

# Deep Learning in Biomedical Optics

Lei Tian, PhD <sup>1\*</sup>, Brady Hunt, PhD,<sup>2</sup> Muyinatu A. Lediju Bell, PhD,<sup>3,4,5</sup> Ji Yi, PhD,<sup>4,6</sup> Jason T. Smith,<sup>7</sup> Marien Ochoa,<sup>7</sup> Xavier Intes, PhD,<sup>7</sup> and Nicholas J. Durr, PhD<sup>3,4</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Boston University, 8 St. Mary's Street, RM 830, Boston, Massachusetts 02215

<sup>2</sup>Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755

<sup>3</sup>Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21218

<sup>4</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21218

<sup>5</sup>Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21218

<sup>6</sup>Department of Ophthalmology, Johns Hopkins University, Baltimore, Maryland 21287

<sup>7</sup>Center for Modeling, Simulation, and Imaging in Medicine, Rensselaer Polytechnic Institute, Troy, New York 12180

This article reviews deep learning applications in biomedical optics with a particular emphasis on image formation. The review is organized by imaging domains within biomedical optics and includes microscopy, fluorescence lifetime imaging, *in vivo* microscopy, widefield endoscopy, optical coherence tomography, photoacoustic imaging, diffuse tomography, and functional optical brain imaging. For each of these domains, we summarize how deep learning has been applied and highlight methods by which deep learning can enable new capabilities for optics in medicine. Challenges and opportunities to improve translation and adoption of deep learning in biomedical optics are also summarized. *Lasers Surg. Med.* © 2021 Wiley Periodicals LLC

**Key words:** biomedical optics; biophotonics; deep learning; machine learning; computer aided detection; microscopy; fluorescence lifetime; *in vivo* microscopy; optical coherence tomography; photoacoustic imaging; diffuse tomography; functional optical brain imaging; widefield endoscopy

## INTRODUCTION

Biomedical optics is the study of biological light-matter interactions with the overarching goal of developing sensing platforms that can aid in diagnostic, therapeutic, and surgical applications [1]. Within this large and active field of research, novel systems are continually being developed to exploit unique light-matter interactions that provide clinically useful signatures. These systems face inherent trade-offs in signal-to-noise ratio (SNR), acquisition speed, spatial resolution, field of view (FOV), and depth of field (DOF). These trade-offs affect the cost, performance, feasibility, and overall impact of clinical systems. The role of biomedical optics developers is to design systems which optimize or ideally overcome these trade-offs in order to appropriately meet a clinical need.

In the past few decades, biomedical optics system design, image formation, and image analysis have primarily been guided by classical physical modeling and signal processing methodologies. Recently, however, deep

learning (DL) has become a major paradigm in computational modeling and demonstrated utility in numerous scientific domains and various forms of data analysis [2,3]. As a result, DL is increasingly being utilized within biomedical optics as a data-driven approach to perform image processing tasks, solve inverse problems for image reconstruction, and provide automated interpretation of downstream images. This trend is highlighted in Figure 1, which summarizes the articles reviewed in this paper stratified by publication year and image domain.

This review focuses on the use of DL in the design and translation of novel biomedical optics systems. While image formation is the main focus of this review, DL has also been widely applied to the interpretation of downstream images, as summarized in other review articles [4,5]. This review is organized as follows. First, a brief introduction to DL is provided by answering a set of questions related to the topic and defining key terms and

**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and have disclosed the following: LT, BH, and JY have no conflict of interest. MALB reports a pending patent on machine learning approaches to beamforming. JTS reports grants from NIH R01EB019443, grants from NIH R01CA207725, grants from NIH R01CA237267, grants from NIH R01CA250636, during the conduct of the study; in addition, JTS has a patent 17/143,448 pending. MO reports funding from grants from: NIH R01EB019443, NIH R01CA207725, NIH R01CA237267, NIH R01CA250636, during the conduct of the study and under Principal Investigator Xavier Intes; in addition, MO has a patent 17/143,448 pending. XT reports grants from NIH R01EB019443, grants from NIH R01CA207725, grants from NIH R01CA237267, grants from NIH R01CA250636, grants from MTEC:20-05-IM-PROVE-004, grants from W912CG2120001, during the conduct of the study; in addition, XT has a patent 17/143,448 pending.

