscopy to boost the performance of virtual staining models. Specifically, the authors employed an LED array reflectance microscope to acquire co-registered label-free reflectance and fluorescence images.¹⁰⁷ This platform collects four dark-field reflectance images using half-annulus LED patterns oriented in different directions (top, bottom, left, and right). These measurements derive two dark-field reflectance differential phase contrast (drDPC) images computed along orthogonal orientations [Fig. 4(e)]. Interestingly, the oblique illumination dark-field and drDPC images provide complementary contrast information. While raw dark-field images highlight subcellular structures, such as nuclei, nucleoli, and hyperreflective areas near the nuclear periphery, the drDPC images emphasize cell membranes with clearly defined boundaries. These images serve as multichannel input for the virtual staining model, which, boosted by the enhanced resolution and sensitivity in the back-scattering data, provide a reliable prediction of subcellular features [Fig. 4(f)].

In a similar vein, Cooke et al.¹⁰⁸ proposed incorporating a physical model of the experimental microscope into the virtual staining model. This approach utilizes a CNN which incorporates a “physical layer” representing the microscope's illumination model. Consequently, during training, the network learns task-specific LED patterns that significantly enhance its ability to infer fluorescence image information from label-free transmission microscopy images. This work, in particular, further underscores the importance of rich input data and highlights the potential combination of programmable optical elements and physics-informed DL to open new possibilities for exploring the structure and function of cells.

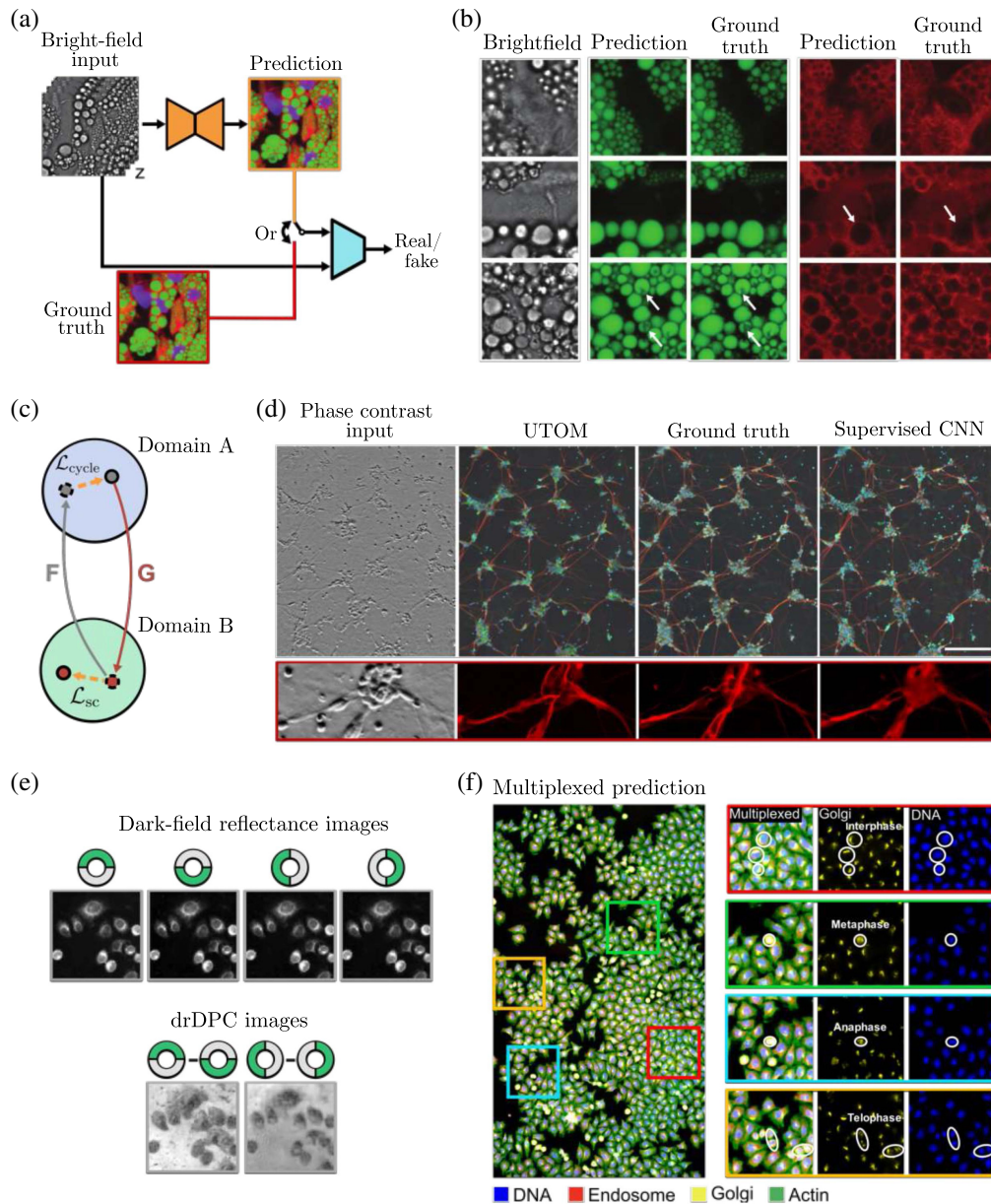


Fig. 4 Virtual cell staining using DL. (a) Helgadottir et al. introduced a cGAN to virtually stain lipid droplets, cytoplasm, and nuclei using bright-field images of human stem-cell-derived fat cells (adipocytes). The U-Net-based generator processes bright-field image stacks captured at various z positions to generate virtually stained fluorescence images. A CNN-based discriminator is trained to differentiate between the virtually generated stains and real fluorescently stained samples, conditioned on the input bright-field image. (b) The virtual staining of lipid droplets (green channel) and cytoplasm (red channel) exhibits a high degree of fidelity, as evidenced in the fine details of the lipid droplet internal structure and the enhanced contrast among distinct cytoplasmic components (highlighted by the arrows). Panels (a) and (b) adapted from Helgadottir et al.⁵⁴ (c) Unsupervised cross-modality image transformation using UTOM. Two GANs, G and F , are trained concurrently to learn bidirectional mappings between image modalities. The model incorporates a cycle-consistency loss (L_{cycle}) to ensure the invertibility of transformations, while a saliency constraint (L_{sc}) preserves key image features and content in the generated outputs. (d) UTOM achieves performance comparable to a supervised CNN trained on paired samples, without requiring paired training data. Panels (c) and (d) adapted from Li et al.⁵⁵ (e) Co-registered label-free reflectance and fluorescence images acquired using a multimodal LED array reflectance microscope. (f) Multiplexed prediction displaying DNA (blue), endosomes (red), the Golgi apparatus (yellow), and actin (green). Zoomed-in views, with white circles, highlight representative cell morphology across different phases of the cell cycle. Adapted from Cheng et al.¹⁰⁶

5 Molecular Imaging

One of the most significant advancements in molecular imaging, which involves the optical imaging of single biological molecules at micro- and nanoscales, has been the introduction of fluorescence microscopy techniques.³ However, unlike in tissue and cellular imaging, the lack of viable techniques for studying molecules without fluorescence currently prevents virtual staining in molecular imaging. Rather, the focus of cross-modality transformations in molecular imaging typically revolves around superresolution microscopy aiming to surpass diffraction-imposed limits in imaging molecules.¹⁶ Traditionally, achieving such high resolutions requires either expensive microscopy setups with specialized objectives, complex numerical estimations of the imaging process, or specific fluorophores.¹⁰⁹ Nevertheless, recent research has indicated that DL-based approaches using generative learning (Sec. 2) can enhance the resolution of images captured with ordinary objectives, comparable to those obtained with costly specialized objectives. Further, DL-driven cross-modality transformations have demonstrated the ability to achieve superresolution across various microscope modalities.^{16,110–112} In this section, we first survey the physics underlying the spatial resolution limitations and the subsequent methods for achieving superresolution, aiming to overcome these constraints. Thereafter, we present an overview of the relatively recent work in applying DL techniques to transform images into their superresolved counterparts, particularly focused on, but not limited to, applications in molecular imaging.

5.1 Physics of Superresolution Microscopy

When light from a point-like light source (an object with diameter D_1) traverses a lens, it undergoes diffraction, producing a characteristic pattern known as the Airy disk. This pattern comprises a bright central region surrounded by concentric rings of diminishing intensity (D_2). The Airy disk represents the smallest focal point achievable by a light beam. Below the object and Airy disk representations, corresponding intensity plots illustrating the point spread functions (PSFs) are displayed. As Airy patterns reach a point of significant interference, causing a reduction in contrast, they merge, becoming indistinguishable and limiting the spatial resolution. Spatial resolution, the shortest physical distance between two points within an image, stands out as the single most important feature in optical microscopy.¹¹³ The primary constraints affecting the achievable spatial resolution stem from an intrinsic phenomenon of diffraction physics. Regardless of lens quality or optical component alignment, a microscope's resolution ultimately correlates

with the wavelength of the detected scattered light and inversely with the numerical aperture (NA) of its objective. This relationship is shown in Fig. 5(a), where light from the sample traverses an objective to the image plane, generating a fundamentally limited diffraction pattern known as a PSF. The PSF inherently limits the minimal distance between two discernible points in the sample, shown in Fig. 5(b). The full width at half-maximum of a PSF in the lateral directions can be approximated as $\text{FWHM} = 0.61 \frac{\lambda}{\text{NA}}$, where λ represents the light's wavelength, and NA denotes the numerical aperture of the objective. Thus, for a typical oil immersion objective with $\text{NA} = 1.4$, the resulting PSF has a lateral size of 200 nm and an axial size of 500 nm, effectively restricting the resolution to this range for visible-light studies.¹¹⁴ Comparing these scales with those depicted in Fig. 1, it becomes evident that the diffraction limit rarely poses a challenge in most imaging at organ, tissue, or even cellular levels. However, in cellular exploration, where subcellular and molecular structures are of interest, issues regarding diffraction limits become prominent. These issues are exacerbated by the typically dense distribution of molecules and subcellular structures, causing their PSFs to overlap, thus blurring many intricate details together. Hence, the development of superresolution techniques that surpass the diffraction limit becomes imperative for further exploration of these structures using noninvasive optical light. Various microscope techniques have been developed to overcome this limitation, including single-molecule localization microscopy (SMLM) methods, such as STORM,¹¹⁵ photo-activated localization microscopy (PALM),¹¹⁶ and fluorescence photoactivation localization microscopy.¹¹⁷ Other methods of transcending the standard resolutions of microscopes exist, including complex numerical estimations of point spread (transfer) functions seeking to estimate the diffraction behavior, illumination pattern engineering methods reducing the PSF size,¹¹⁸ as well as specialized fluorophores.¹⁰⁹ However, these approaches pose their own challenges, including complex and multivariate dependencies on imaging conditions, making solving diffraction integrals of PSFs exceedingly difficult for practically relevant systems,¹¹⁹ as well as increased costs associated with the aforementioned fluorophores.¹²⁰

