INVITED PAPER

Deep-Learning-Based Image Reconstruction and Enhancement in Optical Microscopy

This article provides an overview of efforts to advance the field of computational microscopy and optical sensing systems for microscopy using deep neural networks. First, the work overviews the basics of inverse problems in optical microscopy and then outlines how deep learning can be a framework for solving these problems, typically through supervised methods. Then, there is a discussion of use of deep learning to try to obtain single-image super resolution and image enhancement in these data sets.

By Kevin de Haan, *Member IEEE*, Yair Rivenson, *Member IEEE*, Yichen Wu, and Aydogan Ozcan, *Fellow IEEE*

ABSTRACT | In recent years, deep learning has been shown to be one of the leading machine learning techniques for a wide variety of inference tasks. In addition to its mainstream applications, such as classification, it has created transformative opportunities for image reconstruction and enhancement in optical microscopy. Some of these emerging applications of deep learning range from image transformations between microscopic imaging systems to adding new capabilities to existing imaging techniques, as well as solving various inverse problems based on microscopy image data. Deep learning is helping us move toward data-driven instrument designs that blend microscopy and computing to achieve what neither can do alone. This article provides an overview of some of the recent work using deep neural networks to advance computational microscopy and sensing systems, also covering their current and future biomedical applications.

KEYWORDS | Biomedical imaging; deep learning.

I. INTRODUCTION

Deep learning is a set of machine learning techniques that use multilayered neural networks to automatically analyze signals or data. These deep neural networks consist of several layers of artificial neurons, each of them typically incorporates a nonlinear operation (or activation function), which altogether can approximate an arbitrary function [1]. Deep networks have proven to be very effective for a wide variety of tasks ranging from natural language processing [2] to image classification [3]–[6] and playing games, such as Go [7], among others. Neural networks were first proposed in the 1940s [8] and have continually been developed for decades. Several types of deep networks, such as long short-term memory (LSTM) [9] and convolutional neural networks (CNNs) [10], have been developed over decades of research. Partially due to their shift-invariance property, CNNs have been particularly effective at processing, transforming, and classifying images. CNNs first began to be used for tasks, such as reading documents, in the 1980s and 1990s [11]. However, recently, a "perfect storm" of optimized software

Manuscript received April 22, 2019; revised September 21, 2019; accepted October 22, 2019. Date of publication November 14, 2019; date of current version December 26, 2019. This work was supported by the Koc Group, NSF, and HHMI. *(Corresponding author: Aydogan Ozcan.)*

K. de Haan, **Y. Rivenson**, and **Y. Wu** are with the Electrical and Computer Engineering Department, University of California at Los Angeles, Los Angeles, CA 90095 USA, with the Bioengineering Department, University of California at Los Angeles, Los Angeles, CA 90095 USA, and also with the California NanoSystems Institute (CNSI), University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: kdehaan@ucla.edu).

A. Ozcan is with the Electrical and Computer Engineering Department, University of California at Los Angeles, Los Angeles, CA 90095 USA, with the Bioengineering Department, University of California at Los Angeles, CA 90095 USA, with the California NanoSystems Institute (CNSI), University of California at Los Angeles, Los Angeles, CA 90095 USA, and also with the Department of Surgery, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: ozcan@ucla.edu).

Digital Object Identifier 10.1109/JPROC.2019.2949575

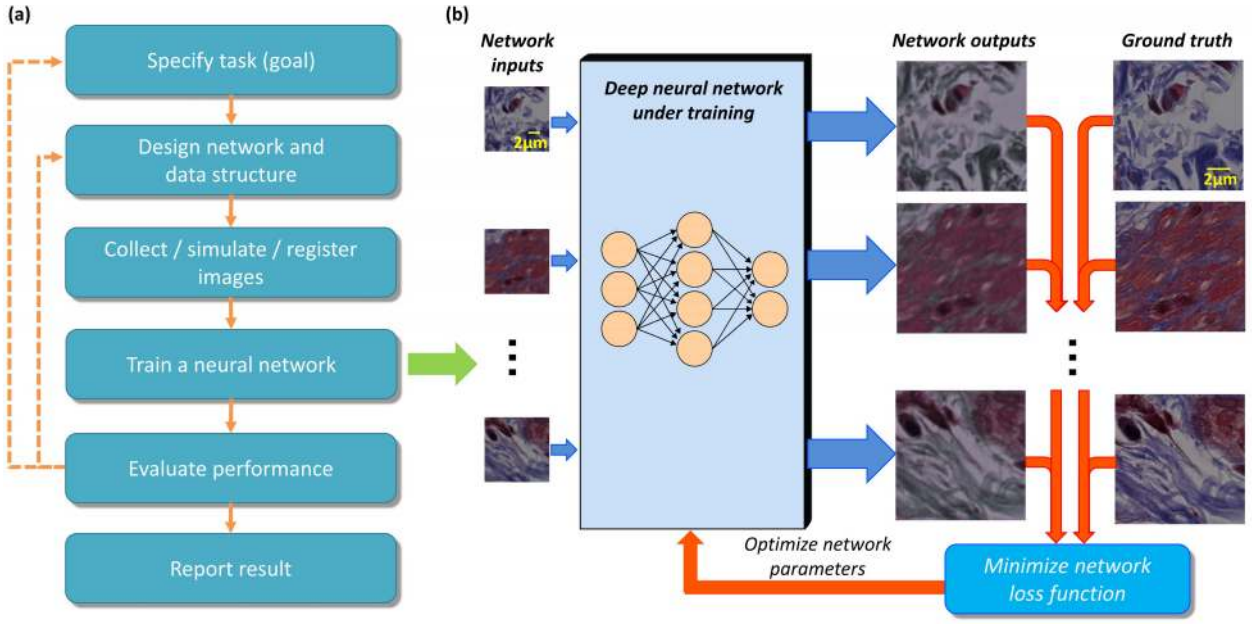

Fig. 1. ***Deep-learning-powered optical microscopy. (a) Typical workflow in constructing a neural network to perform a microscopic image reconstruction or image enhancement task. (b) Typical training flow of a neural network, using bright-field microscopy image super-resolution as an example. Modified from [53].***

[12]–[16], hardware (e.g., increased GPU power), and data availability has allowed deep learning (specifically CNNs) to be used as a recipe to tackle complex problems in many fields of research, most notably in computer vision. Many of these problems have included the task of image classification that is frequently used in various fields [17]–[19] and is one of the most fundamental problems in machine vision.

Beyond classification, deep learning presents many interesting opportunities to solve classical inverse problems in imaging, such as deblurring, super-resolution, denoising, and pixel super-resolution (or geometrical super-resolution). Inverse problems in imaging have a long and rich history, covering various ideas that are related to deep learning. These range from example-based super-resolution and dictionary learning techniques [20]–[23] to deconvolution methods [24]–[28] that require accurate knowledge about the image degradation model. Deep learning has also generated expansive literature in problems that are classically not categorized as inverse problems, such as style transfer [29] and image-to-image transformation [30], among others.

In this article, we will focus on deep neural networks' transformative power to solve inverse problems in microscopy using image data. Microscopy provides unique opportunities for using supervised deep learning to solve inverse problems in imaging. Unlike many other computer vision tasks, in microscopy, variables such as the physical properties of illumination, light-sample interaction, sample preparation, and positioning are fully under the user's control. These controllable degrees of freedom give the user the means to generate high-quality and experimentally obtained data that can be used to train a deep neural network without the need for any assumptions regarding the image formation or degradation model. Therefore, the gold-standard image data can often be experimentally generated rather than simulated. Learning from experimentally generated data has many advantages when compared to using simulations or simplifying assumptions, as in many cases creating an accurate forward model is not tractable.

In order to solve a microscopy-related inverse problem using deep learning, a neural network must be trained using a set of matching input and ground truth (or gold standard) images. Fig. 1 summarizes the basic workflow used to train a neural network to solve an inverse problem in optical microscopy.

In this article, we first discuss the implementation strategies for using deep networks in microscopic image reconstruction and enhancement, the data that are needed to train the network, and preprocessing procedures integral to the training process. We then discuss examples of inverse problems in microscopy that can be solved using deep learning. This article is broken up into the following sections. Section II introduces some of the main inverse problems in microscopic imaging. Section III discusses how deep networks can be used to solve inverse imaging problems. Section IV discusses microscopic imaging data generation and preprocessing. Section V demonstrates how to train a deep network using experimentally obtained image data to perform cross-modality microscopic image transformations through a working example. Section VI demonstrates deep learning-based super-resolution from a single image, where the transformations are between the images acquired by the same microscope. Section VII demon-

strates deep learning-based transformations between different types of microscopes, and Section VIII discusses implementations of deep networks where the gold-standard labels are computationally generated from a single input or a plurality of inputs. Section IX demonstrates some of the recently emerging biomedical applications for deep learning-enabled cross-modality image transformations in which deep networks learn to algorithmically create a physical transformation, e.g., for virtual staining of label-free tissue samples. Section X demonstrates new imaging capabilities that have been enabled by deep learning. Finally, we conclude this article in Section XI and discuss some of the strengths and weaknesses of these emerging techniques as well as possible future directions.

II. BRIEF OVERVIEW OF INVERSE PROBLEMS IN OPTICAL MICROSCOPY

While a number of imaging modalities will be discussed in this review, it will mainly focus upon the following three. The first of these modalities is bright-field microscopy [31], where white light illumination is modulated by the light-sample interaction and is collected by an objective lens. The second is fluorescence microscopy [32], which uses an illumination source to excite the fluorophores in the sample being imaged. The emitted photons from a fluorescent molecule have lower energy compared to the excitation photons, and by using a spectral filter to remove the excitation light, a high contrast fluorescence image of the sample can be acquired. This technique is often used to image samples that are specifically labeled with fluorescent markers. The third form of microscopy commonly referenced in this article is digital holographic microscopy [33]–[36]. Holographic microscopy is typically implemented by illuminating the sample with coherent (or partially coherent) light and uses, e.g., a transmission geometry to create an interference pattern of the sample's transmission function. This, in turn, is reconstructed to extract both the phase and the amplitude information of the object. The phase information channel is related to the optical-path length of the light passing through the sample, whereas the amplitude information channel is related to the absorption and scattering properties of the sample.

For a microscopic imaging system, the discrete imaging forward model, which sets the stage for an inverse imaging problem, can be written as

$$g = Hf + n \tag{1}$$

where $f \in R^{N\times 1}$ and $g \in R^{M\times 1}$ are the object and measurement information, respectively (f and g represent the lexicographical arrangement of a 2-D/3-D signal, which can be complex valued as in the case of coherent microscopy), n is an additive noise term (which can be signal dependent), and H is the mapping operator between the object space and the image/measurement space (which can also be nonlinear). For standard inverse imaging problems, one would like to obtain the optimal object approximation, $\hat{f}$, which satisfies the constraints imposed by (1). In the literature, various optimization techniques have been used for this task [37]–[39]. However, most of these approaches require accurate knowledge of H and often some *a priori* information about the object (e.g., a sparsity constraint [40]).

Microscopic imaging, in general, shares a number of inverse problems that are in common with other computer vision tasks while also presenting its own unique challenges. One of the most well-known of these inverse problems is deconvolution or deblurring

$$g = h^* f + n \tag{2}$$

where h is a low-pass filter and * is the spatial convolution operation. In optical microscopy, the finite aperture of the objective lens limits the extent of the spatial frequencies, which, in many cases, can be estimated as a low-pass filter operator with a Gaussian kernel. Equation (2) assumes a shift-invariant imaging operator; however, in practice, the blurring operator is often shift variant for many microscopic imaging systems due to factors, such as aberrations, introduced by various optical components as well as the mismatch of the sample refractive indices and the optical medium (including the objective lens). This changes the point spread function (blurring kernel) throughout the field of view (FOV). In that case, (1) should be used instead of (2). However, an accurate estimate of H would be tedious and practically impossible to acquire as it also depends on the object properties, which are, by definition, unknown for an unknown object.

Classically, the smallest feature that can be resolved by an imaging system is given by [41]

$$d = \frac{\lambda}{2\text{NA}} \tag{3}$$

where λ is the illumination wavelength and NA is the numerical aperture of the imaging system, which defines the ability of the lens to gather diffracted object light from a fixed distance. In addition to the limits imposed by an objective lens and its NA, microscopy resolution is fundamentally limited by the wavelength of the light itself. This diffraction limit is approximately $\lambda/2$ [41], which restricts modern optical microscopy techniques to have a resolution of $\sim$200–300 nm unless super-resolution methods are employed to beat the diffraction limit of light. For example, optical, computational, and statistical techniques have been developed to break the diffraction limit and achieve super-resolution in fluorescence microscopy [42]–[46]. These super-resolution techniques have significantly expanded the usage of optical microscopes in various fields, such as biology by enabling discoveries at the nanoscale. However, some of these fluorescence super-resolution techniques require specialized and expensive

equipment, with relatively high-power illumination and a large number of image exposures. This can be potentially harmful for live cell imaging and other applications where phototoxicity is a concern [47], [48].