\*Correspondence to: Lei Tian, PhD, Department of Electrical and Computer Engineering, Boston University, 8 St. Mary's Street, RM 830, Boston, MA 02215.

E-mail: leitian@bu.edu

Nicholas J. Durr, PhD, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218.

E-mail: ndurr@jhu.edu

Received 11 January 2021; revised 2 April 2021; Accepted 15 April 2021

Published online 00 Month 2020 in Wiley Online Library (wileyonlinelibrary.com).

DOI 10.1002/lsm.23414

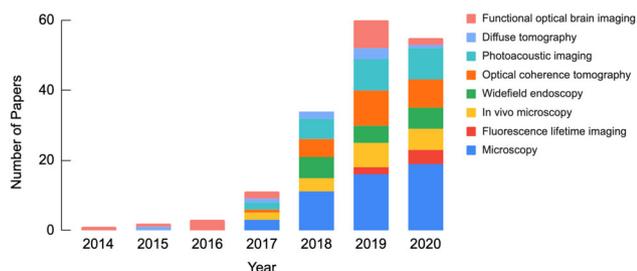


Fig. 1. Number of reviewed research papers which utilize DL in biomedical optics stratified by year and imaging domain. DL, deep learning.

concepts pertaining to the articles discussed throughout this review. Next, recent original research in the following eight optics-related imaging domains is summarized: (i) microscopy, (ii) fluorescence lifetime imaging, (iii) *in vivo* microscopy, (iv) widefield endoscopy, (v) optical coherence tomography, (vi) photoacoustic imaging, (vii) diffuse tomography, and (viii) functional optical brain imaging. Within each domain, state-of-the-art approaches which can enable new functionality for optical systems are highlighted. We then offer our perspectives on the challenges and opportunities across these eight imaging domains. Finally, we provide a summary and outlook of areas in which DL can contribute to future development and clinical impact biomedical optics moving forward.

## DL OVERVIEW

### What is DL?

To define DL, it is helpful to start by defining machine learning (ML), as the two are closely related in their historical development and share many commonalities in their practical application. ML is the study of algorithms and statistical models which computer systems use to progressively improve their performance on a specified task [6]. To ensure the development of generalizable models, ML is commonly broken in two phases: training and testing. The purpose of the training phase is to actively update model parameters to make increasingly accurate predictions on the data, whereas the purpose of testing is to simulate a prospective evaluation of the model on future data.

In this context, DL can be considered a subset of ML, as it is one of many heuristics for development and optimization of predictive, task-specific models [7]. In practice, DL is primarily distinguished from ML by the details of the underlying computational models and optimization techniques utilized. Classical ML techniques rely on careful development of task-specific image analysis features, an approach commonly referred to as “feature engineering” (Fig. 2a). Such approaches typically require extensive manual tuning and therefore have limited generalizability. In contrast, DL applies an “end-to-end” data-driven optimization (or “learning”) of both feature representations *and* model predictions [2] (Fig. 2b). This is achieved through training a type of general

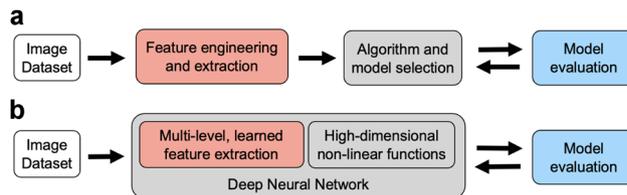


Fig. 2. (a) Classical machine learning uses engineered features and a model. (b) Deep learning uses learned features and predictors in an “end-to-end” deep neural network.

and versatile computational model, termed deep neural network (DNN).