In recent years, another promising avenue for achieving super-resolution has emerged as a consequence of the astounding growth and success of DL-based computer vision algorithms. Analogous to the cross-modality transforms mentioned above, the DL approach to superresolution involves training neural networks to transform one imaging modality (regular-resolution images) to another (superresolved images). Some of these approaches utilize generative learning, effectively learning

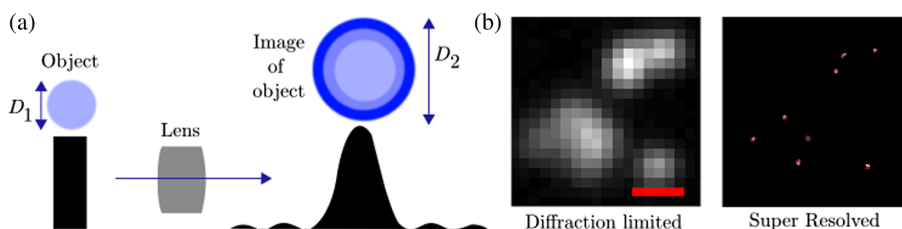


Fig. 5 Superresolution physical principles. (a) Illustration of the PSF resulting from imaging object of diameter D_1 below the diffraction limit of an optical system, leading to an image of diameter D_2 . (b) Low-resolution image of simulated emitters alongside ground-truth emitter positions. Image reproduced with permission from Ref. 56.

the complex interpolation function between regular- and super-resolved images, or through direct supervised learning, for example, by estimating the positions of underlying diffraction-limited emitters. The specific techniques for training these networks vary considerably across applications, as elaborated upon further below.

5.2 DL for Superresolution Microscopy

In general, DL for superresolution can be categorized into two approaches, each with two learning paradigms. The first approach aims to enhance resolution by training end-to-end, directly transforming low-resolution images into high-resolution ones. This can be achieved through supervised learning, using pairs of simulated or experimentally measured images from the same sample at different resolutions to train neural networks, or through unsupervised learning, where only low-resolution or high-resolution images are obtained, either experimentally or through simulations. The other approach seeks resolution enhancement by training a network to output the position of each individual molecule (or equivalent scattering object) within an image, and then reconstructing the high-resolution image from these positions, thus transcending the diffraction limits. This approach can also be trained either in a supervised or unsupervised fashion. A summary of the different models and their characteristics can be found in Table 1 in Sec. 6 guidelines. Once such a network is trained, it can swiftly generate high-resolution images without the need for parameter adjustment, yielding an efficient algorithm for improving image resolution within a specific modality.^{121–126}

5.2.1 End-to-end superresolution mapping

One common approach for supervised end-to-end low- to high-resolution mapping involves pre-upsampling the low-resolution image using a traditional upsampling interpolating algorithm, followed by training a CNN to refine the upsampled image until accurate superresolution is achieved. This approach, initially implemented for single-image superresolution,¹²⁷ has been used in various biological applications, such as enhancing the resolution of magnetic resonance (MR) images^{128–132} and X-ray computed tomography (CT) images.¹³³

Another approach is to apply superresolution to the image after it undergoes computationally intensive CNN layers. This reduces the overall computational burden, as most of the computations are performed on low-resolution images. This approach, known for its efficiency, was first introduced by Dong et al.¹³⁴ and has also been applied in various biological contexts, including superresolution of X-ray images,¹³⁵ endoscopy images,^{136,137} cardiac images,¹³⁸ and MR images.¹³¹

A different strategy for end-to-end low- to high-resolution mapping involves iteratively up- and downsampling the image using downsampling convolutional layers and upsampling transposed convolutional layers. This technique utilized in the “back projection” networks presented by Harris et al.,¹³⁹ incorporates an error feedback mechanism for projection errors at each iteration stage. Each up and downsampling stage is mutually connected through concatenation, reflecting the mutual dependence of low- and high-resolution image pairs, for which the authors demonstrated yield superior results across multiple data sets, as outlined in Table 1 in Sec. 6 guidelines. This approach has also been applied in various biological applications, such as transformation of CT scan brain images into higher-

resolution MRI images^{140,141} for the detection of multiple sclerosis¹⁴² and Alzheimer’s disease,¹⁴³ as well as cardiac MRI scans,¹⁴⁴ and 3D scans.¹⁴⁵

Yet another approach involves sequentially upsampling low-resolution images in several separate steps using separate models, as introduced by Lai et al.¹⁴⁶ This approach offers two main benefits. First, it allows the user to choose desired resolutions for their high-resolution images without retraining models. Second, it simplifies the learning task for individual networks, since their task is simpler than performing full superresolution in a single feedforward pass. This may potentially improve the performance of the final models.

5.2.2 Specific architectures and methods

Although the generic approaches mentioned above can result in a practically infinite variety of specific architectures, many significant superresolution studies in molecular microscopy have been achieved using a small number of named architectures. Structured feature superresolution microscopy model architecture, shown in panels of Fig. 6 as an example, allows for precise live-cell imaging with high spatial and temporal resolution to continuously monitor subcellular dynamics over extended periods. Among these, ANNA-PALM,⁶⁰ a U-Net based cGAN trained solely with experimental data, stands out; Deep-STORM,⁵⁶ based on a CNN encoder-decoder network trained with simulated data; and smNet,⁵⁷ which directly outputs molecule location, dipole orientation, and wavefront distortion from complex and subtle features of the PSF. In single-molecule superresolution microscopy, there is generally a trade-off between throughput and resolution. To construct a high-quality superresolution image, a large number of molecules need to be localized with high precision, requiring sufficient localizations before sampling a structure of interest.

ANNA-PALM accomplishes this by training on a set of blinking single molecules from which a high-quality superresolution image can be experimentally acquired. A subset of these frames is used to generate a (low-quality) “sparse” superresolution image, which, alongside the diffraction-limited image and information about the imaged structure, serves as inputs into ANNA-PALM. The output, or label, is the full superresolution image reconstructed using all frames. Once trained, ANNA-PALM demonstrated the ability to provide high-quality results in imaging mitochondria, the nuclear core complex, and microtubules⁶⁰ at significantly higher speeds than conventional methods.

ANNA-PALM has proven to be a valuable method for accelerating the acquisition of high-density superresolution images by several orders of magnitude. However, there are many other DL models more appropriate for direct single-molecule localization. An important early model for this is Deep-STORM, developed for the acquisition of superresolution images of microtubules with single or multiple overlapping PSFs. While non-DL algorithms exist for this purpose,¹⁴⁸ they typically suffer from high computational costs and require sample-specific parameter tuning. Deep-STORM, an encoder-decoder CNN trained on simulated images, consists of simulated PSFs in various positions in an image on top of experimentally relevant background levels. Said simulated images are thereafter upsampled by a constant factor, constituting a superresolution image. These two versions of the same image are fed as input and output, respectively, to Deep-STORM during training. Such a model has been used to superresolve images of microtubules

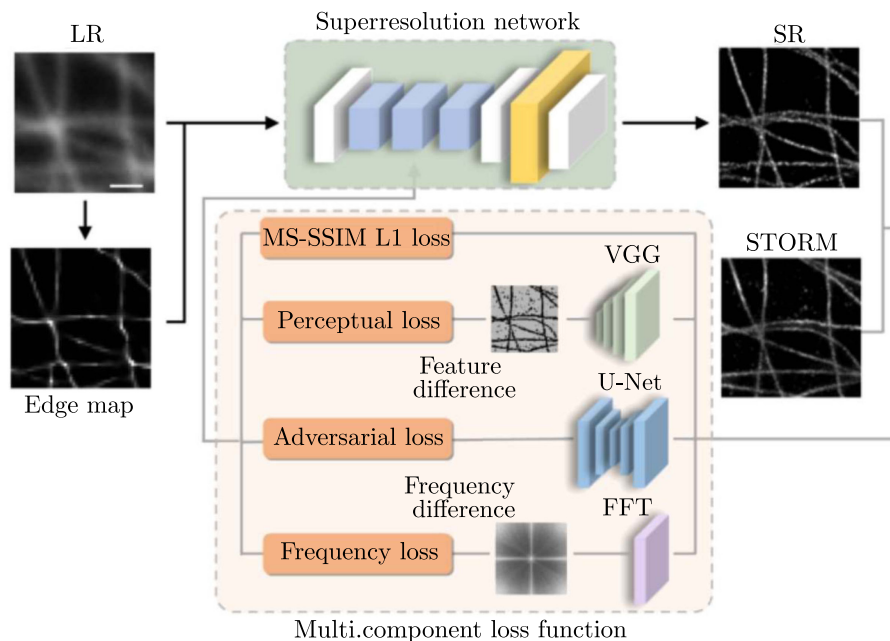


Fig. 6 Superresolution applied architecture. The superresolution network enhances image resolution by training on pairs of simulated low-resolution (LR) and high-resolution ground-truth images or on wide-field (WF) and STORM images from a STORM microscope. First, the LR/WF image undergoes preprocessing through a subpixel edge detector to generate an edge map, both of which serve as inputs to the network. Training is guided by a multi-component loss function that incorporates the combination of multiscale structure similarity index measure and mean absolute error loss (MS-SSIM L1) to capture pixel-level accuracy between the superresolution (SR) and ground-truth/STORM images through multiscale similarity and mean absolute error, perceptual loss to assess feature map differences via the visual geometry group network, adversarial loss using a U-Net discriminator to differentiate ground-truth/STORM images from SR images, and frequency loss to compare differences in the frequency spectrum between SR and ground-truth/STORM images within a specific frequency range using the fast Fourier transform function. This comprehensive loss function helps the superresolution network model achieve precise and perceptually accurate superresolution imaging. Image adapted from Chen et al.¹⁴⁷

and quantum dots,⁵⁶ and inspired works on localizing high-density ultrasound scatterers¹⁴⁹ and Crispr-CAS-protein-DNA binding events.¹⁵⁰