Another inverse problem of interest in optical microscopy is dealiasing or pixel-super-resolution. Following the notation in (1), we can define

$$g = \mathrm{SHf} + n \tag{4}$$

where S is the decimation operator that maps from $R^N \rightarrow R^M$ (where $M < N$). While such pixelation-related problems can be addressed by using higher magnification objective lenses, this also comes with a significant tradeoff in imaging FOV, which generally scales down with the square of the magnification. Therefore, the pixel-super-resolution inverse problem is of broad interest for wide-field imaging applications using low magnification systems, such as lensfree microscopic imaging, where the image resolution is often pixel size limited [35].

In some cases, the image formation model can also be described by a nonlinear relationship (H) between the object and its image, such as

$$g = H'(f) + n. \tag{5}$$

An interesting example of this nonlinear relationship is the mapping from a label-free image of a sample to an image of the same sample following a physical or chemical labeling process taken using a different imaging modality [49]–[52]. In the following, we will demonstrate the effectiveness of data-driven deep networks in solving such nonlinear inverse problems in imaging.

III. DEEP LEARNING AS A FRAMEWORK TO SOLVE INVERSE PROBLEMS IN OPTICAL MICROSCOPY

Here, we provide a brief overview of how deep networks can be used to estimate $\hat{f}$ for a given measurement g by learning a statistical image transformation. A primary example of this technique being used in optical microscopy has been for spatial resolution enhancement using deep learning [53]. Single-image super-resolution or image enhancement is, in general, an ill-posed problem. However, deep learning has been shown to be able to leverage large amounts of well-registered image data to learn statistical transformations through high levels of abstractions in order to improve upon conventional super-resolution and image-enhancement algorithms and achieve superior results [54]. In addition to super-resolution, solutions to other inverse problems in microscopic imaging, such as virtual staining [49] and holographic image reconstruction [55], [56], have also been tackled using deep learning.

Deep neural networks typically learn to solve inverse imaging problems through supervised learning. Supervised learning utilizes "gold-standard" labels that are known in advance and are matched to corresponding input images. These networks are used in a feed-forward fashion, where they are given an input image (to be improved) that passes information from one neural network layer to the next. Feed-forward networks can be made up of several different types of layers, and all of them operate using the same basic principles, where weights, biases, and other trainable parameters are trained using error backpropagation [57]. Once the network has been fully trained, these variables are fixed, and any test image can be inferred by the network in a single feed-forward step, without the need for any iterations. This stands in contrast to many other solutions to inverse problems in microscopy, where hyperparameters need to be carefully hand-tuned, often through an iterative process to achieve optimal convergence. By performing the desired transformation in a single feedforward manner, deep networks, in general, outperform traditional iterative methods in terms of inference speed [55], [58]–[60]. Furthermore, as the parameters are fixed after the training phase, modifications or adjustments are not needed each time it is applied to a new image, increasing its overall usability. Finally, deep neural networks, being data driven, can solve inverse problems for which numerical formulation of a forward model is very difficult or even intractable [61]–[63].

A. Basic Elements of a Neural Network

Several types of layers are used by neural networks. These include fully connected layers that have a connection between each neuron in two consecutive layers: pooling layers, which find, e.g., the maximum or average of adjacent tensor values; and convolutional layers, which provide limited and spatially invariant connectivity across successive layers. Modern deep networks that perform image transformations are typically fully convolutional networks. Using convolutional layers is advantageous as it brings shift-invariance [1], which is of particular use for microscopy applications. As weights and biases are shared, they are applied in the same manner regardless of the sample location. Therefore, due to this parameter sharing, any shift to the input also shifts the output in the same manner. Using only convolutional layers also allows for scaling of the network, where the testing can be performed on images of different sizes. Another desirable property of deep CNNs along with other types of neural networks is that the convolutional layers can create hierarchical data representations rather than relying upon hand-engineered features [64]. Each convolutional layer generates feature maps based on the previous layer. These feature maps are used to extract and preserve important information regarding the input object and are stored as separate channels in a tensor; the general operation of that can be described as

$$v_{\alpha,\beta}^{i,j} = b_{\alpha,\beta} + \sum_{r} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} w_{\alpha,\beta,r}^{p,q} v_{\alpha-1,r}^{i+p,j+q} \tag{6}$$

where $v_{\alpha,\beta}^{i,j}$ represents the value of the pixel at coordinate i, j in feature map β of convolutional layer α. P and Q represent the size of the convolutional kernel (e.g., 3×3), $w_{\alpha,\beta,r}^{p,q}$ is the value of the convolution kernel at position p, q, and $b_{\alpha,\beta}$ is a trainable bias parameter. Finally, r represents the feature maps in layer $\alpha - 1$.

After a given network layer, a nonlinear activation function is typically used. While a number of nonlinear functions have been used in the literature, variants of the rectified linear unit (ReLU) have become among the most popular functions [65]. It is described as

$$\text{ReLU}(x) = \begin{cases} x, & \text{for } x > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

ReLU is typically used because it is easy to calculate and helps to mitigate the vanishing gradient problem [66], which fosters highly efficient training of deep neural networks [1], [65], [67].

B. Neural Network Architecture

A variety of deep neural network architectures have been developed, but there are a few general network types that are among the most commonly used and have been repeatedly shown to be effective. Regardless of which general network architecture is used, several parameters of the network need to be fine-tuned. For example, network size can be changed by varying both the number of layers making up the network as well as the number of channels in each layer of the network. Tuning of these parameters is required to ensure that the network can extract and learn any important features while not being so complicated that it overtrains [68], [69]. Another important consideration when choosing the network architecture and filter sizes is to make sure that the effective receptive field (i.e., the region of the input space that affects a particular unit of the network) is sufficient to contain all the required information [70]. In fact, an active area of research is the development of environments that support automatic and optimal tuning of these hyperparameters of a neural network [71]–[73].

Two of the most popular network architectures for image transformations that have been shown to generate state-of-the-art results are U-net [19], [74] and ResNet [6]. The U-net architecture was originally proposed to perform biological image segmentation and has been broadly applied to image transformation applications since then. This network uses a series of convolutional blocks and skip connections to allow the network to be deeper without having issues related to vanishing gradients. It is made up of a series of downsampling blocks, which use convolutional layers to process the images, and pooling layers to downsample the images. The downsampling blocks are followed by an equal number of upsampling blocks that use convolutions to upsample the images until they reach the original image size. Between each pair of same-sized upsampling and downsampling blocks, a skip connection is used to pass data. This structure allows the network to learn features at different spatial/size levels, and the downsampling allows it to have a large receptive field if needed. ResNet [6], on the other hand, consists of a series of convolutional blocks that maintain the image size, with a residual connection passing the data past the block. Similar to U-net, the residual connections allow the network to have many layers and learn features without having vanishing gradient related issues, making it easier and more effective to train.

C. Loss (Cost) Function

During the training process, the deep network attempts to predict the gold-standard label images from the input images of the training data set by minimizing a cost/loss function. This cost function can be user-defined and is often based on per-pixel differences between the output and ground truth images, such as the mean absolute difference (L1-norm) and mean-squared error (L2-norm). Some of the cost functions can also penalize structural losses, such as the structural-similarity index (SSIM) [75], or can be custom-designed complex functions that are generated by training another network either offline (e.g., perceptual loss) or online, such as a generative adversarial network (GAN) [76], which can adaptively learn the optimal loss function based on the training image data. The GAN framework has played an important role in some of the recent applications of deep learning for optical microscopy and will be explained in greater detail in Section III-D.

D. Generative Adversarial Networks

GANs can be used to improve the overall loss function and allow the network to create realistic-looking images without requiring any specific feature engineering [76]. They were designed to create artificial images that match the feature distribution of a target data set. GANs use two distinct networks. The first network (known as the generator, G) is used to generate images using an input (x), while the second network (known as the discriminator, D) attempts to discriminate between the generated images [$G(x)$] and the ground truth images (z). The discriminator network adds loss to the generator that can be described by the following equation:

$$l_{\text{generator}} = [1 - D(G(x))]^2 \tag{8}$$

where $D(G(x)) = 0$ means that the discriminator can successfully spot a generated image from the ground truth image. This loss function drives the generator to learn how to "trick" the discriminator to classify its outputs as ground truth. On the other hand, the discriminator loss can be described by the following equation:

$$l_{\text{discriminator}} = D(G(x))^2 + (1 - D(z))^2. \tag{9}$$

This loss function is used to teach the discriminator to learn how to distinguish between the generated images and the ground truth label images. For image transformations, GANs typically work best when used in conjunction with either $L1$, $L2$, or perceptual loss. These terms help to regularize the GAN performance if a known gold-standard image data set is available. As an example, for the use of an Lp-norm term, the generator's loss function can be expanded to

$$l_{\text{generator\_total}} = [1 - D(G(x))]^2 + \tau \times L_p\{G(x), z\} \quad (10)$$

where τ is a regularization constant used to balance the relative importance of the generator loss (8) within the total loss function. When using a loss function as in (10), the GAN portion of the loss is used to force the generator to create realistic images, while the second portion of the loss is, in general, used to ensure that the transformation for each pixel is correct and that the GAN is regularized/constrained by the input/output relationship. By balancing the relative importance of these loss functions, together they can train a generator network that performs the desired image transformation in optical microscopy.

IV. CREATING MICROSCOPIC IMAGE DATA SETS FOR TRAINING DEEP NEURAL NETWORK MODELS

While there is a massive amount of biomedical imaging data being created every day, unfortunately, much of it is not labeled suitably to be used for deep learning-based image enhancement or transformations. Since supervised learning requires gold-standard labels, the generation of high-quality data sets to train the deep network is very important. CNNs have proven to generalize very well, but they are much more effective when trained on a data set that is designed specifically for the test data set that it will be used for. For example, when a neural network is trained using one specific microscopy system and applied to another, it will be able to make use of the learned properties that are shared between the two systems. However, it will not learn any properties that are not part of the training data set. This can be mitigated by performing calibration experiments with the new microscopy system of interest and using, e.g., transfer learning [54], [77].

A. Microscopic Image Registration for Training Data

The training image data sets can be created in a few different ways. Most commonly, the gold-standard label and input images are imaged separately (by, e.g., different optical microscopy hardware) and then accurately matched to each other; alternatively, the label images can be acquired using a ground truth microscope, and matching input images can be digitally generated using numerical degradation (following, e.g., a physical model). Following the notation of (1), in this data set preparation mode, we acquire f and then simulate $\tilde{g} = \tilde{H}f + n$, where $\tilde{H}$ represents the approximate forward (sensing) model and $\tilde{g}$ is the approximation of the measured corresponding image, which will be used as input for the deep network training phase. Simulated data sets are typically much easier to create, but the efficacy of the network performance depends upon the accuracy of the forward/sensing models, which, in some cases, is a challenge to obtain beyond a simplified approximation. On the other hand, if both input (f) and ground truth images (g) are experimentally obtained, they must be accurately matched to each other, as the goal of the enhancement network is to predict the output pixel values based on the input pixels. This image registration for the preparation of the training data is a one-time effort and can be done in different stages, which will be detailed in the following.

A third method that can be used to generate training image data is to computationally reconstruct gold-standard label images. A good example of this is the use of a multiframe pixel super-resolution algorithm, where a high-resolution image is synthesized from a set of low-resolution images that are shifted with respect to each other. In this case, we can train a deep network to learn the transformation between a single low-resolution image and the computationally generated high-resolution image that is pixel super-resolved. Additional examples (other than pixel super-resolution) of this method based on training image data generation will be presented in Section VIII.

When both the gold-standard labels and the input images are experimentally acquired, there are three main types of coregistration techniques that are used to spatially match these images to each other: 1) coregistration based on intensity (such as the intensity difference or intensity cross-correlation); 2) coregistration based on features (such as the Euclidean distance between extracted shapes) [78]; and 3) deep learning-based registration [79], [80]. Of the algorithmic registration methods, intensity registration typically works well when the two sets of images are of the same optical modality or of similar modalities. On the other hand, feature-based registration methods can be more general while being less effective at the subpixel level.

Image registration can be further divided into either rigid or nonrigid transformations [81], [82]. Rigid transformations are only capable of rotation and translation, while nonrigid transformations allow the registered images to be deformed. Nonrigid registration allows accurate matching, but distortion can occur if there are differences between the input and ground truth label images. One way to improve the quality of the registration is to use a deep neural network to perform a "soft" form of the desired data transformation and use this network's output image as the target of the registration algorithm [49]. While the network will not produce an ideal image, it can create intermediate images, which can be used for intensity-based cross-registration. This process can be repeated as

needed, improving the transformation network and, therefore, the overall registration accuracy, iteratively.

Depending on how the data were acquired, a combination of registration steps may be needed. There can be a large variation between the images, especially when captured using different microscopy modalities. Therefore, a unique workflow should be tailored for a given application of interest [50], [51], [61]; only a general overview of image coregistration is presented here, but one specific method for integrating image coregistration into the training workflow is demonstrated as an example in Section V.

It is important to note that image transformation networks, such as CycleGANs [83], do not require the input and ground truth images to be coregistered and still have been shown to create highly realistic images. However, without input-ground truth registration, there is a possibility for artifacts, which might be detrimental to various biomedical applications, particularity for clinical use as the outputs need to be highly consistent to ensure accurate diagnoses.