DNNs are composed of multiple layers which are connected through computational operations between layers, including linear weights and nonlinear “activation” functions. Thereby, each layer contains a unique feature representation of the input data. By using several layers, the model can account for both low-level and high-level representations. In the case of images, low-level representations could be textures and edges of the objects, whereas higher level representations would be object-like compositions of those features. The joint optimization of both feature representation at multiple levels of abstraction and predictive model parameters is what makes DNNs so powerful.

### How is DL Implemented?

The majority of the existing DL models in biomedical optics are implemented using the supervised learning strategy. At a high-level, there are three primary components to implement a supervised DL model: (i) labeled data, (ii) model architecture, and (iii) optimization strategy. Labeled data consist of the raw data inputs as well as the desired model output. Large amounts of labeled data are often needed for effective model optimization. This requirement is currently one of the main challenges for utilizing DL on small-scale biomedical datasets, although strategies to overcome this are an active topic in the literature, such as unsupervised [8], self-supervised [9], and semi-supervised learning [10]. For a typical end-to-end DL model, model architecture defines the hypothesis class and how hierarchical information flows between each layer of the DNN. The selection of a DNN architecture depends on the desired task and is often determined empirically through comparison of various state-of-the-art architectures. Three of the most widely used DNN architectures in current biomedical optics literature are illustrated in Figure 3.

The encoder-decoder network [11] shown in Figure 3a aims to establish a mapping between the input and output images using a nearly symmetrically structure with a contracting “encoder” path and an expanding “decoder” path. The encoder consists of several convolutional blocks, each followed by a down-sampling layer for reducing the spatial dimension. Each convolutional block consists of several convolutional layers (Conv2D) that stacks the processed features along the last dimension,

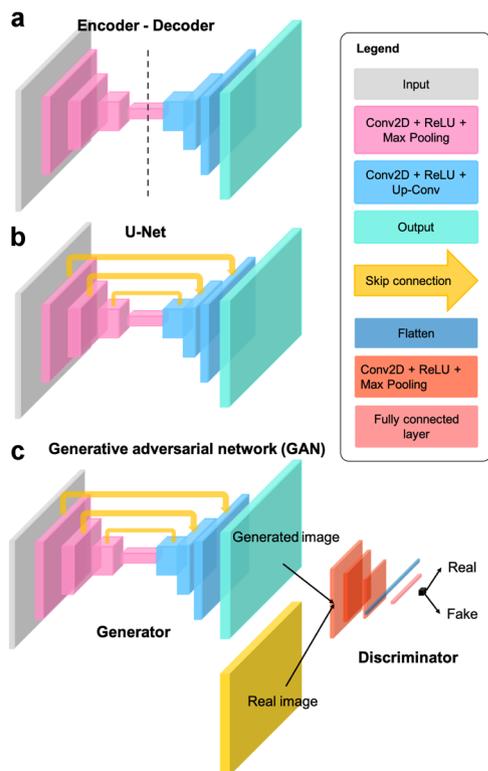


Fig. 3. Three of the most commonly-used DNN architectures in biomedical optics: (a) Encoder-decoder, (b) U-Net, and (c) GAN, generative adversarial network.

among which each layer is followed by a nonlinear activation function, for example, the Rectified Linear Unit (ReLU). The intermediate output from the encoder has a small spatial dimension but encodes rich information along the last dimension. These low-resolution “activation maps” go through the decoder, which consists of several additional convolutional blocks, each connected by a up-sampling convolutional (Up-Conv) layer for increasing the spatial dimension. The output of the network typically has the same dimension as the input image.

The U-Net [12] architecture shown in Figure 3b can be thought of as an extension to the encoder-decoder network. It introduces additional “skip connections” between the encoder and decoder paths so that information across different spatial scales can be efficiently tunneled through the network, which has shown to be particularly effective to preserve high-resolution spatial information [12].

The generative adversarial network (GAN) [13] shown in Figure 3c is a general framework that involves adversarially training a pair of networks, including the “generator” and “discriminator.” The basic idea is to train the generator to make high-quality image predictions that are indistinguishable from the real images of the same class (e.g., H&E stained lung tissue slices).