Another impactful model is the aforementioned smNet, which similarly employs a (ResNet-inspired) CNN trained on simulated images for superresolution. The key distinctions lie in the image recreation of 3D images and the network's outputs of the 3D coordinates and orientations of PSF-convoluted emitters from which the superresolution image can be reconstructed. smNet has been demonstrated to localize highly astigmatic single-molecule PSFs in experimental images of significantly higher quality compared to conventional Gaussian fitting methods.⁵⁷ This approach, reminiscent of DeepLoco,¹⁵¹ was developed around the same time and trains NNs to reconstruct simulated emitters in 3D through well-defined mathematical models of astigmatic PSFs. Other related examples of 3D superresolution are that of Zhou et al., who used a dual-GAN framework to directly super-resolve images of mouse brains and bodies taken with fluorescence microscopy,¹⁵² or Zhang et al.¹⁵³ and Zelger et al.,¹⁵⁴ who used a U-Net-based and CNN-based approach, respectively, for superresolution in SMLM.

For images with higher PSF density, Speiser et al. introduced the method known as DECODE.⁵⁸ This architecture consists of

a stack of two U-Nets, where the first U-Net processes a feature representation of a single frame, and the second U-Net processes feature representations of consecutive frames. The output of this method consists of several channels, each containing information in each pixel of the input image regarding (1) the probability of containing an emitter, (2) its brightness, (3) its 3D coordinates, (4) its background intensity, and (5) epistemic uncertainty of its localization and brightness. In the work of Speiser et al.,⁵⁸ this architecture is trained on simulated PSFs with a loss function connected to all five aforementioned types of pixel-level information and has been successfully applied to microtubules in conditions of low light exposure and ultrahigh sample densities.

5.2.3 Superresolution by emitter localization

Further, there are DL methods revolving around improving the precision of localizing underlying emitters in images. This is particularly relevant in SMLM imaging, where spatial resolution of the microscope is in practice directly correlated with the localization precision of single molecules.¹⁵⁵ Since this localization is normally enabled by conventional heuristic-based fitting algorithms, using DL methods may enhance its performance. BGNet⁵⁹ is one such architecture, designed to accurately identify

the centroid of a PSF. It achieves this by training on (simulated) corrupted PSF images and outputting the background of the image. A trained BGNet can then be used to correct the background of a given image at inference time by subtracting its predicted background. Thus, one obtains background-corrected PSF images, which can be fed into conventional maximum likelihood estimation-fitting algorithms for superresolution, thereby enhancing the overall final output without the need to replace the entire analysis pipeline with an end-to-end DL-based system.

5.2.4 Extracting additional information from PSFs

Further, DL methods aim to extract more information from PSFs themselves.^{10,147,156–160} PSFs contain often unexploited information, such as emission wavelengths, as well as lateral and axial location and identity of emitters. Many deep architectures, including smNET, have been developed to exploit this. smNET outputs not only 3D localization of emitters but also their orientation and wavefront distortions. Another notable architecture is DeepSTORM3D,¹⁰ following the original Deep-STORM, which can identify the location of emitters with multiple overlapping PSFs in highly dense conditions over a large axial range by combining information from multiple 3D PSFs. DeepSTORM3D consists of two components: an image formation model and a decoder CNN. The former takes simulated 3D emitter positions as input and outputs their corresponding low-resolution CCD image, while the latter tries to recover the simulated emitter positions given the low-resolution CCD image. The difference between simulated and predicted positions is then used to optimize the phase mask in the Fourier plane and recovery CNN parameters in tandem. This architecture was used by Nehme et al.¹⁰ to superresolve mitochondria and enable the volumetric imaging of telomeres within cells. Further, there are architectures^{157,158} to classify the color channels of individual PSFs. Hershko et al.¹⁵⁸ exploited the chromatic dependence of the PSF to train a CNN architecture to determine the color of an emitter from a gray-scale camera image from a standard fluorescence microscope.

The progress in DL for superresolution has been astounding in the past half-decade, driven mainly by different forms of CNNs trained in GANs through supervised learning. More recently, there are highly promising developments in few-shot,¹⁶¹ single-shot,¹⁶² zero-shot learning¹⁶³ and even untrained neural networks for image superresolution.¹⁶⁴ Diffusion models have also shown significant promise in improving the fidelity and robustness of image superresolution methods.^{165–168} Diffusion models represent a recent approach in generative modeling. These models utilize probabilistic generative techniques in a two-step process: forward diffusion, where noise is iteratively added to images until they become pure Gaussian noise, and reverse diffusion, where images are iteratively denoised using a neural network. By conditioning the reverse diffusion process, diffusion models effectively handle image-to-image transformation tasks.²¹ While these models can produce higher-quality images compared to GANs, they come with significantly higher computational costs.

Recently, they have been used to generate superresolution images of microtubules,¹⁶⁹ reconstruct authentic images with unseen low-axial resolutions into high-axial resolution of 3D microscopic data,¹⁷⁰ and outperform state-of-the-art in high-fidelity continuous image superresolution.¹⁷¹ Thus, these advancements

suggest that the field will continue to progress significantly in the near future.

6 Guidelines

This section provides detailed recommendations for developing cross-modality transformation models in microscopy, with an emphasis on data quality, model architecture selection, and evaluation metrics. Researchers can use this as a framework to navigate the key decisions and challenges associated with their tasks.

6.1 Data Quality, Augmentations, and Data Normalization

The quality of data plays a critical role in determining the performance of DL models. Two major issues commonly affect model quality: insufficient data to capture the variability within the data set or training data that fail to represent the conditions under which the model will be applied.

To detect the issue of insufficient data, a standard approach is to set aside a validation set that the model never sees during training. If the model's performance on this validation set is significantly worse than on the training set, it likely indicates a lack of sufficient training data to generalize effectively. A common practice is to allocate ~20% to 30% of the data set as a validation set. However, it is important to ensure that the validation set is maximally decorrelated from the training set to avoid misleading results. For example, it is ideal to sample from different locations in the sample or even from entirely different experimental videos. A poor sampling strategy, such as selecting every fifth frame from the same video, would introduce a high correlation between the training and validation sets, resulting in overly optimistic performance estimates.

If detected, data scarcity can be mitigated by data augmentation techniques to synthetically increase the diversity of training data. Techniques such as geometric transformations (rotation, scaling), noise injection, and intensity variation can simulate a broader range of conditions. However, care must be taken to ensure that these transformations can be meaningfully applied across modalities. For instance, intensity variations in quantitative phase contrast imaging hold physical significance and altering them synthetically could distort biologically relevant information. Geometric translations, in most cases, provide limited benefit, as convolutional models are inherently translation-equivariant. However, if the chosen model breaks translation equivariance (such as vision transformers), they may be useful.

Regularization techniques also play an essential role in improving model robustness, especially when data are scarce or noisy. Methods, such as dropout, weight decay, or L_2 regularization, are commonly used to prevent overfitting by penalizing overly complex models that may fit noise in the data rather than in underlying patterns. In scenarios where the model could easily memorize the training data, these techniques ensure that the model learns generalizable features rather than artifacts specific to the training set. Advanced regularization techniques, such as Bayesian regularization, can further improve robustness by incorporating uncertainty into the model's predictions, making it especially useful in tasks where noisy or variable data are expected.

Transfer learning offers another potential solution to address limited data availability. Pretrained models, especially those

trained on large data sets from similar domains, can be fine-tuned to perform specific tasks in microscopy. By leveraging features learned from related tasks, transfer learning reduces the need for extensive training data while still allowing the model to generalize effectively. This approach not only speeds up training but also improves the model's performance on smaller, domain-specific data sets. In some cases, transfer learning from pretrained models in related fields, such as medical imaging, can be more effective than starting from scratch, especially in scenarios where biological structures share visual characteristics across different imaging modalities.

When the training data are nonrepresentative, this issue can be identified by observing a drop in model performance under real-world conditions, even though the performance on the validation set remains strong. This discrepancy often arises due to variations in optical systems, sample preparation protocols, or environmental factors that differ from those present in the training data. For instance, subtle differences in microscope settings, sample staining techniques, or even temperature can cause a shift in the data distribution, leading to poor generalization when the model is applied in different scenarios.

The primary strategy to address issues of representativeness is through effective data normalization. Normalization techniques aim to reduce variability in the data by standardizing features across data sets, such as intensity scaling, contrast adjustment, or color normalization. This can help minimize discrepancies between data sets generated under different conditions. However, caution must be taken in modalities where quantitative relationships between intensity values are critical. In such cases, aggressive normalization may disrupt important mappings between intensity and biological features, potentially degrading the model's ability to learn meaningful cross-modality transformations. Furthermore, domain adaptation techniques can be employed to align the distributions of training and application data, improving the robustness of the model across diverse conditions.

6.2 Model Selection

The choice of model architecture can have a significant impact on the performance of the model and depends on several key factors, such as data availability, target task, and specific requirements. Here, we give a general guideline for choosing an appropriate model.

If your data are not aligned, the CycleGAN is recommended. Aligned data refer to cases where each image in one modality has a direct counterpart in the other modality, meaning that both images capture the same sample part of the sample under the same conditions, making it possible to map pixel-to-pixel relationships between the two. When such paired data are unavailable, CycleGAN is suitable because it learns to map between modalities without requiring this strict correspondence. However, the less restrained training procedure also is likely to result in the model learning transformations that are less precise or biologically relevant, especially when precise quantitative relationships between modalities are required. Careful evaluation and additional constraints may be necessary to ensure that the model's outputs are meaningful and accurate.