B. Quality and Size of the Training Image Data

Poor-quality data can have a dramatic negative effect on the performance of the deep network. With training image data set that is either not relevant or improperly labeled, the error-backpropagation will not have a correct gradient to follow and the network will be trained inaccurately. Even small inconsistencies in the training image data set can cause issues. This is particularly evident when training a GAN as the discriminator loss can collapse, rendering it completely ineffective.

Limited image data can also be a constraint for applications of deep neural networks in optical microscopy. This can be exacerbated for certain microscopy applications where either samples or images are rare and/or challenging to obtain. While deep learning has been shown to be highly effective on small data sets [84], a neural network will generalize better when the training data set is large enough that the network can learn the entire sample space from it. Depending on the sample variation, the required training data size for the examples given in this article can range from a few thousand image patches to hundreds of thousands of patches or more. For example, if the network is being trained on a portion of a sample and being tested on the rest, a small data set can potentially be used. On the other hand, if the network needs to generalize across a diverse set of samples, significantly more image data is typically required. If the network is large and the size of the training data is limited, the network can also overfit to the training data set and not generalize well to test data sets.

One popular method for improving the performance of neural networks with limited training data is transfer learning [77], [85], [86]. This process typically involves training a network on a large data set that has similar features to the data set that the network will be used. Following this training, the last few layers of the network can be retrained using the smaller data set of interest. These retrained layers allow the network to learn the transformations for the specific data set, while the intermediate layers do not need to be retrained as they are simply used to extract common features, which can, in some cases, be generalized from one data set to the next.

Another method to help reduce issues and artifacts related to limited training data is to augment the image data set using techniques, such as image rotation, flipping, shifting, and distortion [87]. However, this is often not enough to overcome the generalization challenge entirely. It will help prevent a network from overfitting, but the network will still not have any new information about the features that were not well represented in the original training data set.

C. Image Data Normalization

Normalization of both the input image data and the ground truth labels can significantly improve the consistency of the network inference. Most deep networks are highly nonlinear, meaning that small differences at the input images can cause relatively large differences at the network output. Therefore, any variations in the illumination source intensity or the exposure time of the microscopy system need to be normalized during the training and testing phases.

V. DEEP LEARNING AS A FRAMEWORK TO SOLVE INVERSE PROBLEMS IN OPTICAL MICROSCOPY

As a working example, in this section, we will explain how to train a network that can transform images taken by a cost-effective smartphone-based microscope into images equivalent to those taken by a high-quality benchtop microscope. While this example will detail a specific transformation, this approach is broadly applicable to other cross-modality image transformation or enhancement applications.

Before the neural network can be trained, an image data set needs to be created. For the transformation from cellphone to benchtop microscopy images, this training data set will consist of inputs made up of the cellphone images and the gold-standard label images of the same samples taken by the benchtop microscope. For proper training of the network, it is vital that these images are accurately registered, especially since, in this transformation, the network's task is to predict pixel values in the ground truth image based on the values of the input pixels. For this image transformation, the training images need to first have the corresponding FOVs matched. This can be done by first digitally stitching the microscope images together into a single large image of the sample and then using a coarse correlation-based registration to find and crop out the area corresponding to each cellphone image.

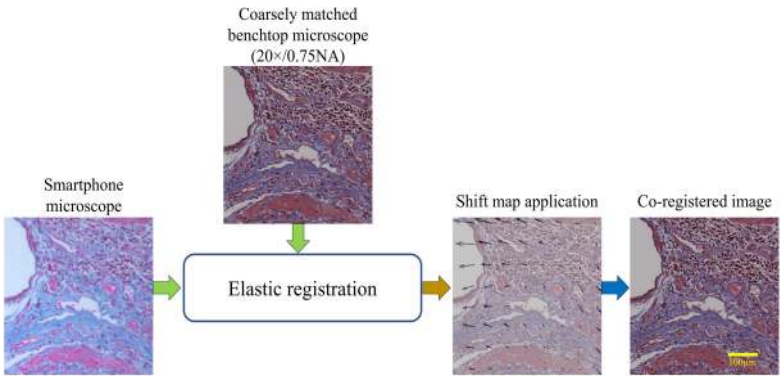

Fig. 2. *Example of the elastic coregistration between smartphone and benchtop microscope images. Modified from [61].*

Next, an elastic registration [88] can be used to ensure that a subpixel-level coregistration is achieved (see Fig. 2).

Elastic registration is preferred in this example, as cellphone-based microscopes will likely have significant aberrations due to inexpensive camera lenses (in comparison to those used in benchtop microscopes), in which a rigid registration will be unable to account for. As both sets of images are captured in the bright-field mode, an intensity-based registration algorithm can be used to achieve a high degree of registration accuracy between the input and label images.

Once the image data have been coregistered (e.g., Fig. 2), it must be normalized and split into separate training and validation sets (the testing data set does not require accurate coregistration). Following this, the network can be created using an architecture that fits the image data set. As discussed in Section III-B, there are several popular architectures that have been proven to be highly effective, with U-net and ResNet architectures being among the most popular. Along with the network architecture, the loss function for the network training also needs to be chosen. For this image transformation, a GAN that uses a U-net as the generator and a VGGNet style network as the discriminator can be chosen (see Fig. 3). The depth and the number of channels within each of these networks depend upon the size of the image data set, as it must be large enough to learn the feature space but not so large that it overstrains. As detailed in Section III-D, in addition to the GAN loss, an L1 loss can be added to ensure that the color and intensity of each pixel are accurately inferred.

Typically, the lowest pixel-based validation loss (in this case, L1) is used as the inference model, and the network training can be stopped when this loss begins to increase as it indicates that the network is overtraining. However, if the transformation of certain features is more important than others, it can be beneficial to compare/quantify how these features behave for different models. These different training stopping conditions can also regularize the deep network [89]. Once the network has been fully trained and the final model has been chosen based on a validation loss, it can be used to blindly enhance the new images taken by the smartphone microscope. The image enhancement results that can be obtained using the training methods outlined in this section will be exemplified in Section VII.

VI. SINGLE-IMAGE SUPER-RESOLUTION IN MICROSCOPY USING DEEP LEARNING

The working example in Section V is just one example of how an inverse problem in microscopic imaging can be solved using deep learning. In this current section and the upcoming ones, a wide variety of such inverse problems will be discussed.

We first examine one of the most widely needed image transformations in microscopy— the conversion of a low-resolution image into a higher resolution image, where both the input and ground truth (higher resolution image) are taken by the same microscope. A training image data set for a network that performs this transformation can be created by, e.g., scanning the same samples with a low- and high-NA objective lens. Following the data acquisition step, the deep network is trained to perform statistical transformation acting on the low-resolution input images to match the corresponding high-resolution labels.

For microscopy applications, this form of super-resolution is far reaching as it allows for larger FOVs to be measured per acquired image, which can be particularly useful for high-throughput, time-lapse imaging. In other words, low-NA objective lenses can image a much larger FOV, as illustrated in Fig. 4(a); thus, the technique improves over the native space-bandwidth product of an objective lens [41]. This super-resolution image enhancement has been demonstrated for fluorescence, bright field, and coherent (holographic) imaging systems, as illustrated in Fig. 4(b)–(d), respectively.

In theory, the missing spatial frequencies (although not detected) in a microscopic image can be extrapolated based on the measured (or *a priori* known) spatial frequencies of an object [90], using, for example, the principle of analytical continuation [91]. However, in practice, the success of such frequency extrapolation methods is fundamentally related to the imaging system's signal-to-noise ratio

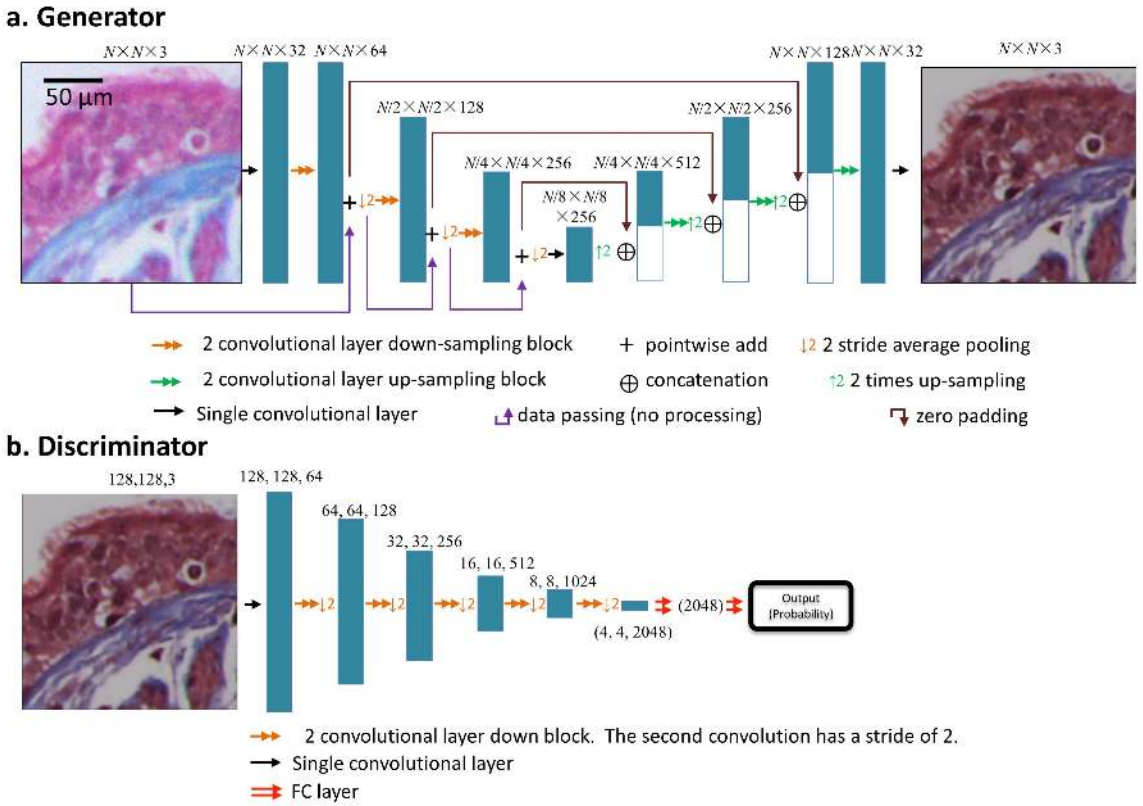

Fig. 3. ***Diagram of example network structures used in GAN. Each upblock and downblock are made up of two convolutional layers with a kernel size of 3 × 3 and followed by ReLU or leaky ReLU activation layers. The downblocks reduce the size of each channel by a factor of two in each dimension and double the number of channels, while the upblocks increase the size and reduce the number of channels by a factor of four. Images were taken from [61].***

(SNR) [90]. While deep networks for deconvolution do not explicitly include any analytical continuation models, the deep network learns to efficiently separate the noise from the signal structures, which effectively improves the frequency extrapolation capability of the network; stated differently, noise features in images are much harder to generalize by a trained network compared to the actual features of the imaged objects. In cases where the feature space of the model (e.g., the number of weights and biases) is much larger than the number of training examples, the network may begin to overfit to noise as well, reducing the overall effectiveness of the network. One of the ways to avoid this problem is to use AutoML framework [92], [93], which learns to automatically optimize the model during the training process.

Another interesting effect of the deep learning-based image enhancement is the extended depth of field (DOF). Just as in photography, the DOF is defined by the distance from the nearest object plane in focus to that of the farthest plane simultaneously in focus [94]. In order to understand the extended DOF (EDOF) effect of deep networks, we must recall that the DOF imaged by a microscope is proportional to λ/NA^2. In other words, for low-NA objective lenses, the DOF will be significantly extended in comparison to the shallower DOF of high-NA lenses. When the deep network learns to enhance the resolution of the images that are acquired using a low-NA objective lens, it does that by enhancing all the spatial details within the DOF. This effect has been validated for both bright field [53], [61] and fluorescence microscopy [54], by comparing the deep network output image to an algorithmically generated EDOF image from a z-scan (depth scan) of the sample using a high-NA objective lens. A demonstration of this EDOF effect can also be seen in Fig. 5. This effect has the potential to further increase the throughput of the imaging system, as the features that are out of focus for a high-NA objective can remain in focus with high resolution at the deep network output.

Interestingly, in many cases, the deep networks were also able to generalize to types of samples that were not part of the training set. For example, super-resolution of bright-field microscopy images that were trained with one tissue type and tested on another type of tissue was successful [53]. Similarly, in fluorescence microscopy, a network trained to super-resolve images of mitochondria was also able to super-resolve blood vessels and actin in different tissue types [54].