To do so, the discriminator is trained to classify that the generator’s output is fake, while the generator is trained to fool the discriminator. Such alternating training steps iterate until a convergence is met when the discriminator

can hardly distinguish if the images produced from the generator are fake or real. When applying to biomedical optics techniques, the generator is often implemented by the U-Net. The discriminator is often implemented using an image classification network. The input image is first processed by several convolutional blocks and down-sampling layers to extract high-level two-dimensional (2D) features. These 2D features are then “flattened” to a 1D vector, which is then processed by several fully connected layers to perform additional feature synthesis and make the final classification.

Once labeled data and model architecture have been determined, optimization of model parameters can be undertaken. Optimization strategy includes two aspects: (i) cost function, and (ii) training algorithm. Definition of a cost function (a.k.a. objective function, error, or loss function) is needed to assess the accuracy of model predictions relative to the desired output and provide guidance to adjust model parameters. The training algorithm iteratively updates the model parameters to improve model accuracy. This training process is generally achieved by solving an optimization problem, using variants of the gradient descent algorithm, for example, stochastic gradient descent and Adam [14]. The optimizer utilizes the gradient of the cost function to update each layer of the DNN through the principle of “error backpropagation” [15]. Given labeled data, a model architecture, and the optimization guides the model parameters toward a local minimum of the cost function, thereby optimizing model performance.

With the recent success of DL, several software frameworks have been developed to enable easier creation and optimization of DNNs. Many of the major technology companies have been active in this area. Two of the front-runners are TensorFlow and PyTorch, which are open-source frameworks published and maintained by Google and Facebook, respectively [16,17]. Both frameworks enable easy construction of custom DNN models, with efficient parallelization of DNN optimization over high-performance graphics computing units (GPUs). These frameworks have enabled non-experts to train and deploy DNNs and have played a large role in the spread of DL research into many new applications, including the field of biomedical optics.

### What is DL Used for in Biomedical Optics?

There are two predominant tasks for which DL has been utilized in biomedical optics: (i) image formation and (ii) image interpretation. Both are important applications; however, image formation is a more central focus of biomedical optics researchers and consequently is the focus of this review.

With regards to image formation, DL has proven very useful for effectively approximating the inverse function of an imaging model in order to improve the quality of image reconstructions. Classical reconstruction techniques are built on physical models with explicit analytical formulations. To efficiently compute the inverse of these analytical models, approximations are often needed to

simplify the problem, for example, linearization. Instead, DL methods have shown to be very effective to directly “learn” an inverse model, in the form of a DNN, based on the training input and output pairs. This in practice has opened up novel opportunities to perform image formation that would otherwise be difficult to formulate an analytical model. In addition, the directly learned DL inverse model can often better approximate the inverse function, which in turn leads to improved image quality as shown in several imaging domains in this review.

Secondly, DL has been widely applied for modeling image priors for solving the inverse problems across multiple imaging domains. Most image reconstruction problems are inherently ill-posed in that the reconstructed useful image signal can be overwhelmed by noise if a direct inversion is implemented. Classical image reconstruction techniques rely on regularization using parametric priors for incorporating features of the expected image. Although being widely used, such models severely limit the type of features that can be effectively modeled, which in turn limit the reconstruction quality. DL-based reconstruction bypasses this limitation and does not rely on explicit parameterization of image features, but instead represents priors in the form of a DNN which is optimized (or “learned”) from a large data set that is of the same type of the object of interest (e.g., endoscopic images of esophagus). By doing so, DL enables better quality reconstructions.

Beyond achieving higher quality reconstructions, there are other practical benefits of DNNs in image formation. Classical inversion algorithms typically require an iterative process that can take minutes to hours to compute. Furthermore, they have stringent sampling requirements, which if lessened, make the inversion severely ill-posed. Due to more robust “learned” priors, DL-based techniques can accommodate highly incomplete or undersampled inputs while still providing high-quality reconstructions. Additionally, although DNNs typically require large data sets for training, the resulting models are capable of producing results in real time with a GPU. These combined capabilities allow DL-based techniques to bypass physical trade-offs (e.g., acquisition speed and imaging quality) and enable novel capabilities beyond existing solutions.