Assuming your data are paired and aligned, we recommend starting with a conditional GAN architecture, specifically using a U-Net-like generator and a spatial discriminator. A well-

established configuration for this setup is the pix2pix model. Conditional GANs are optimized to generate quantitative, physically meaningful images by leveraging paired data to learn a direct mapping between input and output modalities. The U-Net generator was originally developed for biomedical images and is one of the most proven and widely adopted architectures for tasks involving fine-scale structural details. Spatial discriminators, in turn, evaluate the realism of local regions of the image rather than assessing it as a whole, often resulting in more detailed and accurate outputs.

However, depending on the specific requirements of your task, other architectures may be better suited. For example, if the goal is simply to enhance the contrast of specific substructures without requiring physical realism in the produced images, it may be more practical to forego generative models entirely. In such cases, direct supervised training of a U-Net can offer a simpler and stabler solution. The drawback of using a purely supervised U-Net is that it may lack the ability to generate the nuanced, high-fidelity details that generative models, particularly GANs, are capable of producing. However, for applications where interpretability and stability are more important than photorealism, this trade-off can be worthwhile.

On the other hand, if maximal photorealism is required, diffusion models are worth considering. These models have consistently been shown to produce highly realistic images, often outperforming GANs in terms of image quality and stability. Diffusion models work by iteratively denoising random noise to generate an image, which allows them to better capture fine-grained details and complex textures. However, diffusion models are typically much more computationally expensive, both to train and to evaluate, compared to GANs. Moreover, one should be careful not to conflate photorealism with better quantitative performance on downstream tasks.

Another important consideration is the spatial distribution of information in the image. The U-Net generator is highly effective for analyzing local, position-invariant features, making it ideal for tasks where the meaning of a structure does not depend on its specific location within the image. However, for data where the spatial context is crucial, such as brain scans, attention-based models may be more suitable. Attention mechanisms allow the model to focus on specific regions of the image while considering their global relationships, enabling more context-aware analysis. This makes attention-based architectures a better choice for tasks that require understanding both local features and their larger spatial context.

Finally, for more specialized applications, more complex, hybrid models may be necessary. For tasks where interpretability is a priority, incorporating latent-space constraints can improve both stability and clarity in the results. For example, using a Wasserstein GAN (WGAN) with a carefully designed loss function can provide more control over the training process and generate smoother, more interpretable transformations. In addition, hybrid models that combine multiple architectures, such as variational autoencoders (VAEs) with GANs, can provide both generative flexibility and the ability to impose structural constraints, improving the model's capacity to generate accurate, interpretable results for complex tasks. In superresolution tasks, specialized models such as Deep-STORM or DECODE utilize domain knowledge to far outperform what standard cGANs can achieve.

Table 1 Overview of the key parameters for common approaches of DL in microscopy across scales.

Method	Architecture		Data sets	Learning type	Significant aspect
	Discriminator	Generator			
Tissue					
Conditional GAN (cGAN) ³⁹	CNN	U-Net	Paired labeled/annotated images	Supervised	Conditioning mechanism based on the additional input information
CycleGAN ³⁵	CNN	U-Net	Unpaired histology images	Unsupervised	Cycle-consistency loss enforces consistency and unsupervised translation
StarGAN ⁴⁰	PatchGAN	U-Net/ResNet	Unpaired images of tissue structures from multiple domains	Unsupervised	Unified architecture for a single model across multiple domains
Cellular and subcellular structures					
Conditional GAN (cGAN) ⁵⁴	CNN	U-Net	Paired images from fluorescence, confocal electron microscopy	Supervised	Conditioning mechanism based on the additional input information
CycleGAN ⁵⁵	CNN	U-Net	Unpaired images of bright-field, phase-contrast, fluorescence, and DIC microscopy	Unsupervised	Cycle-consistency loss enforces consistency and unsupervised translation
Molecular structures					
Deep-STORM ⁵⁶	Encoder–decoder CNN		Fluorescent images from techniques like STORM, PALM, or dSTORM	Supervised, labeled	Trained on simulated data to enhance resolution in SMLM, enabling superresolution imaging of molecular structures with improved accuracy
smNet ⁵⁷	ResNet-inspired CNN		Fluorescent images from techniques such as STORM, PALM, or SIM	Supervised, labeled	Simulated PSFs and ground-truth 3D position labels training, accurately localized astigmatic single-molecule PSFs
DECODE ⁵⁸	Stack of two U-Nets		Images captured via optical diffraction tomography	Supervised, labeled	Stacked U-Net to process single and consecutive frames, improved accuracy, and resolution under low-light conditions
DeepSTORM3D ¹⁰	Encoder–decoder CNN		Simulated images of fluorescent emitters noise, and optical properties of the microscope with known positions, including PSFs, background	Supervised, labeled	Image formation model and decoder CNN to pinpoint 3D emitter coordinates from simulated PSFs, enabling high-resolution volumetric molecular imaging
BGNet ⁵⁹	CNN		Fluorescent images from fluorescence microscopy or SMLM	Supervised, labeled	Identifying PSF centroids for background correction, improving single-molecule localization and overall imaging resolution
ANNA-PALM ⁶⁰	U-Net-based cGAN		Images of photoactivated single molecules captured by PALM or STORM	Supervised, labeled	Trained on experimental data to rapidly acquire high-density superresolution images, especially for mitochondria, nuclear core complexes, and microtubules

6.3 Evaluation Metrics

Evaluating the performance of a cross-modality transformation model can be challenging. Typical strategies involve measuring the visual fidelity of the images, but these measures may not fully correlate with the retention of biologically relevant information. Some examples of evaluation metrics include the following.

- **SSIM and PSNR:** These metrics are simple to calculate and provide an overall assessment of image quality. However, they may not capture fine-grained differences in features critical to biological research.
- **MS-SSIM:** Multiscale SSIM evaluates image quality at different scales, which may make it more suitable for capturing hierarchical structures.
- **NIQE:** Natural image quality evaluator is a no-reference metric that measures perceptual quality, making it useful when ground truth is not available. However, the metric is optimized for natural images, and its applicability to microscopy images should be questioned.
- **Frequency domain analysis:** For specific tasks, such as superresolution, analyzing the frequency components of an image can give insights into how well high-frequency details are preserved.

Another approach is to evaluate the biological relevance of the generated images by performing downstream analyses, such as cell counting or feature segmentation, and comparing the results to known quantities or those obtained from real experimental data. This method more directly assesses the retention of biologically meaningful information, but it introduces additional uncertainties. For example, inaccuracies in the downstream task, such as errors in the cell-counting algorithm, can confound the evaluation of the model's performance, making it difficult to disentangle the model's contributions from errors in postprocessing or analysis pipelines.

6.4 Ethical Considerations

Ethical considerations are essential for ensuring responsible and fair use of AI in the image analysis of biological samples. A key concern is protecting the privacy of patients and donors, as these samples often contain sensitive personal information. Handling biological samples must comply with data protection laws, which require informed consent from all parties involved and transparency about how the samples will be used. These laws vary by region, with the General Data Protection Regulation (GDPR) in the European Union,¹⁷² the Health Insurance Portability and Accountability Act (HIPAA) in the United States,¹⁷³ the Data Protection Act 2018 in the United Kingdom,¹⁷⁴ and the Personal Information Protection Law (PIPL) in China.¹⁷⁵ International efforts, such as those led by the World Health Organization,^{176,177} along with national initiatives, such as India's data protection frameworks,¹⁷⁸ continue to evolve these regulations to keep pace with the rapid growth of AI technologies.^{179–181} In addition, a growing body of literature addresses these developments across various countries,¹⁸² stages of database handling,¹⁸³ and specific fields.¹⁸⁴

These regulations address issues such as privacy breaches due to improper data use and cybersecurity threats.¹⁸⁵ Furthermore, the responsibility for AI models used in diagnostics is a major concern, particularly in relation to bias mitigation. Biased data sets can result in inaccurate or discriminatory

outcomes, especially in healthcare applications. Rigorous validation of AI models is critical to ensure accuracy, reproducibility, and the prevention of errors that may lead to misdiagnosis or flawed scientific conclusions. Transparency is also crucial, requiring clear documentation of model training, data sources, usage frameworks, and decision-making processes.¹⁸⁶ Guidelines should promote not only sharing data sets but also the trained model weights, enabling researchers to independently validate and replicate findings. Lastly, accountability frameworks are necessary to ensure that researchers and developers are held responsible for the ethical use of AI, with proper oversight to enforce compliance with established guidelines.

7 Perspectives

Cross-modality transformations in biological microscopy present advanced techniques with important implications for biology, medicine, and materials science. Although these advances suggest exciting opportunities where AI is set to increase diagnostic accuracy and improve workflow efficiency, it still faces ongoing challenges that require innovative solutions for them to be implicated and come to societal use. Figure 7 summarizes the ongoing developments and potential outcomes of combining AI with novel imaging modalities.

For example, the synthesis of high-resolution images from less invasive imaging methods such as MRI and CT scans provides tissue insights without the need for biopsies,^{37,49} illustrated in Fig. 7(a). This approach provides another important advantage: access to living tissue data that can have a significant impact on clinical studies. Similar strategies may also advance cellular *in vitro* culture studies in preclinical settings. Such strategies may advance preclinical *in vitro* cell culture studies in creating physically relevant environments with 3D settings, such as promoting cell growth into spheroids or inoculating them on a microphysiology platform.¹⁹⁰ Recent advances of this technology have enabled co-culture of single or multiple cell types that mimic closer *in vivo* conditions by implementing 3D architectures, fluid dynamics, and the gradient of materials contained in living tissues. In these environments, extracting probe-free information on cellular behavior, metabolic states, or migration is currently not feasible, but may soon be achievable through AI and various imaging modalities, as presented in the bottom right panel.