For most microscopic image enhancement techniques, the goal is to solve a deconvolution/deblurring problem, which can often be estimated as a convolution with a Gaussian kernel [95]. For this form of simple

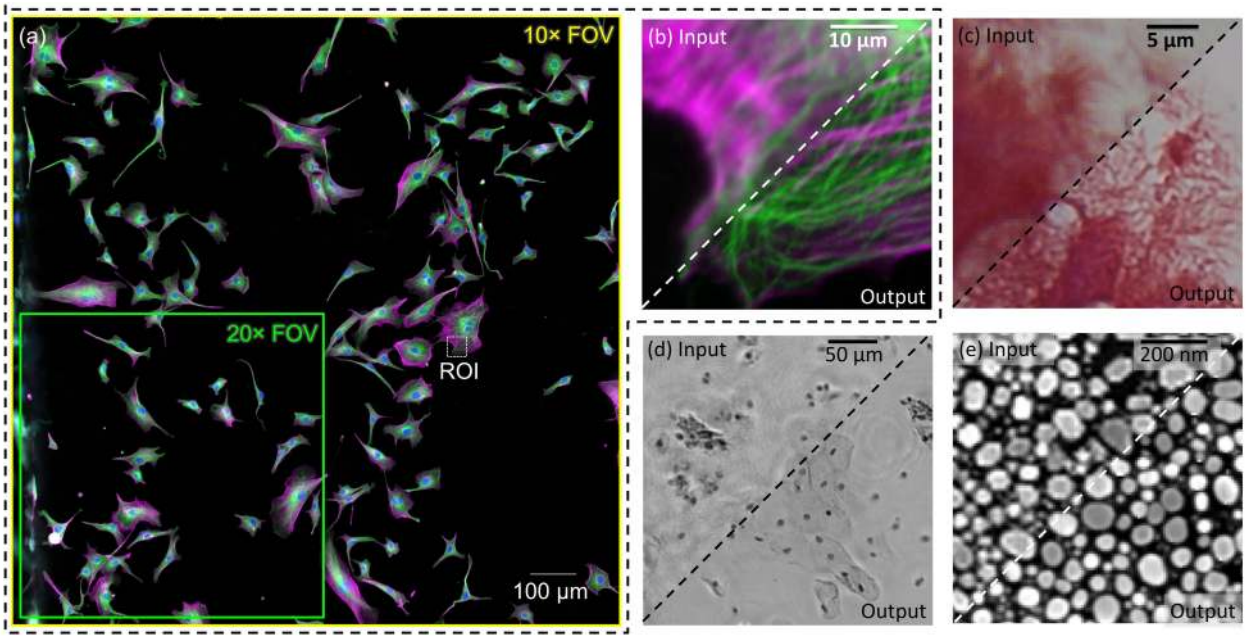

Fig. 4. *Super-resolution and image quality enhancement using deep learning. (a) Fluorescence microscopy image captured with a 20× objective lens and 10× FOV marked. (b) Zoomed-in region of (a) showing a network perform super-resolution of a 10×/0.4 NA input image. Images were taken from [54]. (c) Super-resolution of a bright-field microscope image. Images were taken from [53]. Input image is taken using a 40×/0.95 NA microscope objective. (d) Super-resolution of an image created using digital holographic microscopy [59]. Input is imaged using a 4×/0.13 NA microscope objective. Images were taken from [59]. (e) Super-resolution of an SEM image. Images were taken from [95]. Input is imaged using 10 000× magnification.*

deconvolution, it is possible to train the deep network using simulated image data [96], i.e., we can assume an accurate knowledge about the blurring kernel and simulate the low-resolution data from the high-resolution images to create the training image set. However, in most practical cases, this estimation will not yield satisfactory results. A more accurate model can be a shift-variant convolution, where different parts of the image FOV correspond to different convolution kernels; in fact, this spatial variance depends on the imaging system as well as the sample preparation itself. When trained using experimentally generated input image data and labels, neural networks are capable of learning to perform this shift variant deconvolution [54] (see Fig. 6). If simulated data were to

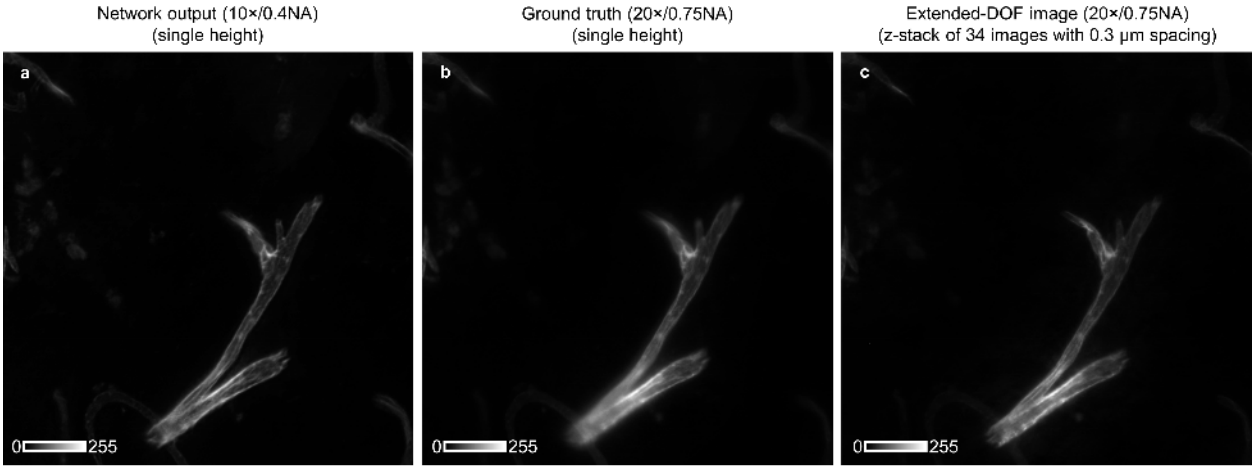

Fig. 5. *Demonstration of the EDOF using deep learning. (a) Output of the network in response to a low-resolution image. (b) Reference image taken at a single focal depth of the microscope with a higher numerical aperture objective (higher resolution and shallower DOF). (c) EDOF image created using 34 images from different focal depths (with a step size of 300 nm) taken using the same high numerical aperture objective. Images were taken from [54].*

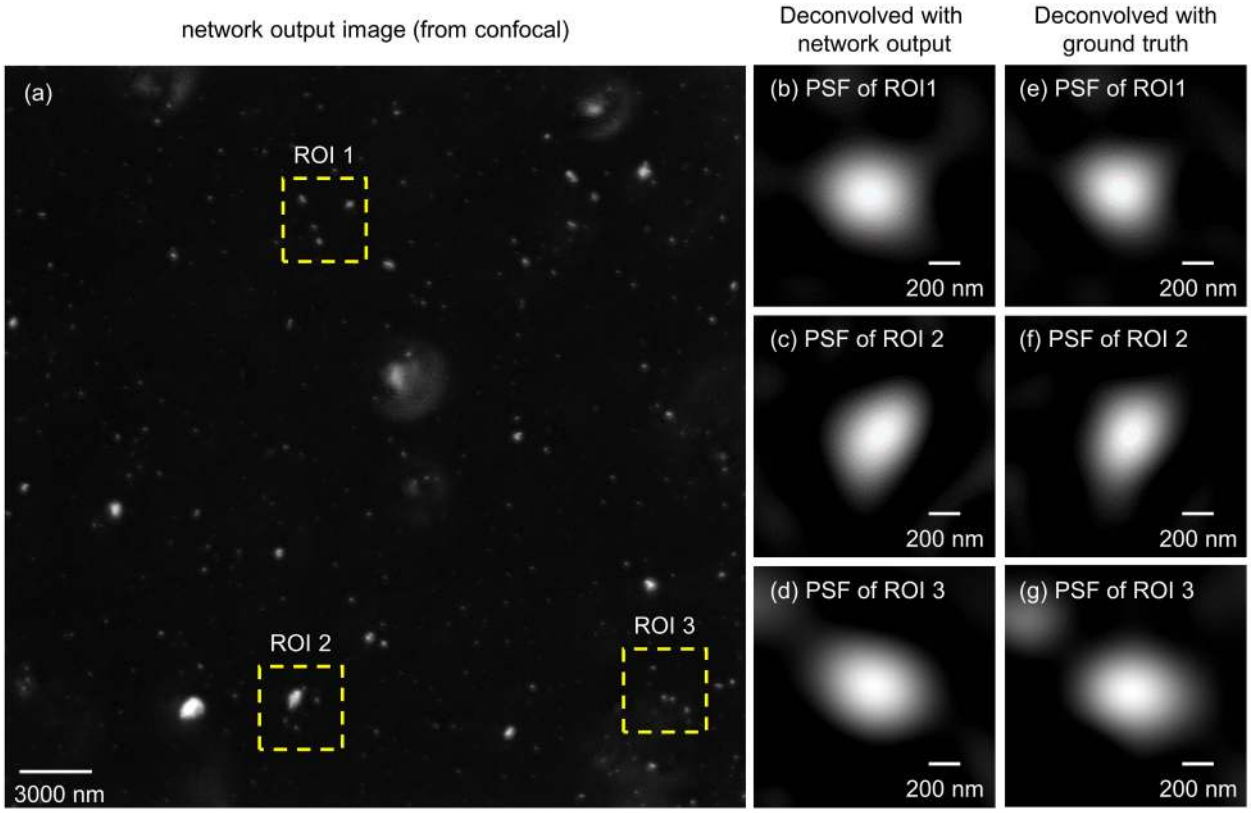

Fig. 6. *Examples of the spatially varying point spread functions (blurring kernels) that are learned by the network for different regions of interest (ROIs). (a) Network output image. (b)–(d) Demonstrations of the PSF for various marked ROIs created using a deconvolution between the network input and output. (e)–(g) Demonstrations of the PSF for these ROIs created using a deconvolution between the network input and ground truth image. The network input for these images is generated using a confocal microscope, while the ground truth images are captured using a STED microscope. This demonstrates that the network is capable of learning to deconvolve spatially variant PSFs from image data only. Images were taken from [54].*

be used to train the network, it would be unable to learn such spatial variations unless an accurate account of how the point spread function varies over the FOV is available/known.

In addition to light microscopy, similar techniques have also been applied to enhance images acquired by a scanning electron microscope (SEM) [97] [see Fig. 4(e)]. SEM imaging is not limited by the wavelength of electrons but instead by aberrations and pixel size of the imaging system [see (4)]. For the latter, dealiasing through deep learning can be used to perform pixel super-resolution of electron microscopy images. One of the main advantages of using this approach to enhance SEM images is to reduce the electron beam's radiation, as it can be damaging to materials that are soft or poor conductors, such as biological tissue. This is because a high electron density is required at high magnifications since each pixel must be exposed to a certain number of electrons to ensure an adequate SNR. By using computational super-resolution in conjunction with imaging the sample at a lower magnification, the beam dwell time and, therefore, electron exposure of each unit area can be reduced without a reduction in SNR. Therefore, charging and electron beam damage can be reduced. SEM can also be expensive and slow to image large areas at the nanoscale level, which could be improved using the same method.

VII. TRANSFORMATIONS BETWEEN MICROSCOPY SYSTEMS

While enhancing the input images to match higher resolution images that are natively taken using the same microscopy platform can increase the imaging system's throughput, another interesting avenue is to learn a statistical transformation between two different microscopic imaging modalities. Learning these cross-modality transformations where the ground truth image set is made up of images taken by a different microscope allows the network to achieve results that are not possible using standard forward model-based inverse problem solutions (see Fig. 7). Using this deep learning-based approach, cost-effective or simpler microscopes can take the same quality of measurements as the gold-standard microscopes, helping to democratize microscopy-related research and innovation. Several examples of this exciting opportunity will be discussed in this section.

A prime example of a system that could benefit from image enhancement is mobile phone microscopy. The general procedure for training a deep neural network

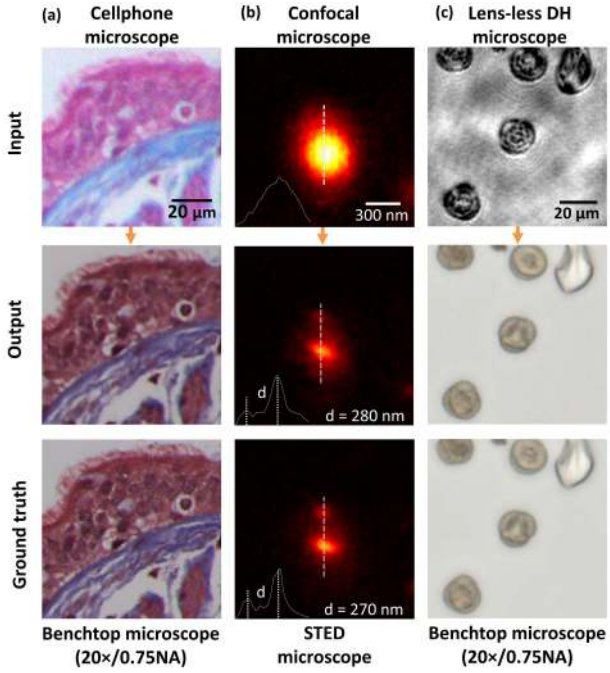

Fig. 7. *Transforming images between different microscopy modalities. (a) Cellphone-based microscope image transformed into a benchtop microscope equivalent image. Images were taken from [61]. (b) Confocal microscope image transformed into a STED equivalent image. Images were taken from [54]. (c) Lensless digital holographic microscope (DHM) image transformed into a benchtop bright-field microscope image. Images were taken from [106].*

using experimentally obtained image data was discussed in Section V; in this section, we will provide examples of the image enhancement that one can obtain in mobile phone microscopy using deep learning.