By leveraging these unique capabilities of DL methods, innovative techniques have been broadly reported across many imaging domains in biomedical optics. Examples include improving imaging performance, enabling new imaging functionalities, extracting quantitative microscopic information, and discovering new biomarkers. These and other technical developments have the potential to significantly reduce system complexity and cost, and may ultimately improve the quality, affordability, and accessibility of biomedical imaging in health care.

## DL APPLICATIONS IN BIOMEDICAL OPTICS

### Microscopy

**Overview.** Microscopy is broadly used in biomedical and clinical applications to capture cellular and tissue structures based on intrinsic (e.g., scattering, phase, and

autofluorescence) or exogenous contrast (e.g., stains and fluorescent labels). Fundamental challenges exist in all forms of microscopy because of the limited information that can be extracted from the instrument. Broadly, the limitations can be categorized based on two main sources of origin. The first class is due to the physical tradeoffs between multiple competing performance parameters, such as SNR, acquisition speed, spatial resolution, FOV, and DOF. The second class is from the intrinsic sensitivity and specificity of different contrast mechanisms. DL-augmented microscopy is a fast-growing area that aims to overcome various aspects of conventional limitations by synergistically combining novel instrumentation and DL-based computational enhancement. This section focuses on DL strategies for bypassing the physical tradeoffs and augmenting the contrast in different microscopy modalities.

**Overcoming physical tradeoffs.** An ideal microscopy technique often needs to satisfy several requirements, such as high resolution in order to resolve the small features in the sample, low light exposure to minimize photo-damage, and a wide FOV in order to capture information from a large portion of the sample. Traditional microscopy is fundamentally limited by the intrinsic tradeoffs between various competing imaging attributes. For example, a short light exposure reduces the SNR; a high spatial resolution requires a high-magnification objective lens that provides a small FOV and shallow DOF. This section summarizes recent achievements in leveraging DL strategies to overcome various physical tradeoffs and expand the imaging capabilities.

**Denoising:** Enhancing microscopy images by DL-based denoising has been exploited to overcome the tradeoffs between light exposure, SNR, and imaging speed, which in turn alleviates photo-bleaching and photo-toxicity. The general strategy is to train a supervised network that takes a noisy image as the input and produces the SNR-enhanced image output. Weigert et al. [18] demonstrated a practical training strategy of a U-Net on experimental microscopy data that involves taking paired images with low and high light exposures as the noisy input and high-SNR output of the network [19–23] (Fig. 4a). This work showed that the DNN can restore the same high-SNR images with 60-fold fewer photons used during the acquisition. Similar strategies have been applied to several microscopy modalities, including widefield, confocal, light-sheet [18], structured illumination [24], and multi-photon microscopy [25].

**Image reconstruction:** Beyond denoising, the imaging capabilities of several microscopy techniques can be much expanded by performing image reconstruction. To perform reconstruction by DL, the common framework is to train a DNN, such as the U-Net and GAN, that takes the raw measurements as the input and the reconstructed image as the output. With this DL framework, three major benefits have been demonstrated. First, Wang et al. showed that GAN-based super-resolution reconstruction

allows recovering high-resolution information from low-resolution measurements, which in turn provides an enlarged FOV and an extended DOF [19] (Fig. 4b). For widefield imaging, [19] demonstrated super-resolution reconstruction using input images from a  $\times 10/0.4$ -NA objective lens and producing images matching a  $\times 20/0.75$ -NA objective lens. In a cross-modality confocal-to-STED microscopy transformation case, [19] showed resolution improvement from 290 to 110 nm. Similar results have also been reported in label-free microscopy modalities, including brightfield [26], holography [27], and quantitative phase imaging [20] (Fig. 4c).

Second, DL-based 3D reconstruction technique allows drastically extending the imaging depth from a single-shot and thus bypasses the need for physical focusing. In [21], Wu et al. demonstrated  $\times 20$  DOF extension in widefield

fluorescence microscopy using a conditional GAN (Fig. 4d). Recent work on DL-based extended DOF has also shown promising results on enabling rapid slide-free histology [28].