It is also likely that new imaging technologies and implementations with AI will be able to integrate data across scales to provide unprecedented perspective on diseases at the cellular and molecular levels, portrayed in Fig. 7(b). This approach would produce realistic models that combine visual, molecular, and genomic information. By implementing data from a variety of modalities, AI will enable a more comprehensive analysis of biological models including the practical information required for the diagnosis of complex diseases where tissue morphology and function are crucial, exemplified in Fig. 7(d). In addition, longitudinal disease monitoring could be more sensitive, allowing clinicians to track tissue changes over time, shown in Fig. 7(c), and tailor treatment responses accordingly. This possibility extends to the field of transplantation biology, where the cellular integration, biocompatibility, and biodegradation of transplanted tissues, synthetic materials, or prostheses can be monitored over time. In summary, AI-powered tools will enable faster and more accurate diagnostics with reduced bias, thus minimizing the time required for human review. Over time, these advancements have the potential to make high-quality histological analysis more accessible, particularly in areas with

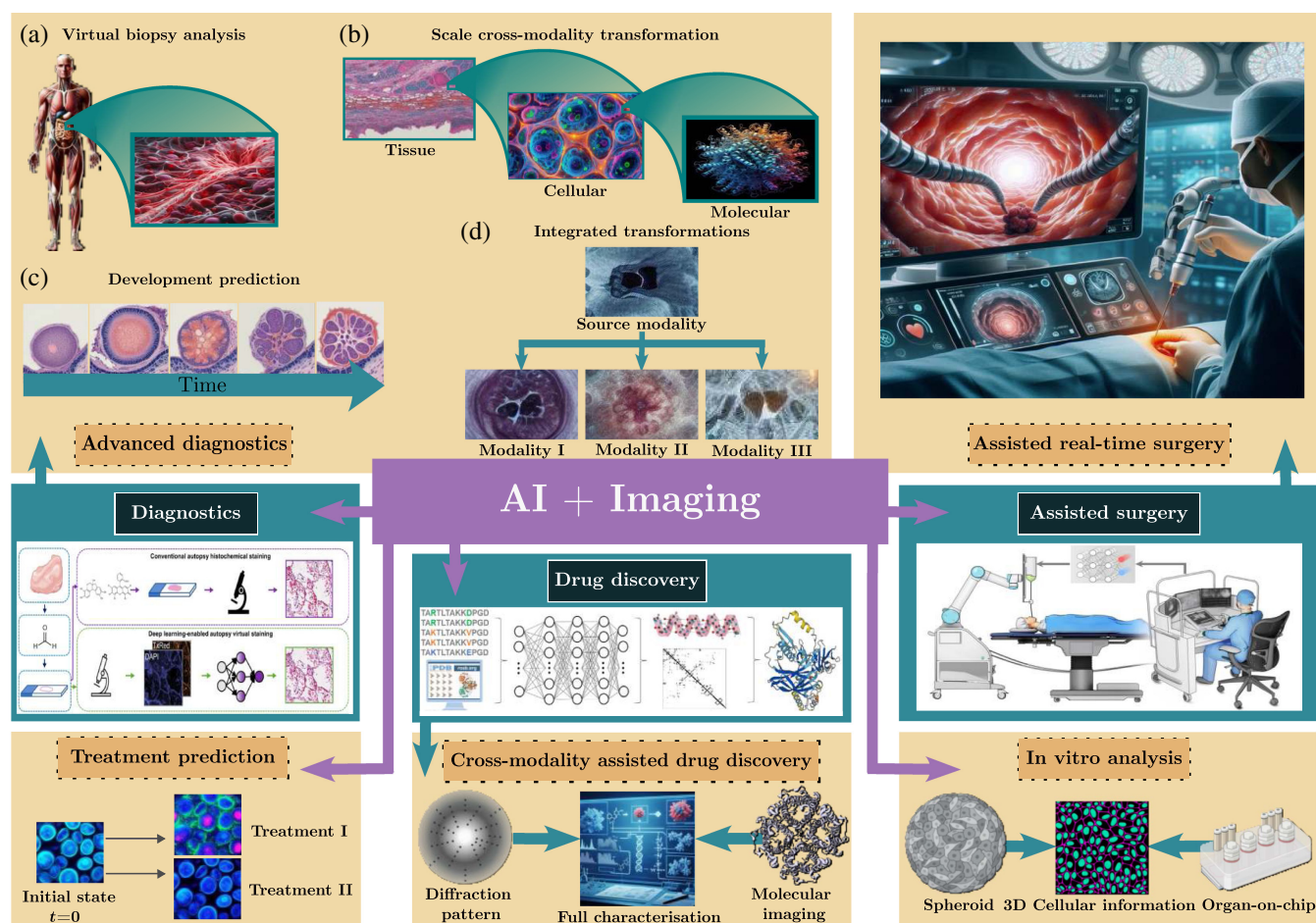


Fig. 7 Potential application perspectives of AI on biological samples imaging. Current developments found in the literature are contained in green boxes, while speculative prospects for the future are contained in yellow boxes. Starting from the top left, AI is extensively used in diagnostics and other cross-modality transformations (image in the green panel adapted from Li et al.¹⁸⁷). (a) In the future, this could lead to *in vivo* analysis of virtual biopsies instead of performing tissue extraction and preparation. (b) Cross-modality transform could progressively transition the limits of the different scales outlined in this review. (c) As a consequence of broader data availability and *in vivo* imaging, AI models could also transform and predict sample evolution over time. (d) Further, cross-modality transformations could also become more accessible and be routinely integrated in the sampling process, obtaining simultaneously different modalities with one single measurement. In the bottom left panel, these ideas could be integrated in treatment prediction and design for biological samples, leading to personally tailored therapy. The potential of superresolution to better characterize molecules structures is nowadays used for protein-folding determination (image in the green panel adapted from Kumar et al.¹⁸⁸), but implementing other information input from different modalities could enhance the reach of our knowledge (centered, bottom panel). The application of AI extracting cellular information from 3D structures hosted in increasingly complex *in vitro* systems that better replicate the dynamic conditions of *in vivo* systems (bottom right panel), is highlighted. AI is already implemented in assisted-surgery settings and equipment (right green panel adapted from Zhang et al.¹⁸⁹), which could be greatly improved by including real-time contrast enhancement and segmentation from cross-modality transformations. Images in the yellow boxes were created with the assistance of Designer, using DALL-E 3 technology, and BioRender. They have demonstrative purposes and do not hold real scientific meaning beyond the visualization of the ideas expressed.

limited pathology expertise, while also standardizing diagnostic protocols across institutions.

Beyond diagnostics, AI is already being used to guide physicians during advanced surgeries, directing the surgeon's

movements¹⁸⁹ with accuracy and precision. Future imaging cross-modality transformations may facilitate real-time tissue mapping during surgery, giving surgeons immediate insights from various imaging techniques, represented in the top right

panels. AI-powered methods are also being used in the research for drug discovery to identify new drug candidates and their potential folding structures,¹⁸⁸ as demonstrated in the bottom central panels. Subsequent preclinical studies to predict how tissues will respond to novel treatments could be envisioned using cross-modality transformations, as seen in the bottom left panel. In the future, AI may simulate the effects of drugs on a patient's tissue, aiding in the development of personalized therapies. However, challenges remain, including the need for high-quality multimodal data sets to train AI systems and the development of interpretable AI models that biologists and clinicians can trust. In addition, integrating AI into clinical workflows requires careful consideration to ensure these new technologies are used effectively and ethically by healthcare professionals. Despite these hurdles, the future of AI in cross-modality transformations in biology is promising, with profound implications for both biomedical research and clinical diagnostics.

8 Conclusions

The incorporation of DL techniques in biological microscopy represents a significant advancement, with the potential to enhance our understanding of histology, cellular structures, and molecular imaging. While these technologies offer promise, it is essential to acknowledge that the field is still evolving. The current state of these methods often involves grappling with their black-box nature, necessitating further refinement and investigation. Researchers continue to address challenges related to interpretability and the need for extensive developments to unlock the full transformative potential of DL in biological microscopy. Beyond technological advancement, these methods offer a paradigm shift by enabling imaging without the reliance on chemical stains and fluorescence, both simplifying experimental processes and preserving sample integrity. This marks a pivotal shift in microscopy, offering a noninvasive and label-free alternative that preserves the integrity of the specimens under investigation. Cross-modality transformations have a significant impact not only in laboratory settings but also in clinical diagnostics and fundamental biological research, opening new avenues for discoveries and breakthroughs. Furthermore, these techniques are becoming more accessible and affordable, democratizing access to microscopic exploration and empowering researchers across disciplines.

Code and Data Availability

This review article did not generate any original data or code. All data discussed are available in the cited publications.

Acknowledgments

Jesus Manuel Antunez, Caroline Beck Adiels, and Giovanni Volpe acknowledge support from the MSCA-ITN-ETN project ActiveMatter sponsored by the European Commission (Horizon 2020, Project No. 812780). Giovanni Volpe acknowledges support from the ERC-CoG project MAPEI sponsored by the European Commission (Horizon 2020, Project No. 101001267) and from the Knut and Alice Wallenberg Foundation (Grant No. 2019.0079). Finally, Caroline Beck Adiels and Giovanni Volpe acknowledge the Swedish Foundation for Strategic Research (Grant No. ITM17-0384).