Mobile phone-based imaging has emerged as an alternative to traditional benchtop microscopes for various applications targeting low-resource settings, ranging from imaging blood smears to providing a quantitative readout for various diagnostic assays [98]–[101]. Mobile phone microscopes typically have low-NA lenses; furthermore, due to the portable nature of these microscopes and the demand to keep the device cost-effective and lightweight, the design constraints dictate a compromise on the optical components that are used in mobile phone-based microscopes. In addition, the pixel size of a CMOS image sensor used in a mobile phone is much smaller than that of a typical CCD-based microscope camera, which also results in a lower SNR. Thus, mobile microscopes often produce relatively noisy images that are spectrally and spatially aberrated.

Recently, deep learning has been demonstrated to bridge this gap between mobile microscopes and their benchtop diffraction-limited counterparts [61]. Since the images taken by mobile microscopes can show fluctuations due to, e.g., sample misalignment, illumination nonuniformities, or diminishing battery power, input images need to be either normalized or be trained to learn the distribution of such fluctuations. If trained correctly, the network can learn how to mitigate these aberrations and create an image with similar quality to a benchtop microscope image from a low-cost, portable microscope, as illustrated in Fig. 7(a). This example demonstrates the power of the data-driven image enhancement, as a numerical modeling of the image formation process or the forward model, in this case, was intractable, due to aberrations, noise, and mechanical instability of the sample scanning module in the smartphone microscope (all in comparison to a high-end benchtop microscope).

As another example, cross-modality image transformations can be used to enhance the resolution of a microscopy modality beyond the diffraction limit. Some examples of this have been demonstrated by transforming diffraction-limited confocal microscopy images into stimulated emission depletion (STED) microscopy equivalent images [see Fig. 7(b)] [102] as well as by transforming total internal reflection fluorescence (TIRF) microscopy images [103] into TIRF-based structured illumination microscopy (TIRF-SIM) equivalent images [54], [104]. In both cases, the ground truth images used to train the neural network were obtained using a super-resolution microscopic imaging modality (STED and SIM, respectively) revealing fine features beyond the classical limit of diffraction. These transformations create images that match the corresponding images obtained with these advanced microscopy techniques, without the requirement of the extensive and expensive setups nor exposing the samples to increased doses of radiation causing adverse photobleaching and phototoxicity effects [47]. Computational super-resolution through such a cross-modality transformation also eliminates some of the required hardware, such as additional lasers, filters, and specialized fluorophores, which are often used for these techniques [105].

In addition to the lateral super-resolution and contrast enhancement shown earlier, deep networks performing cross-modality transformations can also be used to enhance the axial sectioning ability of the imaging system. For example, a deep neural network has been trained to take a single numerically backpropagated (refocused) hologram and output an image that is equivalent to a bright-field microscope image of the matching FOV, as demonstrated in Fig. 7(c) [106]. By performing this transformation, holographic images can match the corresponding images of the same samples imaged by a high-NA bright-field microscope, which allows the holographic imaging system to obtain the colorization and sectioning ability of a high-NA bright-field microscope while also eliminating coherence-related artifacts, such as speckle, twin image, and self-interference distortions. Since holographic images can be numerically backpropagated to different depths and focused upon different portions of the image, these different depths can be transformed into distinct bright-field equivalent images imaged at the desired sample depth. Using a series of consecutive depths that are digitally probed, a 3-D bright-field equivalent image can be created using an acquired hologram, which makes it ideal for applications where high-throughput imaging and screening of samples are needed. One such

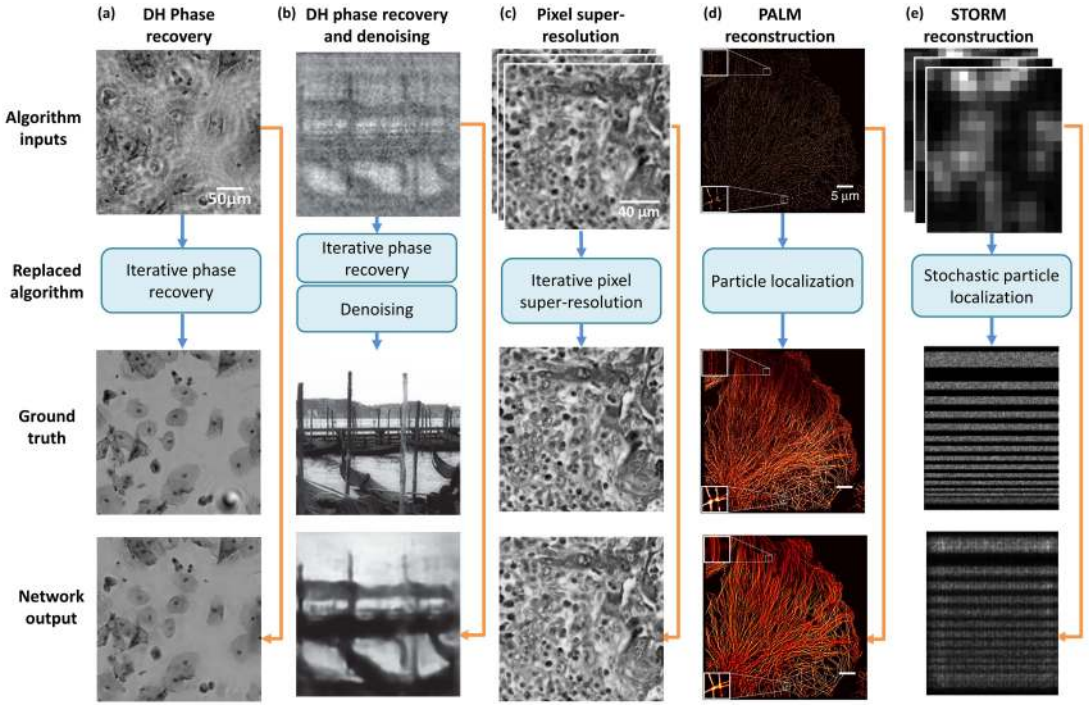

Fig. 8. ***Various image processing tasks performed by deep neural networks. Orange lines with arrows refer to the operation of each neural network (connecting the input image to its output). (a) In-line holography phase recovery. Input is a backpropagated (numerically refocused) hologram (missing the phase recovery step), and the ground truth is a phase recovered image. Images were taken from [55]. (b) In-line holographic phase recovery under low-light conditions. Input is a backpropagated hologram with significant noise, and the ground truth is a digital representation of the same image with high SNR. Reprinted with permission from Goy et al. [119]. Copyright Optical Society of America 2018. (c) Pixel super-resolution. Input is an image reconstruction from pixel-size limited low-resolution holograms, and the ground truth is a reconstruction from higher resolution pixel super-resolved holograms. Images were taken from [59]. (d) Artificial neural network accelerated PALM (ANNA-PALM). The input is an incomplete PALM image with fewer localizations, and the ground truth is a complete PALM image with more localizations. Reprinted with permission from Macmillan Publishers Ltd. from Ouyang et al. [60]. (e) Deep-STORM. Input is a simulated set of stochastic blinking frames, and the ground truth is a simulated reconstructed super-resolved image. Reprinted with permission from Nehme et al. [95]. Copyright Optical Society of America 2018. The neural network bypasses these complicated algorithms and saves both the number of required measurements and the computation time (marked by the yellow arrows, see Table I for details). In all cases, the output image is calculated by the trained network using the corresponding input (following the orange lines). DH: digital holography.***

application is the label-free imaging flow cytometer, where the throughput can be substantially expanded using cross-modality transformation-based image reconstructions that permit larger channel heights and samples volumes to be screened rapidly [107].

VIII. PHYSICAL MODEL-BASED IMAGE RECONSTRUCTION USING DEEP LEARNING

Sections V–VII have demonstrated deep networks that were trained with optically acquired inputs and image labels. However, for many imaging modalities, it is difficult or sometimes impossible to directly acquire input-ground truth image pairs, and instead, we rely on numerical tools to generate training data sets. Some examples of deep learning-based reconstruction methods that either used numerical tools or physics-based reconstruction methods to generate their gold-standard data or degraded their gold-standard data to generate simulated input data can be seen in Fig. 8.

Phase recovery and reconstruction of holographic images are the applications where this form of image transformations can be particularly useful [55]. The ground truth holographic images rely on coherent illumination to create a quantitative phase microscopy image of the specimen, which signifies the optical-path delay through the sample. In the optical part of the spectrum, optoelectronic sensors can only detect the intensity of the optical beam. Therefore, the classical way to obtain the object's field is to utilize an interference [108] between the object's wave and a reference wave (a) [108], that is

$$g = |a + L\{f\}|^2 = |a| + |L\{f\}|^2 + a(L\{f\})^* + a^* L\{f\} \tag{11}$$

where g is the intensity of the detected interference pattern. The image information is encoded in the expression $a^* L\{f\}$, where L is a linear operator (e.g., free space propagation operator [41]) that relates the wavefield f that is scattered off the object to the wavefield impinging on the optoelectronic sensor plane. The other terms in

this expression distort this image information and should ideally be removed in order to accurately reveal the object's image. One method to achieve this is to use off-axis holography [41], [109], with the tradeoff of reducing the space-bandwidth product of the imaging system. In an alternative holography approach (in-line holography), the scattered object field and the reference wave propagate along the same direction. In fact, as a result of its hardware simplicity, in-line holography is often preferred for various applications, especially for field measurements [110], [111]. Therefore, many methods have been developed to solve the "missing phase" problem and numerically remove the distortion terms from (11) using physical constraints [35]. This usually requires multiple hologram measurements to be made at, e.g., multiple sample-to-sensor distances, multiple angles of illumination, or multiple illumination wavelengths [112]–[115]. The need for this measurement diversity introduces constraints on the imaging hardware and its design and necessitates computationally intensive algorithms to accurately recover the signal.

As shown in Fig. 8(a), this hologram reconstruction process can be simplified by recovering both the phase and amplitude images of an object with only a single hologram measurement by using a deep neural network [55]. This makes both the imaging and reconstruction processes significantly faster than the iterative phase recovery-based methods [116], [117] that were used to generate the gold-standard images for the training phase. The problem of recovering the object's complex-valued image can be better constrained by performing free-space backpropagation (L^{-1}), which can be easily obtained by an autofocusing algorithm [118]. Thus, we can transform (11) into another form

$$g_L = a^* f + n(f) \tag{12}$$

where $g_L = L^{-1}\{g\}$ and $n(f)$ is the object-dependent noise [resulting from the remaining terms embedded in (11)]. This means that an alternative way of looking at holographic image reconstruction is to treat it as a sample-dependent denoising problem, for which the deep network is trained for using pairs of input–output images. This initial condition created by $L^{-1}\{g\}$ operation is used to remove input ambiguities and foster high accuracy object recovery, even for highly connected samples, such as tissue slides [55], as demonstrated in Fig. 8(a). It is interesting to note that the network also learns to reject out-of-focus particles even though they were physically located in the imaging path. This is very beneficial as the network learns to reject objects that are part of the hologram formation but not in the sample plane/volume. One application of this technique that made use of this feature is a portable imaging flow cytometer [107], where deep neural networks enabled reconstruction of holographic images taken by the cytometer for high-throughput testing of water samples in real time using a GPU-equipped laptop.

In addition to reconstructing these holograms, deep learning has been shown to improve the phase retrieval when the images are taken in low-light situations [119], [120]. In this case, a deep neural network can significantly reduce speckle noise and, therefore, improve the SNR of the reconstructed image, as demonstrated in Fig. 8(b).

Enhancing images that have their resolution limited by the detector's pixel size is not common in lens-based microscopic imaging since it is quite easy to add magnification in the light collection path (typically at the cost of a reduced FOV). However, it is one of the limiting factors in lensless holographic microscopy as the FOV is as large as the sensor size (typically dozens of mm^2), which enables high-throughput imaging. To get around this pixel-size limited resolution in lensless holographic microscopy, different super-resolution techniques have been applied [35], [36], [121], most of them are based on subpixel shift integration. Deep learning has also been demonstrated to efficiently enhance the resolution of lensless microscopy systems using significantly fewer images compared to these earlier pixel super-resolution methods, which further increases the throughput of these lensless systems and relaxes some of their hardware design constraints [see Fig. 8(c)] [59].

Beyond holography, deep learning has been shown to be also effective at reconstructing single-molecule localization images. Techniques such as photoactivated localization microscopy (PALM) [45] and stochastic optical reconstruction microscopy (STORM) [46] are able to accurately determine the positions of individual molecules within a sample. These techniques work by taking many images in which only a random subset of the sample's fluorophores can be emissive in each frame. If these molecules are sufficiently sparse and enough photons per molecule are collected, the centroid of each molecule can be fit to determine its location. When thousands of these images are taken together, they can be used to reconstruct a full image of the sample. However, since many images are needed, the imaging process is usually time consuming. By using deep learning, both PALM [referred to as ANNA-PALM, demonstrated in Fig. 8(d)] [60] and STORM [referred to as Deep-STORM, demonstrated in Fig. 8(e)] [95] techniques can be performed significantly faster without compromising spatial resolution. Both of these techniques were trained using a set of simulated ground truth image data.