Third, DL significantly improves both the imaging acquisition and reconstruction speeds and reduces the number of measurements for microscopy modalities that intrinsically require multiple measurements for the image formation, as shown in quantitative phase microscopy [20,29,30] (Fig. 4c), single molecule localization microscopy [31–33], and structured illumination microscopy [24]. For example, in [20], a 97% data reduction as compared to the conventional sequential acquisition scheme was achieved for gigapixel-scale phase reconstruction based on a multiplexed acquisition scheme using a GAN.

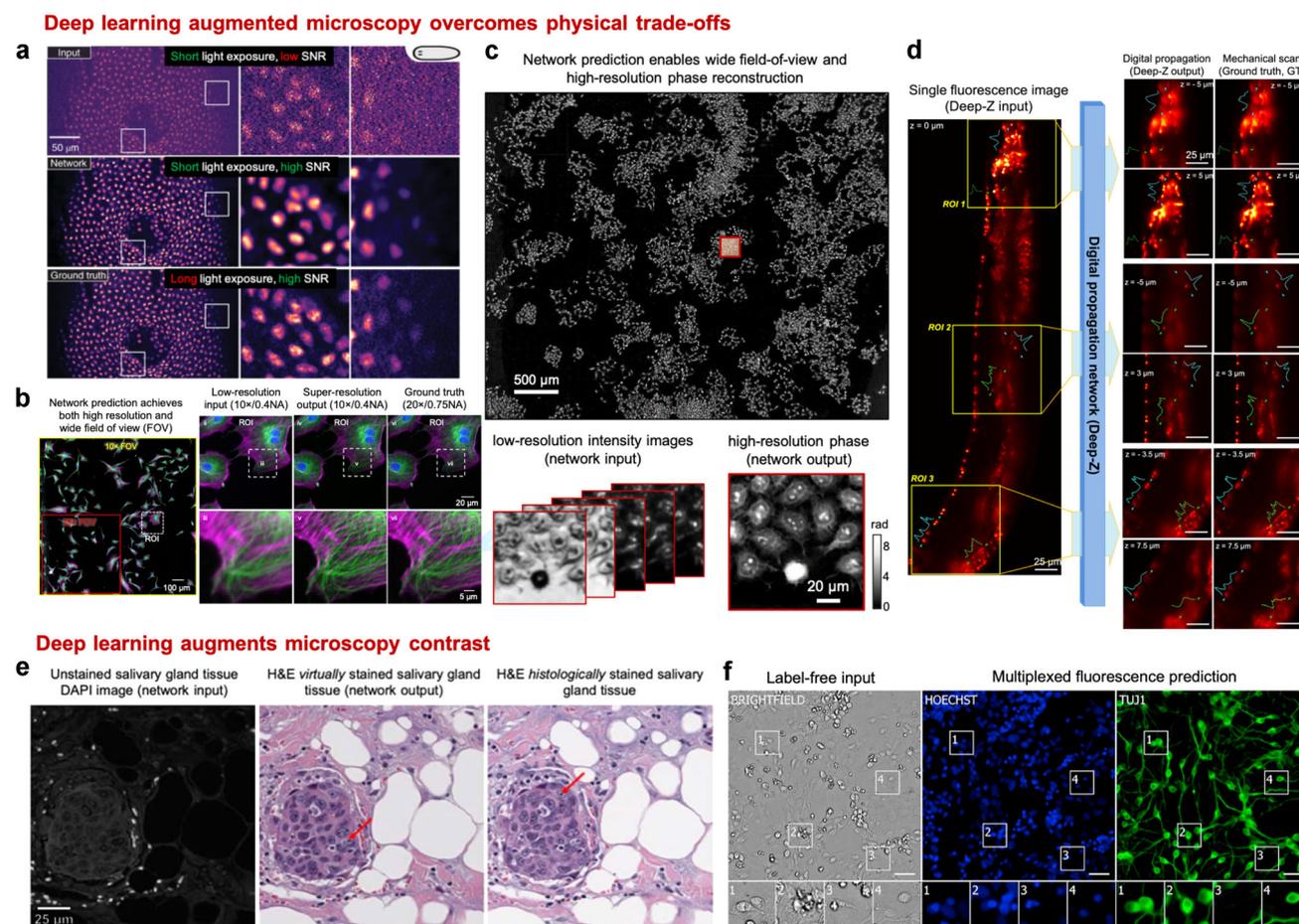


Fig. 4. DL overcomes physical tradeoffs and augments microscopy contrast. (a) CARE network achieves higher SNR with reduced light exposure (with permission from the authors [18]). (b) Cross-modality super-resolution network reconstructs high-resolution images across a wide FOV [19] (with permission from the authors). (c) DL enables wide-FOV high-resolution phase reconstruction with reduced measurements (adapted from [20]). (d) Deep-Z network enables digital 3D refocusing from a single measurement [21] (with permission from the authors). (e) Virtual staining GAN transforms autofluorescence images of unstained tissue sections to virtual H&E staining [22] (with permission from the authors) (f) DL enables predicting fluorescent labels from label-free images [23] (Reprinted from Christiansen et al. [23], Copyright (2020), with permission from Elsevier). 3D, three dimensional; DL, deep learning; FOV, field of view; GAN, generative adversarial network; SNR, signal-to-noise ratio.

**Augmenting contrasts.** The image contrast used in different microscopy modalities can be broadly categorized into endogenous and exogenous. For example, label-free microscopy captures endogenous scattering and phase contrast, and is ideal for imaging biological samples in their natural states, but suffers from lack of molecular specificity. Specificity is often achieved by staining with absorbing or fluorescent labels. However, applications of exogenous labeling are limited by the physical staining/labeling process and potential perturbation to the natural biological environment. Recent advances in DL-augmented microscopy have the potential to achieve the best of both label-free and labeled microscopy. This section summarizes two most widely used frameworks for augmenting microscopy contrast with DL.