References

1. D. Murphy and M. Davidson, *Fundamentals of Light Microscopy and Electronic Imaging*, John Wiley & Sons, Ltd., Hoboken, New Jersey (2012).
2. K. Suvana, C. Layton, and J. Bancroft, *Bancroft's Theory and Practice of Histological Techniques E-Book*, Elsevier Health Sciences, London, England (2012).
3. J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nat. Methods* **2**(12), 910–919 (2005).
4. I. Johnson, "Molecular probes handbook: a guide to fluorescent probes and labeling technologies," Life Technologies Corporation, Carlsbad, California (2010).
5. H. Sahoo, "Fluorescent labeling techniques in biomolecules: a flashback," *RSC Adv.* **2**(18), 7017–7029 (2012).
6. S. Shashkova and M. Leake, "Single-molecule fluorescence microscopy review: shedding new light on old problems," *Biosci. Rep.* **37**, BSR20170031 (2017).
7. M. Gurcan et al., "Histopathological image analysis: a review," *IEEE Rev. Biomed. Eng.* **2**, 147–171 (2009).
8. F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nat. Methods* **2**(12), 932–940 (2005).
9. Y.-F. He et al., "Deep-learning driven, high-precision plasmonic scattering interferometry for single-particle identification," *ACS Nano* **18**(13), 9704–9712 (2024).
10. E. Nehme et al., "DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning," *Nat. Methods* **17**(7), 734–740 (2020).
11. J. Ferreira and L. Groc, *Surface Glutamate Receptor Nanoscale Organization with Super-Resolution Microscopy (dSTORM)*, pp. 35–52, Springer US, New York, New York (2024).
12. M. D. Wilkinson et al., "The fair guiding principles for scientific data management and stewardship," *Sci. Data* **3**(1), 160018 (2016).
13. Y. Rivenson et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nat. Biomed. Eng.* **3**(6), 466–477 (2019).
14. S. Shafi and A. V. Parwani, "Artificial intelligence in diagnostic pathology," *Diagn. Pathol.* **18**(1), 109 (2023).
15. E. M. Christiansen et al., "In silico labeling: predicting fluorescent labels in unlabeled images," *Cell* **173**(3), 792–803 (2018).
16. H. Wang et al., "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods* **16**(1), 103–110 (2019).
17. X. Cao et al., "Deep learning based inter-modality image registration supervised by intra-modality similarity," in *Lecture Notes in Computer Science*, pp. 55–63, Springer International Publishing, Cham, Switzerland (2018).
18. L. Latonen et al., "Virtual staining for histology by deep learning," *Trends Biotechnol.* **42**(9), 1177–1191 (2024).
19. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Lecture Notes in Computer Science*, pp. 694–711, Springer International Publishing, Cham, Switzerland (2016).
20. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, pp. 15640–15649, IEEE (2017).
21. K. Zhang et al., "Negative-aware attention framework for image-text matching," in *IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, IEEE (2022).
22. T. S. Gurina and L. Simms, "Histology, staining," StatPearls Publishing, Treasure Island, Florida (2023).
23. C. L. Chen et al., "Deep learning in label-free cell classification," *Sci. Rep.* **6**, 21471 (2016).
24. M. Hermsen et al., "Deep learning-based histopathologic assessment of kidney tissue," *J. Am. Soc. Nephrol.* **30**(10), 1968–1979 (2019).

25. B. Bai et al., "Label-free virtual HER2 immunohistochemical staining of breast tissue using deep learning," *BME Frontiers* **2022**, 9786242 (2022).
26. S. Soltani et al., "Prostate cancer histopathology using label-free multispectral deep-UV microscopy quantifies phenotypes of tumor aggressiveness and enables multiple diagnostic virtual stains," *Sci. Rep.* **12**(1), 9329 (2022).
27. N. Pillar and A. Ozcan, "Virtual tissue staining in pathology using machine learning," *Expert Rev. Molecular Diagnostics* **22**(11), 987–989 (2022).
28. D. S. Richardson and J. W. Lichtman, "Clarifying tissue clearing," *Cell* **162**(2), 246–257 (2015).
29. X. Wang et al., "Single-shot isotropic differential interference contrast microscopy," *Nat. Commun.* **14**, 2063 (2023).
30. F. Pan et al., "Accurate detection and instance segmentation of unstained living adherent cells in differential interference contrast images," *Comput. Biol. Med.* **182**, 109151 (2024).
31. M. Z. Hoque et al., "Stain normalization methods for histopathology image analysis: a comprehensive review and experimental comparison," *Inf. Fusion* **102**, 101997 (2024).
32. A. Tomczak et al., "Multi-task multi-domain learning for digital staining and classification of leukocytes," *IEEE Trans. Med. Imaging* **40**(10), 2897–2910 (2021).
33. J. Park et al., "Artificial intelligence-enabled quantitative phase imaging methods for life sciences," *Nat. Methods* **20**(11), 1645–1660 (2023).
34. E. Breznik et al., "Cross-modality sub-image retrieval using contrastive multimodal image representations," *Sci. Rep.* **14**, 18798 (2024).
35. A. Lahiani et al., "Virtualization of tissue staining in digital pathology using an unsupervised deep learning approach," in *Lecture Notes in Comput. Sci.*, pp. 47–55, Springer International Publishing, Warwick, England (2019).
36. G. Zhang et al., "Image-to-images translation for multiple virtual histological staining of unlabeled human carotid atherosclerotic tissue," *Molecular Imaging Biol.* **24**(1), 31–41 (2022).
37. J. Li et al., "Biopsy-free *in vivo* virtual histology of skin using deep learning," *Light Sci. Appl.* **10**(1), 233 (2021).
38. T. M. Abraham et al., "Label- and slide-free tissue histology using 3D epi-mode quantitative phase imaging and virtual hematoxylin and eosin staining," *Optica* **10**(12), 1605–1618 (2023).
39. A. Rana et al., "Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis," *JAMA Network Open* **3**(5), e205111 (2020).
40. D. Li et al., "Deep learning for virtual histological staining of bright-field microscopic images of unlabeled carotid artery tissue," *Molecular Imaging Biol.* **22**(5), 1301–1309 (2020).
41. E. A. Burlingame et al., "Shift: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning," *Sci. Rep.* **10**(1), 17507 (2020).
42. B. Shen et al., "Deep learning autofluorescence-harmonic microscopy," *Light Sci. Appl.* **11**(1), 76 (2022).
43. Y. Zhang et al., "Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue," *Light Sci. Appl.* **9**(1), 78 (2020).
44. K. de Haan et al., "Deep learning-based transformation of h&e stained tissues into special stains," *Nat. Commun.* **12**(1), 4884 (2021).
45. Y. Rivenson et al., "Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light Sci. Appl.* **8**(1), 23 (2019).
46. M. Boktor et al., "Virtual histological staining of label-free total absorption photoacoustic remote sensing (TA-PARS)," *Sci. Rep.* **12**(1), 10296 (2022).
47. J. J. Levy et al., "A large-scale internal validation study of unsupervised virtual trichrome staining technologies on non-alcoholic steatohepatitis liver biopsies," *Mod. Pathol.* **34**(4), 808–822 (2021).
48. Y. Hong et al., "Deep learning-based virtual cytokeratin staining of gastric carcinomas to measure tumor–stroma ratio," *Sci. Rep.* **11**(1), 19255 (2021).
49. Q. Dou et al., "PNP-Adanet: plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation," *IEEE Access* **7**, 99065–99076 (2019).
50. J. Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2242–2251, Institute of Electrical and Electronics Engineers Inc. (2017).
51. M. Kawai et al., "Virtual multi-staining in a single-section view for renal pathology using generative adversarial networks," *Comput. Biol. Med.* **182**, 109149 (2024).
52. Y. Rivenson et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nat. Biomed. Eng.* **3**(6), 466–477 (2019).
53. S. Koivukoski et al., "Unstained tissue imaging and virtual hematoxylin and eosin staining of histologic whole slide images," *Lab. Invest.* **103**(5), 100070 (2023).
54. S. Helgadottir et al., "Extracting quantitative biological information from bright-field cell images using deep learning," *Biophys. Rev.* **2**(3), 031401 (2021).
55. X. Li et al., "Unsupervised content-preserving transformation for optical microscopy," *Light Sci. Appl.* **10**(1), 44 (2021).
56. E. Nehme et al., "Deep-storm: super-resolution single-molecule microscopy by deep learning," *Optica* **5**(4), 458–464 (2018).
57. P. Zhang et al., "Analyzing complex single-molecule emission patterns with deep learning," *Nat. Methods* **15**(11), 913–916 (2018).
58. A. Speiser et al., "Deep learning enables fast and dense single-molecule localization with high accuracy," *Nat. Methods* **18**(9), 1082–1090 (2021).
59. L. Möckl et al., "Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network BGnet," *Proc. Natl. Acad. Sci.* **117**(1), 60–67 (2019).
60. W. Ouyang et al., "Deep learning massively accelerates super-resolution localization microscopy," *Nat. Biotechnol.* **36**(5), 460–468 (2018).
61. R. Sanyal, D. Kar, and R. Sarkar, "Carcinoma type classification from high-resolution breast microscopy images using a hybrid ensemble of deep convolutional features and gradient boosting trees classifiers," *IEEE/ACM Trans. Comput. Biol. Bioinf.* **19**, 2124–2136 (2021).
62. D. Zhang et al., "Cross-modality deep feature learning for brain tumor segmentation," *Pattern Recognit.* **110**, 107562 (2021).
63. Z. Wang et al., "Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning," *Nat. Methods* **18**(5), 551–556 (2021).
64. M. S. Durkee et al., "Artificial intelligence and cellular segmentation in tissue microscopy images," *Am. J. Pathol.* **191**(10), 1693–1701 (2021).
65. Q. Dou et al., "Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss," in *Int. Joint Conf. Artif. Intell.*, pp. 691–697 (2018).
66. Q. Li et al., "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Trans. Med. Imaging* **35**(1), 109–118 (2016).
67. S. Klein et al., "Deep learning predicts HPV association in oropharyngeal squamous cell carcinomas and identifies patients with a favorable prognosis using regular H&E stains," *Clin. Cancer Res.* **27**(4), 1131–1138 (2021).
68. W. Xie et al., "Prostate cancer risk stratification via nondestructive 3D pathology with deep learning–assisted gland analysis," *Cancer Res.* **82**(2), 334–345 (2022).
69. H. Lee et al., "Revisiting the use of structural similarity index in Hi-C," *Nat. Genet.* **55**(12), 2049–2052 (2023).