Similar to the above-mentioned examples, the Fourier ptychographic microscopy images have also been reconstructed using deep neural networks. Using deep learning, the main benefits of the Fourier phytographic microscopy were achieved (extending the space-bandwidth product of a microscope objective lens) while reducing the number of images required for the reconstruction by sixfold and decreasing the processing time by 50-fold [122]. This enabled the reconstruction of large-scale spatial and temporal information quickly, which is important for imaging of live cells.

A summary of some of these deep learning-powered reconstruction methods and the time that is saved for both imaging and inference are presented in Table 1.

Table 1 Data (Number of Measurements) and Runtime (Computation) Efficiency Improvement Enabled by Image Processing Algorithm Abstraction by Deep Neural Networks

	Computational task	Conventional method		Deep learning-based method		Reported efficiency improvement	
		Number of captured images	*Runtime*	*Number of captured images*	*Runtime*	*Number of captured images*	*Runtime*
Rivenson et al. (2018)[55]	DH phase recovery	2 – 8 holograms	30 – 50 s (2048^2 pixels)	Single hologram	7 s (2048^2 pixels)	2 – 3 ×	4 – 7 ×
Wu et al. (2018)[58]	Phase recovery, Auto-focusing	2 – 8 holograms	3 - 6 s (512^2 pixels)	Single hologram	0.1 s (512^2 pixels)	2 – 3 ×	30 - 60 ×
Liu et al. (2018)[59]	Pixel super-resolution	36 raw holograms	30 – 60 s (1940^2 pixels)	1 – 9 raw holograms	1 s (1940^2 pixels)	4 – 36 ×	30 - 60 ×
Ouyang et al. (2018)[60]	PALM image reconstruction	30,000 – 60,000 localizations	Not specified	300 – 800 localizations	Not specified	100 ×	Not specified

Table 1 indicates that for many reconstruction methods, inference through the neural network is significantly faster than the traditional reconstruction algorithms and much fewer data are required.

IX. *IN SILICO* LABELING USING DEEP LEARNING

In addition to improving the quality of images, deep learning can also perform many other image transformations, including virtually replicating physical changes made to a sample using, e.g., specific sample preparation processes. In this research direction, one of the most interesting applications for deep learning is the replacement of labeling techniques that are used to add contrast to tissue samples, which, otherwise, have almost no interpretable contrast in many biomedical applications.

Virtual staining of histological tissue sections used to diagnose various diseases is one such method [123]. Tissue sections are either frozen or embedded in paraffin and then sectioned into thin (typically 2–5 μm) slices that are mounted onto microscope slides and imaged after a staining/labeling process. They use the cellular and subcellular chemical environments to bind chromophores or fluorophores that, in turn, introduce exogenous contrast to different tissue constituents. This process is time consuming, which delays any diagnosis and may result in anxiety for the patient, puts financial stress on the health care system, requires trained staff and chemical reagents, and does not support tissue preservation for advanced molecular analysis. For most cases, the tissue section is imaged using a standard bright-field microscope following its histological staining (see Fig. 9).

Over the last few decades, researchers have designed new imaging methods to introduce contrast to these tissue sections. These techniques use rapid labeling or no labeling at all, where the latter uses endogenous contrast agents that are naturally embedded in the tissue section. Following that, linearized numerical approximations have been used to map the obtained contrast distribution to a virtual image of the hematoxylin and eosin (H&E) stain, which is a standard stain used in diagnosis, specifically for cancer screening and tumor margin estimation [124], [125].

Recently, deep learning has been demonstrated to be an effective virtual staining method. The success of this virtual staining has been illustrated using several different microscopy modalities as input images [49], [51], [126], [127]. Rivenson *et al.* [49] demonstrated the efficacy of the deep learning-based staining using images of a single autofluorescence channel of an unstained tissue and that the technique can be applied to several different tissue types and stains (Masson's Trichrome and Jones stain in addition to the standard H&E). Furthermore, this study was able to validate the technique using a panel of pathologists who determined that the quality of the virtual stain was equivalent to that of a histologically stained slide and that diagnoses could be made accurately using a virtually stained slide. Examples of these virtually stained whole slide images can be seen in Fig. 9(d) and (e). It was also demonstrated that additional excitation channels or imaging modalities can also be used to increase the staining quality, in cases where some of the tissue constituents do not provide meaningful contrast at a single band [49], [127]. These techniques can also be used with other contrasts introducing imaging techniques, such as optical coherence tomography (OCT) [128], [129], quantitative phase imaging [51], the Raman microscopy [130], [131], as well as other rapid staining methods [124], [132], [133].

Deep learning can even improve upon the quality of physical stains; for example, neural networks create more consistent staining than standard histochemical

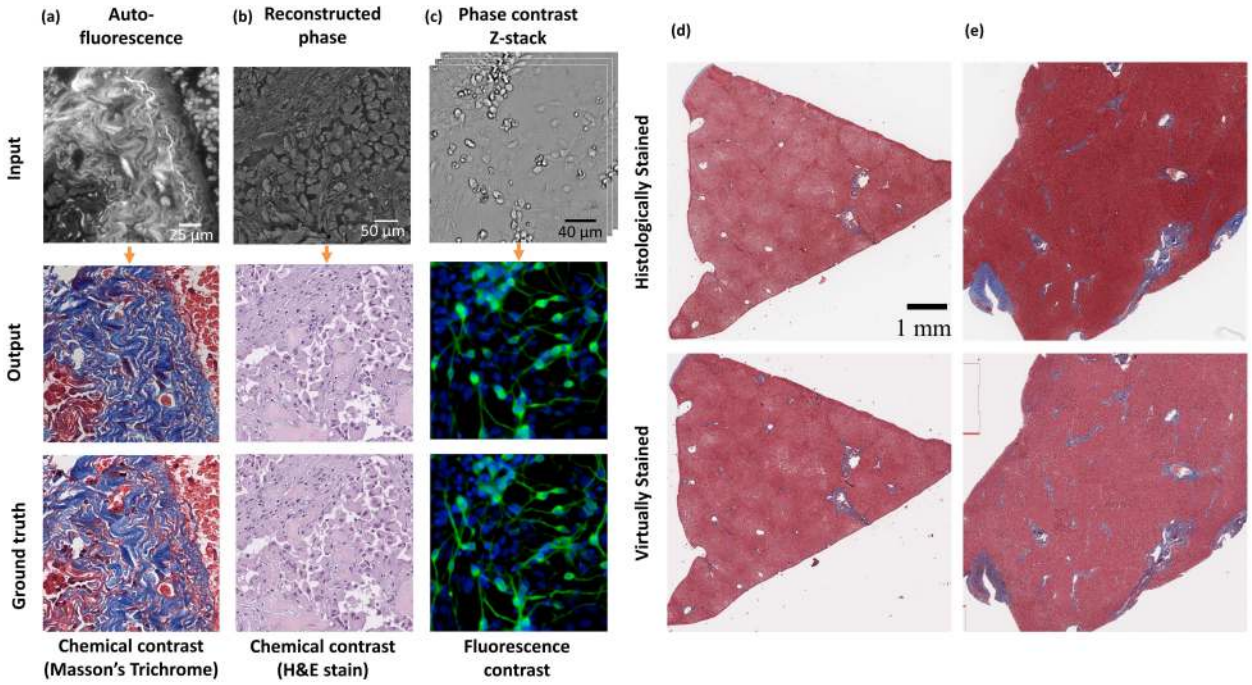

Fig. 9. ***Transforming microscopy images between different contrast mechanisms. (a) Label-free autofluorescence image to histologically stained bright-field equivalent image. Images were taken from [49]. (b) Label-free quantitative phase image to histologically stained bright-field equivalent image. Images were taken from [51]. (c) Label-free phase-contrast image z-stack (covering 13 axial steps) to a fluorescently labeled image. Reprinted with permission from Christiansen et al. [52]. (d) and (e) Examples of whole slide image transformations that have been virtually stained with Masson's Trichrome using a deep neural network. Histologically stained versions of the same samples, imaged under bright-field microscopy, are also provided for comparison.***

staining, which exhibits lab-to-lab and technician-to-technician variations. An example of this stain quality normalization or standardization can be seen in Fig. 9(d) and (e). This feature of deep learning-based virtual staining can increase the diagnostic accuracy of downstream analysis of histological images inspected by humans and/or machines [134].

Another *in silico* labeling technique that deep learning has successfully reproduced is the prediction of fluorescently tagged labels using unlabeled images [52]. An example of this is shown in Fig. 9(c), where Christiansen *et al.* [52] acquired different forms of transmission light microscopy images, such as bright field and phase contrast, forming an axial image stack used as input to a deep neural network to segment the images based on the expected fluorescence tags. By performing these transformations, they were able to detect specific markers without needing to go through the potentially lengthy and costly process of external cell labeling.

X. DEEP LEARNING ENABLES NEW IMAGING MODALITIES

Deep learning can also be used to add new capabilities to microscopic imaging techniques, achieving results that cannot be directly obtained through current instrumentation or physics-based forward models or assumptions. An example of this can be found in holographic imaging, where the DOF can be significantly increased using deep learning [58]. Using this method, different object features within the DOF can be refocused, all in parallel, with a single feed-forward pass through a trained network to get a result that cannot be obtained using standard holographic reconstruction methods (see Fig. 10).

Stated differently, in this case, there is no equivalent ground truth image that can be used as a label since this effect cannot be achieved using traditional holographic reconstruction methods or algorithms. In addition to the benefits of the increased DOF, parallelization makes the holographic reconstruction significantly faster as well as robust to any change in focus; computational complexity of the reconstruction is also drastically simplified since autofocusing and phase recovery steps are merged into the same computational step, and all the parts of the sample volume are brought into focus simultaneously, i.e., without the need for a separate autofocusing step for different regions of interest (see Fig. 10). All these properties have already made the algorithm useful for applications, such as particle aggregation-based sensing of viruses [135] and automatic detection/counting of bioaerosols [136], in which a trained network was used to improve both the quality of the autofocusing and phase recovery.

In one of the earlier works that applied neural networks to microscopic imaging [137], it was shown that a network can perform 3-D tomographic inference from a set of

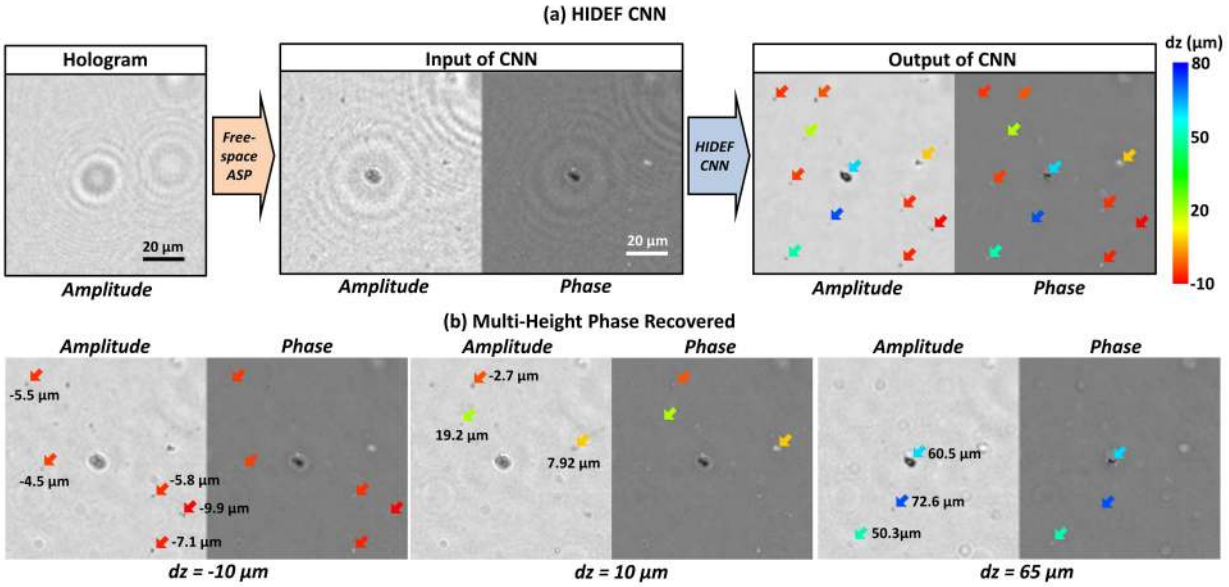

Fig. 10. ***Holographic imaging of aerosols using deep learning with EDOF (HIDEF). (a) Workflow for the HIDEF network. A raw hologram is backpropagated and input into the HIDEF network. (b) Same FOV reconstructed using multiheight phase recovery. Modified from [58]. Copyright Optical Society of America 2018.***

2-D projections. This is a classic problem in optical imaging and has some differences from standard tomographic reconstruction techniques used in, e.g., computed tomography (CT). In many cases, however, when performing tomographic reconstructions in the optical regime, any scattering feature in the sample is often approximated to introduce only a single scattering event; stated differently, different parts of the sample volume are assumed to have the same illumination pattern, independent of the object. This simplification allows us to get an analytical solution based on the first Born approximation [138] and then apply standard filtered backprojection (FBP) techniques for 3-D image reconstruction. By using a neural network, this approximation is no longer necessary as the network is constructed to learn the refractive indices of the specimen as bias terms, and the weights of the network were fixed and given by a numerical expression of the free space propagation between the different specimen planes.