**Virtual staining/labeling:** The main idea of virtual staining/labeling is to digitally transform the captured label-free contrast to the target stains/labels. DL has been shown to be particularly effective to perform this “cross-modality image transformation” task. By adapting this idea to different microscopy contrasts, two emerging applications have been demonstrated. First, virtual histological staining has been demonstrated for transforming a label-free image to the brightfield image of the histologically-stained sample (Fig. 4e). The label-free input utilized for this task include autofluorescence [22,34], phase [35,36], multi-photon and fluorescence lifetime [37]. The histological stains include H&E, Masson’s Trichrome and Jones’ stain. Notably, the quality of virtual staining on tissue sections from multiple human organs of different stain types was assessed by board-certified pathologists to show superior performance [22]. A recent cross-sectional study has been carried out for clinical evaluation of unlabeled prostate core biopsy images that have been virtually stained [38]. The main benefits of the virtual staining approach include saving time and cost [22], as well as facilitating multiplexed staining [34]. Interested readers can refer to a recent review on histopathology using virtual staining [39].

Second, digital fluorescence labeling has been demonstrated for transforming label-free contrast to fluorescence labels [23,40–43] (Fig. 4f). In the first demonstration [23], Christiansen et al. performed 2D digital labeling using transmission brightfield or phase contrast images to identify cell nuclei (accuracy quantified by Pearson correlation coefficient  $PCC = 0.87\text{--}0.93$ ), cell death ( $PCC = 0.85$ ), and to distinguish neuron from astrocytes and immature dividing cells ( $PCC = 0.84$ ). A main benefit of digital fluorescence labeling is digital multiplexing of multiple subcellular fluorescence labels, which is particularly appealing to kinetic live cell imaging. This is highlighted in [40], 3D multiplexed digital labeling using transmission brightfield or phase contrast images on multiple subcellular components are demonstrated, including nucleoli ( $PCC\sim 0.9$ ), nuclear envelope, microtubules, actin filaments ( $PCC\sim 0.8$ ), mitochondria, cell membrane, Endoplasmic reticulum, DNA+ ( $PCC\sim 0.7$ ),

DNA ( $PCC\sim 0.6$ ), Actomyosin bundles, tight junctions ( $PCC\sim 0.5$ ), Golgi apparatus ( $PCC\sim 0.2$ ), and Desmosomes ( $PCC\sim 0.1$ ). Recent advances further exploit other label-free contrasts, including polarization [41], quantitative phase map [43], and reflectance phase-contrast microscopy [42]. Beyond predicting fluorescence labels, recent advances further demonstrate multiplexed single-cell profiling using the digitally predicted labels [42].