70. D. Gu et al., "An artificial-intelligence-based age-specific template construction framework for brain structural analysis using magnetic resonance images," *Hum. Brain Mapp.* **44**(3), 861–875 (2023).
71. Z. Yu et al., "Need for objective task-based evaluation of deep learning-based denoising methods: a study in the context of myocardial perfusion spect," *Med. Phys.* **50**(7), 4122–4137 (2023).
72. M. Dohmen et al., "Similarity metrics for MR image-to-image translation," arXiv: 2405.08431 (2024).
73. J. Ke et al., "Artifact detection and restoration in histology images with stain-style and structural preservation," *IEEE Trans. Med. Imaging* **42**(12), 3487–3500 (2023).
74. R. Houhou et al., "Comparison of denoising tools for the reconstruction of nonlinear multimodal images," *Biomed. Opt. Express* **14**(7), 3259–3278 (2023).
75. G. Litjens et al., "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Sci. Rep.* **6**, 26286 (2016).
76. M. Luella, B. Paul, and A. Javad, "Generative AI in medical imaging and its application in low dose computed tomography (CT) image denoising," in *Applications of Generative AI*, pp. 387–401, Springer International Publishing (2024).
77. Y. Choi et al., "StarGAN: unified generative adversarial networks for multi-domain image-to-image translation," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 8789–8797 (2018).
78. J. Vasiljević et al., "HistostarGAN: a unified approach to stain normalisation, stain transfer and stain invariant segmentation in renal histopathology," *Knowl.-Based Syst.* **277**, 110780 (2023).
79. C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: a survey," *Med. Image Anal.* **67**, 101813 (2021).
80. J. van der Laak, G. Litjens, and F. Ciompi, "Deep learning in histopathology: the path to the clinic," *Nat. Med.* **27**(5), 775–784 (2021).
81. A. Echle et al., "Deep learning in cancer pathology: a new generation of clinical biomarkers," *Br. J. Cancer* **124**, 686–696 (2020).
82. S. Banerji and S. Mitra, "Deep learning in histopathology: a review," *WIREs Data Min. Knowl. Discovery* **12**, e1439 (2022).
83. J. Xu et al., *Deep Learning for Histopathological Image Analysis: Towards Computerized Diagnosis on Cancers*, pp. 73–95, Springer International Publishing (2017).
84. Y. He et al., "PST-Diff: achieving high-consistency stain transfer by diffusion models with pathological and structural constraints," *IEEE Trans. Med. Imaging* **43**(10), 3634–3647 (2024).
85. M. J. Fanous and G. Popescu, "GANscan: continuous scanning microscopy using deep learning deblurring," *Light Sci. Appl.* **11**(1), 265 (2022).
86. Y. Rivenson and A. Ozcan, "Deep learning accelerates whole slide imaging for next-generation digital pathology applications," *Light Sci. Appl.* **11**(1), 300 (2022).
87. A. Su et al., "A deep learning model for molecular label transfer that enables cancer cell identification from histopathology images," *NPJ Precis. Oncol.* **6**(1), 14 (2022).
88. S. M. Hickey et al., "Fluorescence microscopy: an outline of hardware, biological handling, and fluorophore considerations," *Cells* **11**, 35 (2022).
89. P. A. Moghadam et al., "A morphology focused diffusion probabilistic model for synthesis of histopathology images," in *IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, IEEE Computer Society, Los Alamitos, California, pp. 1999–2008 (2023).
90. Y. Shen and J. Ke, "StainDiff: transfer stain styles of histology images with denoising diffusion probabilistic models and self-ensemble," in *Med. Image Computing and Computer Assisted Intervention*, A. Greenspan et al., Eds., pp. 549–559, Springer International Publishing, Cham, Switzerland (2023).
91. T. Kataria, B. Knudsen, and S. Y. Elhabian, "StainDiffuser: multitask dual diffusion model for virtual staining," arXiv:2403.11340 (2024).
92. S. Dubey et al., "VIMS: virtual immunohistochemistry multiplex staining via text-to-stain diffusion trained on uniplex stains," in *Mach. Learn. Med. Imaging*, X. Xu et al., Eds., pp. 143–155, Springer International Publishing, Cham, Switzerland (2024).
93. T. M. Abraham and R. Levenson, "A comparison of diffusion models and CycleGANs for virtual staining of slide-free microscopy images," in *ICML 2023 Workshop Deployable Generative AI*, pp. 1–6 (2023).
94. R. Rizzuto et al., "Chimeric green fluorescent protein as a tool for visualizing subcellular organelles in living cells," *Curr. Biol.* **5**(6), 635–642 (1995).
95. O. Kepp et al., "Cell death assays for drug discovery," *Nat. Rev. Drug Disc.* **10**(3), 221–237 (2011).
96. E. Moen et al., "Deep learning for cellular image analysis," *Nat. Methods* **16**(12), 1233–1246 (2019).
97. V. Lulevich et al., "Cell tracing dyes significantly change single cell mechanics," *J. Phys. Chem. B* **113**(18), 6511–6519 (2009).
98. A. J. Hobro and N. I. Smith, "An evaluation of fixation methods: spatial and compositional cellular changes observed by Raman imaging," *Vib. Spectrosc.* **91**, 31–45 (2017).
99. J. Hira et al., "From differential stains to next generation physiology: chemical probes to visualize bacterial cell structure and physiology," *Molecules* **25**(21), 4949 (2020).
100. E. C. Jensen, "Overview of live-cell imaging: requirements and methods used," *Anat. Rec.* **296**(1), 1–8 (2013).
101. S. N. Chandrasekaran et al., "Image-based profiling for drug discovery: due for a machine-learning upgrade?" *Nat. Rev. Drug Disc.* **20**(2), 145–159 (2021).
102. T. C. Nguyen et al., "Virtual organelle self-coding for fluorescence imaging via adversarial learning," *J. Biomed. Opt.* **25**(9), 096009 (2020).
103. M. E. Kandel et al., "Multiscale assay of unlabeled neurite dynamics using phase imaging with computational specificity," *ACS Sens.* **6**(5), 1864–1874 (2021).
104. Y. N. Nygate et al., "Holographic virtual staining of individual biological cells," *Proc. Natl. Acad. Sci.* **117**(17), 9223–9231 (2020).
105. C. Ounkomol et al., "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nat. Methods* **15**(11), 917–920 (2018).
106. S. Cheng et al., "Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy," *Sci. Adv.* **7**(3), eabe0431 (2021).
107. W. Song et al., "Led array reflectance microscopy for scattering-based multi-contrast imaging," *Opt. Lett.* **45**(7), 1647–1650 (2020).
108. C. L. Cooke et al., "Physics-enhanced machine learning for virtual fluorescence microscopy," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, pp. 3803–3813 (2021).
109. V. Mannam et al., "Deep learning-based super-resolution fluorescence microscopy on small datasets," *Proc. SPIE* **11650**, 116500O (2021).
110. L. Xu et al., "Deep learning enables stochastic optical reconstruction microscopy-like superresolution image reconstruction from conventional microscopy," *iScience* **26**(11), 108145 (2023).
111. F. Zhao et al., "Deep-learning super-resolution light-sheet add-on microscopy (Deep-SLAM) for easy isotropic volumetric imaging of large biological specimens," *Opt. Express* **11**(12), 7273–7285 (2020).
112. R. J. G. van Sloun et al., "Super-resolution ultrasound localization microscopy through deep learning," *IEEE Trans. Med. Imaging* **40**(3), 829–839 (2021).
113. X. Zhuang, "Nano-imaging with storm," *Nat. Photonics* **3**(7), 365–367 (2009).
114. B. Huang, M. Bates, and X. Zhuang, "Super-resolution fluorescence microscopy," *Annu. Rev. Biochem.* **78**(1), 993–1016 (2009).

115. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**(10), 793–796 (2006).
116. H. Shroff, H. White, and E. Betzig, "Photoactivated localization microscopy (PALM) of adhesion complexes," *Curr. Protoc. Cell Biol.* **41**, 4.21.1–4.21.27 (2013).
117. S. T. Hess, T. P. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.* **91**(11), 4258–4272 (2006).
118. M. Jung, D. Kim, and J. Y. Mun, "Direct visualization of actin filaments and actin-binding proteins in neuronal cells," *Front. Cell Dev. Biol.* **8**, 588556 (2020).
119. E. Wolf, *Progress in Optics*, Elsevier Science & Technology (2008).
120. M. Lelek et al., "Single-molecule localization microscopy," *Nat. Rev. Methods Primers* **1**(1), 39 (2021).
121. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 105–114 (2016).
122. J. Grant-Jacob et al., "A neural lens for super-resolution biological imaging," *J. Phys. Commun.* **3**, 7 (2019).
123. C. Chen et al., "Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation," in *Proc. AAAI Conf. Artif. Intell.*, Vol. 33, pp. 865–872 (2019).
124. D. Maleki and H. Tizhoosh, "LILE: look in-depth before looking elsewhere: a dual attention network using transformers for cross-modal information retrieval in histopathology archives," arXiv:2203.01445 (2022).
125. Q. Yang et al., "MRI cross-modality image-to-image translation," *Sci. Rep.* **10**(1), 3753 (2020).
126. R. Naseem et al., "Cross modality guided liver image enhancement of CT using MRI," in *8th Eur. Workshop Visual Inf. Process.*, pp. 46–51 (2019).
127. C. Dong et al., "Learning a deep convolutional network for image super-resolution," in *Computer Vision*, D. Fleet et al., Eds., pp. 184–199, Springer International Publishing, Cham, Switzerland (2014).
128. Y. Zheng et al., "A hybrid convolutional neural network for super-resolution reconstruction of MR images," *Med. Phys.* **47**(7), 3013–3022 (2020).
129. J. Chun et al., "MRI super-resolution reconstruction for MRI-guided adaptive radiotherapy using cascaded deep learning: in the presence of limited training data and unknown translation model," *Med. Phys.* **46**(9), 4148–4164 (2019).
130. J. Lv et al., "Reconstruction of undersampled radial free-breathing 3D abdominal MRI using stacked convolutional auto-encoders," *Med. Phys.* **45**(5), 2023–2032 (2018).
131. H. Li et al., "Fast and accurate super-resolution of MR images based on lightweight generative adversarial network," *Multimedia Tools Appl.* **82**(2), 2465–2487 (2023).
132. L. Kang et al., "Super-resolution method for MR images based on multi-resolution CNN," *Biomed. Signal Process. Control* **72**, 103372 (2022).
133. E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.* **44**(10), e360–e375 (2017).
134. C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision—ECCV 2016*, J. M. Leibe et al., Eds., pp. 391–407, Springer International Publishing, Cham, Switzerland (2016).
135. Y.-B. Du et al., "X-ray image super-resolution reconstruction based on a multiple distillation feedback network," *Appl. Intell.* **51**(7), 5081–5094 (2021).
136. Y. Almalioglu et al., "EndoL2H: deep super-resolution for capsule endoscopy," *IEEE Trans. Med. Imaging* **39**(12), 4297–4309 (2020).
137. D. Ravi et al., "Effective deep learning training for single-image super-resolution in endomicroscopy exploiting video-registration-based reconstruction," *Int. J. Comput. Assisted Radiol. Surgery* **13**(6), 917–924 (2018).
138. M. Srikrishna et al., "Deep learning from MRI-derived labels enables automatic brain tissue classification on human brain CT," *NeuroImage* **244**, 118606 (2021).
139. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 1664–1673 (2018).
140. O. Oktay et al., "Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imaging* **37**(2), 384–395 (2018).
141. M. Srikrishna et al., "Comparison of two-dimensional- and three-dimensional-based U-net architectures for brain tissue classification in one-dimensional brain CT," *Front. Comput. Neurosci.* **15**, 785244 (2022).
142. A. Mani et al., "Applying deep learning to accelerated clinical brain magnetic resonance imaging for multiple sclerosis," *Front. Neurol.* **12**, 685276 (2021).
143. K. Chui et al., "An MRI scans-based Alzheimer's disease detection via convolutional neural network and transfer learning," *Diagnostics* **12**(7), 1531 (2022).
144. J.-Y. Lin, Y.-C. Chang, and W. H. Hsu, "Efficient and phase-aware video super-resolution for cardiac MRI," in *Medical Image Comput. Comput. Assisted Interv. – MICCAI*, Springer, Cham, Switzerland, Vol. 12264, pp. 66–76 (2020).
145. Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding," in *IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, pp. 5787–5796 (2017).
146. W.-S. Lai et al., "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 624–632 (2017).
147. R. Chen et al., "Single-frame deep-learning super-resolution microscopy for intracellular dynamics imaging," *Nat. Commun.* **14**(1), 2854 (2023).
148. R. J. Marsh et al., "Artifact-free high-density localization microscopy analysis," *Nat. Methods* **15**(9), 689–692 (2018).
149. J. Youn et al., "Detection and localization of ultrasound scatterers using convolutional neural networks," *IEEE Trans. Med. Imaging* **39**(12), 3855–3867 (2020).
150. D. Allen et al., "High-throughput imaging of CRISPR- and recombinant adeno-associated virus-induced DNA damage response in human hematopoietic stem and progenitor cells," *CRISPR J.* **5**(1), 80–94 (2022).
151. N. Boyd et al., "DeepLOCO: fast 3D localization microscopy using neural networks," *bioRxiv* (2018).
152. H. Zhou et al., "3D high resolution generative deep-learning network for fluorescence microscopy imaging," *Opt. Lett.* **45**(7), 1695–1698 (2020).
153. W. Zhang et al., "High-axial-resolution single-molecule localization under dense excitation with a multi-channel deep U-Net," *Opt. Lett.* **46**(21), 5477–5480 (2021).
154. P. Zelger et al., "Three-dimensional localization microscopy using deep learning," *Opt. Express* **26**(25), 33166–33179 (2018).
155. Y. Li et al., "Real-time 3D single-molecule localization using experimental point spread functions," *Nat. Methods* **15**(5), 367–369 (2018).
156. Z. Zhang et al., "Machine-learning based spectral classification for spectroscopic single-molecule localization microscopy," *Opt. Lett.* **44**(23), 5864–5867 (2019).
157. T. Kim, S. Moon, and K. Xu, "Information-rich localization microscopy through machine learning," *Nat. Commun.* **10**(1), 1996 (2019).

158. E. Hershko et al., "Multicolor localization microscopy and point-spread-function engineering by deep learning," *Opt. Express* **27**(5), 6158–6183 (2019).
159. J. Liu et al., "Deep learning-enhanced fluorescence microscopy via degeneration decoupling," *Opt. Express* **28**(10), 14859–14873 (2020).
160. J. Li et al., "Spatial and temporal super-resolution for fluorescence microscopy by a recurrent neural network," *Opt. Express* **29**(10), 15747–15763 (2021).
161. Y. Ma et al., "Cascade neural approximating for few-shot super-resolution photoacoustic angiography," *Appl. Phys. Lett.* **121**(10), 103701 (2022).
162. M. Guo et al., "Single-shot super-resolution total internal reflection fluorescence microscopy," *Nat. Methods* **15**(6), 425–428 (2018).
163. J. Soh, S. Cho, and N. Cho, "Meta-transfer learning for zero-shot super-resolution," in *IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, IEEE Computer Society, Los Alamitos, California, pp. 3513–3522 (2020).
164. Z. Burns and Z. Liu, "Untrained, physics-informed neural networks for structured illumination microscopy," *Opt. Express* **31**, 8714–8724 (2023).
165. H. Sahak et al., "Denoising diffusion probabilistic models for robust image super-resolution in the wild," arXiv:2302.07864 (2023).
166. S. Gao et al., "Implicit diffusion models for continuous super-resolution," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 10021–10030 (2023).
167. J. Ho et al., "Cascaded diffusion models for high fidelity image generation," *J. Mach. Learn. Res.* **23**(47), 1–33 (2021).
168. R. Rombach et al., "High-resolution image synthesis with latent diffusion models," in *IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, pp. 10674–10685 (2021).
169. A. Saguy et al., "This microtubule does not exist: super-resolution microscopy image generation by a diffusion model," *Small Methods*, 2400672 (2024).
170. M. Pan et al., "DiffuseIR: diffusion models for isotropic reconstruction of 3D microscopic images," in *Med. Image Comput. Comput. Assisted Interv.*, H. Greenspan et al., Eds., Springer Nature, Switzerland, pp. 323–332, (2023).
171. S. Gao et al., *GDPR Requirements for Biobanking Activities Across Europe*, pp. 10021–10030 (2023).
172. V. Colcelli et al., "GDPR requirements for biobanking activities across europe," Springer International Publishing, Cham, Switzerland (2023).
173. U.S. Department of Health and Human Services, "Health insurance portability and accountability act of 1996 (HIPAA)," Public Law, pp. 104–191 (1996).
174. U.K. Government, Data Protection Act 2018, c. 12 (2018).
175. Y. Yao and F. Yang, "Overcoming personal information protection challenges involving real-world data to support public health efforts in China," *Front. Public Health* **11**, 1265050 (2023).
176. D. Dhingra and A. Dabas, "Global strategy on digital health," *Indian Pediatrics* **57**(4), 356–358 (2020).
177. W. R. O. for Europe, "The protection of personal data in health information systems—principles and processes for public health," WHO Regional Office for Europe, Copenhagen (2020).
178. D. Jain, "Regulation of digital healthcare in India: ethical and legal challenges," *Healthcare* **11**(6), 911 (2023).
179. D. B. Larson et al., "Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework," *Radiology* **295**, 675–682 (2020).
180. A. S. Pillai, "Utilizing deep learning in medical image analysis for enhanced diagnostic accuracy and patient care: challenges, opportunities, and ethical implications," *J. Deep Learn. Genomic Data Anal.* **1**, 1–17 (2021).
181. R. Boudierhem, "Shaping the future of AI in healthcare through ethics and governance," *Humanities Soc. Sci. Commun.* **11**(1), 416 (2024).
182. N. Forgó et al., "Big data, AI and health data: between national, european, and international legal frameworks," in *Legal Challenges in the New Digital Age*, pp. 358–394. Edward Elgar Publishing, Cheltenham, England (2023).
183. S. T. Padmapriya and S. Parthasarathy, "Ethical data collection for medical image analysis: a structured approach," *Asian Bioethics Rev.* **16**(1), 95–108 (2024).
184. T. Willem et al., "Risks and benefits of dermatological machine learning health care applications: an overview and ethical analysis," *J. Eur. Acad. Dermatol. Venereol.* **36**(9), 1660–1668 (2022).
185. M. Jeyaraman et al., "Unraveling the ethical enigma: artificial intelligence in healthcare," *Cureus* **15**, e43262 (2023).
186. K. Grünberg et al., "Ethical and privacy aspects of using medical image data," in *Cloud-Based Benchmarking of Medical Image Analysis*, pp. 33–43, Springer International Publishing, Cham, Switzerland (2017).
187. Y. Li et al., "Virtual histological staining of unlabeled autopsy tissue," *Nat. Commun.* **15**(1), 1684 (2024).
188. H. Kumar and P. Kim, "Artificial intelligence in fusion protein three-dimensional structure prediction: review and perspective," *Clin. Transl. Med.* **14**, e1789 (2024).
189. J. Zhang et al., "Ai co-pilot bronchoscope robot," *Nat. Commun.* **15**(1), 241 (2024).
190. A. A. Banaeiyan et al., "Design and fabrication of a scalable liver-lobule-on-a-chip microphysiological platform," *Biofabrication* **9**(1), 15014 (2017).

Jesús Manuel Antúnez Domínguez is a biophysicist with expertise in microscopic approaches to bacterial collective behaviour. Holding an industrial PhD in biophysics, he has experience in both academia and industry, notably at the Innovation Unit of Elvys in Paris and the Biophysics Lab in the Department of Physics at the University of Gothenburg. His research interests span microfluidics, active matter, and advanced image analysis.

Giovanni Volpe is a professor of physics at the University of Gothenburg, with expertise in artificial intelligence, complex systems, and active matter. He leads interdisciplinary research exploring AI-driven solutions to understand emergent behaviors in biological and synthetic systems. He is currently co-authoring the book *Deep Learning Crash Course* for No Starch Press. His work integrates experimental, theoretical, and computational approaches, contributing widely to scientific literature and innovation.

Caroline B. Adiels is an associate professor of biophysics at the University of Gothenburg, with a focus on microfluidics, biology, and artificial intelligence applications. She leads an interdisciplinary research group dedicated to advancing single-cell analysis and communication studies using optics and microfluidics, which extends to organ-on-a-chip technology. Her work integrates AI-based image analysis software tailored for life sciences, creating a research portfolio that bridges the fields of physics and biology.

Biographies of the other authors are not available.