These recent results and others [62] show how deep networks enable novel image reconstruction and transformation methods that are not possible with current imaging models or instrumentation. Some of these networks defy classical inverse problem solutions and reduce the need for inaccurate model approximations that have been made for the sake of making the numerical inversion tractable.

XI. CONCLUSION

In this article, we discussed a variety of ways that deep learning has been used to transform and reconstruct biomedical images in optical microscopy. These techniques can be used with various systems and optical modalities to either improve, reconstruct, or transform images and have the potential to revolutionize the way that microscopic imaging is done for medical and clinical applications. Systems that perform image super-resolution can reduce the cost of imaging systems and enable new samples to be characterized. Other techniques, such as virtual staining, can be used to improve the speed and accuracy of diagnoses using histopathology.

Microscopy is an ideal field to apply deep learning, as experimental data sets are typically obtained under highly controlled conditions. The user is often a trained technician/professional who can ensure that the sample is prepared correctly and that the variables, such as the illumination intensity and focus, are well tuned. Microscopy instruments are also highly repeatable in terms of illumination light properties and sample positioning, which is controlled with precise 3-D stages. This is a stark contrast to many consumer applications of deep learning involving macroscale imaging using cameras, where the imaging systems are used by untrained operators in highly diverse, uncontrolled, or poor imaging conditions.

However, there are also limitations to these emerging deep learning-based microscopy approaches presented here. Most notably, it can be challenging to create a matching data set to train these networks if the sample is dynamic and rapidly changing because a coregistered training data set cannot be created easily. Even for static samples, data set creation can be time consuming or expensive when a sample-specific data set is required. The performance of these networks also depends on the quality and breadth of the data used during the training phase. As such, the methods described here cannot be performed in a reliable manner without this high-quality data being available.

While the majority of the work to date has been performed to improve existing imaging modalities, computational microscopy is no longer limited to imaging systems that rely on constraints, such as the ability to numerically formulate a sensing model [62]. Instead, deep learning can be used to transform and analyze the images, without the need for forward model approximations. For example, even though physical models for reconstructing images that have passed through a multimode fiber exist only for specific and restrictive cases, the actual transformation and its reversal can be learned through image data and deep learning [63]. Such an imaging system that is able to reconstruct the images that passed through a fiber optic cable has the potential to be used as a thin, flexible endoscope [139].

One of the most interesting directions for the use of big data in computational imaging is to codesign the sensing system along with the inference algorithm. Deep learning-inspired imaging instrumentation can be used to engineer new types of optics for specific tasks, such as imaging or classification [140]. Alternatively, it could be used to design optical components and imaging tactics with an optical front end and an algorithmic back end [141]. This could be used to resolve the optimally encoded [142]–[145] patterns for image reconstruction or the extraction of a meaningful metric [141], [146]. This kind of holistic design is considered as the holy grail of computational imaging. We believe that in the future, new optical architectures and deep networks might fully unlock the potentials of task-specific microscopy and help us design low-cost and/or high-throughput imaging and sensing modalities. Along these lines, one of the more interesting directions is to create a "thinking" imaging system [147], which can decide what measurement should be taken next, based on previous data rather than *a priori* design. This has been recently shown for an unsupervised MRI sensing [148], and similar ideas will likely find their way to microscopic imaging. ■

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[2] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [review article]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

[5] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 1–9.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[7] D. Silver *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[8] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[12] M. Abadi *et al.* "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," Mar. 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467

[13] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. Adv. Neural Inf. Process. Syst. Workshop*, 2017, pp. 1–2.

[14] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," Jun. 2014, *arXiv:1408.5093*. [Online]. Available: https://arxiv.org/abs/1408.5093

[15] The Theano Development Team *et al.*, "Theano: A Python framework for fast computation of mathematical expressions," May 2016, *arXiv:1605.02688*. [Online]. Available: https://arxiv.org/abs/1605.02688

[16] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A MATLAB-like environment for machine learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011.

[17] Y. Jo *et al.*, "Holographic deep learning for rapid optical screening of anthrax spores," *Sci. Adv.*, vol. 3, no. 8, p. e1700606, Aug. 2017.

[18] Y. Liu *et al.*, "Artificial intelligence-based breast cancer nodal metastasis detection: Insights into the black box for pathologists," *Arch. Pathol. Lab. Med.*, vol. 143, no. 7, pp. 859–868, Oct. 2018.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. New York, NY, USA: Springer, 2015, pp. 234–241.

[20] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1182–1189.

[21] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar. 2002.

[22] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 349–356.

[23] C.-Y. Yang, C. Ma, and M.-H. Yang, "Single-image super-resolution: A benchmark," in *Proc. ECCV*, 2014. pp. 372–386.

[24] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, p. 55, 1972.

[25] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, no. 6, p. 745, 1974.

[26] T. J. Holmes, "Maximum-likelihood image restoration adapted for noncoherent optical imaging," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 5, no. 5, pp. 666–673, May 1988.

[27] J.-A. Conchello and E. W. Hansen, "Enhanced 3-D reconstruction from confocal scanning microscope images. 1: Deterministic and maximum likelihood reconstructions," *Appl. Opt.*, vol. 29, no. 26, pp. 3795–3804, Sep. 1990.

[28] J.-B. Sibarita, "Deconvolution microscopy," in *Microscopy Techniques*, vol. 95, J. Rietdorf, Ed. Berlin, Germany: Springer, 2005, pp. 201–243.

[29] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision—ECCV*. New York, NY, USA: Springer, 2016, pp. 694–711.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[31] D. B. Murphy, *Fundamentals of Light Microscopy and Electronic Imaging*. Hoboken, NJ, USA: Wiley, 2002.

[32] Nikon's MicroscopyU. *Introduction to Fluorescence Microscopy*. Accessed: Mar. 19, 2019. [Online]. Available: https://www.microscopyu.com/techniques/fluorescence/introduction-to-fluorescence-microscopy

[33] E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," *Appl. Opt.*, vol. 38, no. 34, pp. 6994–7001, Dec. 1999.

[34] G. Popescu, *Quantitative Phase Imaging of Cells and Tissues*. New York, NY, USA: McGraw-Hill, 2011.

[35] A. Greenbaum *et al.*, "Imaging without lenses: Achievements and remaining challenges of wide-field on-chip microscopy," *Nature Methods*, vol. 9, no. 9, pp. 889–895, Sep. 2012.

[36] A. Greenbaum *et al.*, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," *Sci. Transl. Med.*, vol. 6, no. 267, p. 267ra175, Dec. 2014.

[37] P. Boccacci and M. Bertero, *Introduction to Inverse Problems in Imaging*. New York, NY, USA: Taylor & Francis, 1998.

[38] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA, USA: SIAM, 2002.

[39] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[40] D. L. Donoho, "Superresolution via sparsity constraints," *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309–1331, Sep. 1992.

[41] J. W. Goodman, *Introduction to Fourier Optics*. Englewood, CO, USA: Roberts and Company Publishers, 2005.

[42] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 15, pp. 8206–8210, Jul. 2000.

[43] M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited

resolution," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 37, pp. 13081–13086, Sep. 2005.
[44] B. Huang, M. Bates, and X. Zhuang, "Super resolution fluorescence microscopy," *Annu. Rev. Biochem.*, vol. 78, pp. 993–1016, Mar. 2009.
[45] E. Betzig *et al.*, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, Sep. 2006.
[46] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–796, Oct. 2006.
[47] S. Wäldchen, J. Lehmann, T. Klein, S. van de Linde, and M. Sauer, "Light-induced cell damage in live-cell super-resolution microscopy," *Sci. Rep.*, vol. 5, p. 15348, Oct. 2015.
[48] *ZEISS Microscopy Online Campus|Practical Aspects of Photoactivated Localization Microscopy (PALM)*. Accessed: Mar. 9, 2019. [Online]. Available: http://zeiss-campus.magnet.fsu.edu/print/superresolution/palm/practicalaspects-print.html
[49] Y. Rivenson *et al.*, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nature Biomed. Eng.*, vol. 3, pp. 466–477, Mar. 2019.
[50] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," *Nature Methods*, vol. 15, no. 11, p. 917, Nov. 2018.
[51] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: The digital staining of label-free quantitative phase microscopy images using deep learning," *Light, Sci. Appl.*, vol. 8, no. 1, p. 23, Feb. 2019.
[52] E. M. Christiansen *et al.*, "*In silico* labeling: Predicting fluorescent labels in unlabeled images," *Cell*, vol. 173, no. 3, pp. 792–803, Apr. 2018.
[53] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, vol. 4, no. 11, pp. 1437–1443, Nov. 2017.
[54] H. Wang *et al.*, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, pp. 103–110, Jan. 2019.
[55] Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light, Sci. Appl.*, vol. 7, p. 17141, Feb. 2018.
[56] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, pp. 1117–1125, Sep. 2017.
[57] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proc. Int. Joint Conf. Neural Netw.*, vol. 1, 1989, pp. 593–605.
[58] Y. Wu *et al.*, "Extended depth-of-field in holographic image reconstruction using deep learning based auto-focusing and phase-recovery," *Optica*, vol. 5, no. 6, p. 704, Jun. 2018.
[59] T. Liu *et al.*, "Deep learning-based super-resolution in coherent imaging systems," *Sci. Rep.*, vol. 9, no. 1, p. 3926, Mar. 2019.
[60] W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nature Biotechnol.*, vol. 36, no. 5, pp. 460–468, Apr. 2018.
[61] Y. Rivenson *et al.*, "Deep learning enhanced mobile-phone microscopy," *ACS Photon.*, vol. 5, no. 6, pp. 2354–2364, Jun. 2018.
[62] Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, and A. Ozcan, "Three-dimensional propagation and time-reversal of fluorescence images," Jan. 2019, *arXiv:1901.11252*. [Online]. Available: https://arxiv.org/abs/1901.11252
[63] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," *Light, Sci. Appl.*, vol. 7, no. 1, p. 69, Oct. 2018.
[64] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
[65] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. USA*, 2010, pp. 807–814.
[66] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[67] D. Pedamonti, "Comparison of non-linear activation functions for deep neural networks on MNIST classification task," Apr. 2018, *arXiv:1804.02763*. [Online]. Available: https://arxiv.org/abs/1804.02763
[68] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.
[69] B. Neyshabur, S. Bhojanapalli, D. Mcallester, and N. Srebro, "Exploring generalization in deep learning," in *Advances in Neural Information Processing Systems*, I. Guyon, Eds. *et al.* Red Hook, NY, USA: Curran Associates, 2017, pp. 5947–5956.
[70] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2016, pp. 4898–4906.
[71] K. O. Stanley and R. Miikkulainen, "Evolving neural networks through augmenting topologies," *Evol. Comput.*, vol. 10, no. 2, pp. 99–127, Jun. 2002.
[72] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 2951–2959.
[73] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–16.
[74] T. Falk *et al.*, "U-net: Deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, p. 67, Jan. 2019.
[75] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[76] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 2672–2680.
[77] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Cambridge, MA, USA, vol. 2, 2014, pp. 3320–3328.
[78] A. A. Goshtasby, *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. New York, NY, USA: Wiley, 2005.
[79] H. Sokooti, B. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. Med. Image Comput. Comput. Assist. Intervent (MICCAI)*, 2017, pp. 232–239.
[80] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1075–1078.
[81] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration: A review," *Comput. Methods Biomech. Biomed. Eng.*, vol. 17, no. 2, pp. 73–93, 2014.
[82] W. R. Crum, T. Hartkens, and D. L. G. Hill, "Non-rigid image registration: Theory and practice," *Brit. J. Radiol.*, vol. 77, no. 2, pp. S140–S153, 2004.
[83] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
[84] M. Olson, A. Wyner, and R. Berk, "Modern neural networks generalize on small data sets," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2018, pp. 3619–3628.
[85] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 759–766.
[86] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.
[87] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," Dec. 2017, *arXiv:1712.04621*. [Online]. Available: https://arxiv.org/abs/1712.04621
[88] S. Culley *et al.*, "Quantitative mapping and minimization of super-resolution optical imaging artifacts," *Nature Methods*, vol. 15, no. 4, pp. 263–266, Apr. 2018.
[89] L. Prechelt, "Early stopping—But when?" in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, K.-R. Müller, Eds., 2nd ed. Berlin, Germany: Springer, 2012, pp. 53–67.
[90] I. J. Cox and C. J. R. Sheppard, "Information capacity and resolution in an optical system," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 3, no. 8, pp. 1152–1158, Aug. 1986.
[91] Y. Katznelson, *An Introduction to Harmonic Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[92] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and Robust Automated Machine Learning," in *Advances in Neural Information Processing Systems*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2015, pp. 2962–2970.
[93] Y. He, J. Lin, Z. Liu, H. Wang, L.-J. Li, and S. Han, "AMC: AutoML for model compression and acceleration on mobile devices," Feb. 2018, *arXiv:1802.03494*. [Online]. Available: https://arxiv.org/abs/1802.03494
[94] Nikon's MicroscopyU. *Depth of Field and Depth of Focus*. Accessed: Mar. 8, 2019. [Online]. Available: https://www.microscopyu.com/microscopy-basics/depth-of-field-and-depth-of-focus
[95] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, Apr. 2018.
[96] H. Zhang *et al.*, "High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network," *Biomed. Opt. Express*, vol. 10, no. 3, pp. 1044–1063, Mar. 2019.
[97] K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, "Resolution enhancement in scanning electron microscopy using deep learning," Jan. 2019, *arXiv:1901.11094*. [Online]. Available: https://arxiv.org/abs/1901.11094
[98] D. N. Breslauer, R. N. Maamari, N. A. Switz, W. A. Lam, and D. A. Fletcher, "Mobile phone based clinical microscopy for global health applications," *PLoS ONE*, vol. 4, no. 7, p. e6320, Jul. 2009.
[99] A. Ozcan, "Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools," *Lab. Chip*, vol. 14, no. 17, pp. 3187–3194, Sep. 2014.
[100] D. Tseng *et al.*, "Lensfree microscopy on a cellphone," *Lab Chip*, vol. 10, no. 14, pp. 1787–1792, Jul. 2010.
[101] Y. Lu, W. Shi, J. Qin, and B. Lin, "Low cost, portable detection of gold nanoparticle-labeled microfluidic immunoassay with camera cell phone," *Electrophoresis*, vol. 30, no. 4, pp. 579–582, Feb. 2009.
[102] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.*, vol. 19, no. 11, pp. 780–782, Jun. 1994.
[103] D. Axelrod, "Cell-substrate contacts illuminated by total internal reflection fluorescence," *J. Cell Biol.*,

vol. 89, no. 1, pp. 141–145, Apr. 1981.
[104] D. Li *et al.*, "Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics," *Science*, vol. 349, no. 6251, p. aab3500, Aug. 2015.
[105] S. W. Hell, "Far-field optical nanoscopy," *Science*, vol. 316, no. 5828, pp. 1153–1158, May 2007.
[106] Y. Wu *et al.*, "Bright-field holography: Cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," *Light, Sci. Appl.*, vol. 8, no. 1, p. 25, Mar. 2019.
[107] Z. Göröcs *et al.*, "A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples," *Light, Sci. Appl.*, vol. 7, no. 1, p. 66, Sep. 2018.
[108] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, no. 4098, pp. 777–778, May 1948.
[109] E. N. Leith and J. Upatnieks, "Holographic imagery through diffusing media," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 56, no. 4, p. 523, Apr. 1966.
[110] E. McLeod and A. Ozcan, "Unconventional methods of imaging: Computational microscopy and compact implementations," *Rep. Progr. Phys.*, vol. 79, no. 7, p. 076001, May 2016.
[111] Y. Wu and A. Ozcan, "Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring," *Methods*, vol. 136, pp. 4–16, Mar. 2018.
[112] R. W. Gerchberg and A. W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237–250, 1971.
[113] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.*, vol. 3, no. 1, pp. 27–29, Jul. 1978.
[114] Z. Zalevsky, D. Mendlovic, and R. G. Dorsch, "Gerchberg–Saxton algorithm applied in the fractional Fourier or the Fresnel domain," *Opt. Lett.*, vol. 21, no. 12, pp. 842–844, Jun. 1996.
[115] V. Elser, "Solution of the crystallographic phase problem by iterated projections," *Acta Crystallogr.*, vol. 59, no. 3, pp. 201–209, May 2003.
[116] D. L. Misell, "An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations," *J. Phys. D, Appl. Phys.*, vol. 6, no. 18, pp. 2200–2216, Dec. 1973.
[117] A. Greenbaum and A. Ozcan, "Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy," *Opt. Express*, vol. 20, no. 3, pp. 3129–3143, Jan. 2012.
[118] Y. Zhang, H. Wang, Y. Wu, M. Tamamitsu, and A. Ozcan, "Edge sparsity criterion for robust holographic autofocusing," *Opt. Lett.*, vol. 42, no. 19, pp. 3824–3827, Oct. 2017.
[119] A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Phys. Rev. Lett.*, vol. 121, no. 24, p. 243902, Dec. 2018.
[120] W. Jeon, W. Jeong, K. Son, and H. Yang, "Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks," *Opt. Lett.*, vol. 43, no. 17, pp. 4240–4243, Sep. 2018.
[121] W. Bishara, T.-W. Su, A. F. Coskun, and A. Ozcan, "Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution," *Opt. Express*, vol. 18, no. 11, pp. 11181–11191, May 2010.
[122] T. Nguyen, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, "Deep learning approach for Fourier ptychography microscopy," *Opt. Express*, vol. 26, no. 20, pp. 26470–26484, Oct. 2018.
[123] R. Rubin, D. S. Strayer, E. Rubin, and J. M. McDonald, *Rubin's Pathology: Clinicopathologic Foundations of Medicine*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2008.
[124] Y. K. Tao *et al.*, "Assessment of breast pathologies using nonlinear microscopy," *Proc. Nat. Acad Sci. USA*, vol. 111, no. 43, pp. 15304–15309, Oct. 2014.
[125] M. G. Giacomelli *et al.*, "Virtual hematoxylin and eosin transillumination microscopy using Epi-fluorescence imaging," *PLoS ONE*, vol. 11, no. 8, p. e0159337, Aug. 2016.
[126] A. Rana, G. Yauney, A. Lowe, and P. Shah, "Computational histological staining and destaining of prostate core biopsy RGB images with generative adversarial neural networks," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 828–834.
[127] N. Borhani, A. J. Bower, S. A. Boppart, and D. Psaltis, "Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy," *Biomed. Opt. Express*, vol. 10, no. 3, pp. 1339–1350, Mar. 2019.
[128] B. J. Vakoc *et al.*, "Three-dimensional microscopy of the tumor microenvironment *in vivo* using optical frequency domain imaging," *Nat. Med.*, vol. 15, no. 10, pp. 1219–1223, Oct. 2009.
[129] E. Dalimier and D. Salomon, "Full-field optical coherence tomography: A new technology for 3D high-resolution skin imaging," *Dermatology*, vol. 224, no. 1, pp. 84–92, 2012.
[130] F.-K. Lu *et al.*, "Label-free DNA imaging *in vivo* with stimulated Raman scattering microscopy," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 37, pp. 11624–11629, Sep. 2015.
[131] D. A. Orringer *et al.*, "Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy," *Nat. Biomed. Eng.*, vol. 1, Feb. 2017, Art. no. 0027.
[132] F. Fereidouni *et al.*, "Microscopy with ultraviolet surface excitation for rapid slide-free histology," *Nature Biomed. Eng.*, vol. 1, pp. 957–966, Dec. 2017.
[133] A. K. Glaser *et al.*, "Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens," *Nat. Biomed. Eng.*, vol. 1, no. 7, p. 0084, Jul. 2017.
[134] Y. Liu *et al.*, "Detecting cancer metastases on gigapixel pathology images," Mar. 2017. *arXiv:1703.02442*. [Online]. Available: https://arxiv.org/abs/1703.02442
[135] Y. Wu *et al.*, "Deep learning enables high-throughput analysis of particle-aggregation-based biosensors imaged using holography," *ACS Photon.*, vol. 6, no. 2, pp. 294–301, Feb. 2019.
[136] Y. Wu *et al.*, "Label-free bioaerosol sensing using mobile microscopy and deep learning," *ACS Photon.*, Oct. 2018.
[137] U. S. Kamilov *et al.*, "Learning approach to optical tomography," *Optica*, vol. 2, no. 6, pp. 517–522, Jun. 2015.
[138] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Philadelphia, PA, USA: SIAM, 2001.
[139] F. Koenig, J. Knittel, and H. Stepp, "Diagnosing cancer *in vivo*," *Science*, vol. 292, no. 5520, pp. 1401–1403, May 2001.
[140] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, Sep. 2018.
[141] D. Mengu, Y. Luo, Y. Rivenson, and A. Ozcan, "Analysis of diffractive optical neural networks and their integration with electronic neural networks," Oct. 2018, *arXiv:1810.01916v2*. [Online]. Available: https://arxiv.org/abs/1810.01916
[142] S. Elmalem, R. Giryes, and E. Marom, "Learned phase coded aperture for the benefit of depth of field extension," *Opt. Express*, vol. 26, no. 12, pp. 15316–15331, Jun. 2018.
[143] A. Robey and V. Ganapati, "Optimal physical preprocessing for example-based super-resolution," *Opt. Express*, vol. 26, no. 24, pp. 31333–31350, Nov. 2018.
[144] M. R. Kellman, E. Bostan, N. Repina, and L. Waller, "Physics-based learned design: Optimized coded-illumination for quantitative phase imaging," Aug. 2018, *arXiv:1808.03571*. [Online]. Available: https://arxiv.org/abs/1808.03571
[145] E. Hershko, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Multicolor localization microscopy and point-spread-function engineering by deep learning," *Opt. Express*, vol. 27, no. 5, pp. 6158–6183, Mar. 2019.
[146] J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," *Sci. Rep.*, vol. 8, no. 1, p. 12324, Aug. 2018.
[147] Y. Rivenson and A. Ozcan, "Toward a thinking microscope," *Opt. Photon. News*, vol. 29, no. 7, pp. 34–41, Jul. 2018.
[148] K. H. Jin, M. Unser, and K. M. Yi, "Self-supervised deep active accelerated MRI," Jan. 2019, *arXiv:1901.04547*. [Online]. Available: https://arxiv.org/abs/1901.04547

ABOUT THE AUTHORS

Kevin de Haan (Member, IEEE) received the B.A.Sc. degree in engineering physics from Queen's University, Kingston, ON, Canada, in 2015, and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2017. He is currently working toward the Ph.D. degree at the University of California at Los Angeles (UCLA), Los Angeles, CA, USA.

His current research interests include developing new data sets and developing networks to perform image transformations with deep learning.

Yair Rivenson (Member, IEEE) was born in Be'er Sheva, Israel, in 1981. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering and computer engineering from the Ben-Gurion University of the Negev, Be'er Sheva, in 2003, 2009, and 2014, respectively.

In 2013, he joined the Engineering Faculty, Bar-Ilan University, Ramat Gan, Israel, as a Postdoctoral Scholar. In 2015, he joined the Electro-Optics Unit, Ben-Gurion University of the Negev, as a Lecturer. From 2015 to 2018, he was a Marie Sklodowska-Curie Postdoctoral Fellow with

Bar-Ilan University and the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, where he has been an Assistant Adjunct Professor with the Department of Electrical and Computer Engineering since 2019. His current research interests include computational imaging and sensing for biomedical, life sciences and environmental applications, and physics-inspired machine learning.

Dr. Rivenson is a member of the Optical Society of America (OSA), the Society of Photographic Instrumentation Engineers (SPIE), and the Digital Pathology Association (DPA) and an Associate Member of the IEEE Bio Imaging and Signal Processing (BISP) Technical Committee. His research in the field of computational imaging has received numerous awards and distinctions. In 2018, he received the OSA Outstanding Reviewer of Year Award.

Yichen Wu received the B.S.E. degree in information engineering from Zhejiang University, Hangzhou, China, in 2014. He is currently working toward the Ph.D. degree at the Bio-Photonics Laboratory led by Prof. Aydogan Ozcan at the Electrical and Computer Engineering Department, University of California at Los Angeles (UCLA), Los Angeles, CA, USA.

His previous research contributions include machine learning, digital holography, environmental sensing, nanophotonics, and spectroscopy. His current research interests include computational imaging and machine learning for volumetric and fast biomedical imaging applications.

Mr. Wu was a recipient of the Vodafone Americas Wireless Innovation Award, the SPIE John Kiel Scholarship, and the UCLA EE Departmental Fellowship. He has served as the President of the OSA/SPIE Student Chapter at UCLA.

Aydogan Ozcan (Fellow, IEEE) is currently the Chancellor's Professor with the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, an HHMI Professor with the Howard Hughes Medical Institute, leading the Bio- and Nano-Photonics Laboratory, UCLA School of Engineering, and the Associate Director of the California NanoSystems Institute. He is also the Founder and a member of the Board of Directors of Lucendi, Inc., Los Angeles, and Holomic/Cellmic LLC, Los Angeles, which was named a Technology Pioneer by The World Economic Forum in 2015. He is also the author of one book and the coauthor of more than 700 peer-reviewed publications in major scientific journals and conferences. He holds 41 issued patents and more than 20 pending patent applications.

Dr. Ozcan is a Fellow of the American Association for the Advancement of Science (AAAS), the International Photonics Society (SPIE), the Optical Society of America (OSA), the American Institute for Medical and Biological Engineering (AIMBE), the Royal Society of Chemistry (RSC), the American Physical Society (APS), and the Guggenheim Foundation. He is also an elected Fellow of the National Academy of Inventors (NAI). He has received major awards, including the Presidential Early Career Award for Scientists and Engineers, the International Commission for Optics Prize, the Biophotonics Technology Innovator Award, the Rahmi M. Koc Science Medal, the International Photonics Society Early Career Achievement Award, the Army Young Investigator Award, the NSF CAREER Award, the NIH Director's New Innovator Award, the Navy Young Investigator Award, the IEEE Photonics Society Young Investigator Award, the Distinguished Lecturer Award, the National Geographic Emerging Explorer Award, the National Academy of Engineering's The Grainger Foundation Frontiers of Engineering Award, and the MIT's TR35 Award for his seminal contributions to computational imaging, sensing and diagnostics.