In both virtual histopathological staining and digital fluorescence labeling, the U-Net forms the basic architecture to perform the image transformation. GAN has also been incorporated to improve the performance [22,38].

**Classification:** Instead of performing pixel-wise virtual stain/label predictions, DL is also very effective in holistically capturing complex “hidden” image features for classification. This has found broad applications in augmenting the label-free measurements and provide improved specificity and classify disease progression [44,45] and cancer screening [46–48], as well as detect cell types [49,50], cell states [44,51], stem cell lineage [52–54], and drug response [55]. For example, in [44], Eulenberg *et al.* demonstrated a classification accuracy of 98.73% for the G1/S/G2 phase, which provided 6× improvement in error rate as compared to the previous state-of-the-art method based on classical ML techniques.

**Opportunities and challenges.** By overcoming the physical tradeoffs in traditional systems, DL-augmented microscopy achieves unique combinations of imaging attributes that are previously not possible. This may create new opportunities for diagnosis and screening. By augmenting the contrast using virtual histological staining techniques, DL can open up unprecedented capabilities in label-free and slide-free digital pathology. This can significantly simplify the physical process and speed up the diagnosis. By further advancing the digital fluorescence labeling techniques, it can enable high-throughput and highly multiplexed single-cell profiling and cytometry. Beyond clinical diagnoses, this may find applications in drug screening and phenotyping.

In addition, several emerging DL techniques can further enhance the capabilities of microscopy systems. First, DL can be applied to optimize the hardware parameters used in microscopy experiments. In quantitative phase microscopy, DL was applied to optimized the illumination patterns to reduce the data requirement [30,56]. In single molecule localization microscopy, DL was used to optimize the point spread functions to enhance the localization accuracy [33,57]. DL has also been used to optimize the illumination power [58] and focus positions [59–61].

Second, new DL frameworks are emerging to significantly reduce the labeled data requirements in training, which is particularly useful in biomedical microscopy since acquiring a large-scale labeled training data set is often impractical. For example, a novel denoising approach, known as Noise2Noise [62], has been developed that can be trained using only independent

pairs of noisy images, and bypasses the need for ground-truth clean images. Following this work, self-supervised denoising DL approaches have been advanced to further alleviate the training data requirement. Techniques, such as Noise2Void, Noise2Self, and their variants, can be directly trained on noisy data set without the need for paired noisy images [63–65]. In addition, semi-supervised and unsupervised DL approaches have also been developed to reduce or completely remove the need for labeled training data during training, which have been demonstrated for vessel segmentation [66,67]. Lastly, physics-embedded DL opens up a new avenue for reducing training requirements by incorporating the physical model of the microscopy technique [68,69].

Finally, uncertainty quantification techniques address the need for assessing the reliability of the DL model by quantifying the confidence of the predictions, and has recently been applied in quantitative phase reconstruction [20].

### Fluorescence Lifetime Imaging (FLI)

**Overview.** Fluorescence imaging has become a central tool in biomedical studies with high sensitivity to observe endogenous molecules [70,71] and monitor important biomarkers [72]. Increasingly, fluorescence imaging is not limited to intensity-based techniques but can extract additional information by measuring fluorophore lifetimes [73–75]. FLI has become an established technique for monitoring cellular micro-environment via analysis of various intracellular parameters [76], such as metabolic state [77,78], reactive oxygen species [79], and/or intracellular pH [80]. FLI is also a powerful technique for studying molecular interactions inside living samples, via Förster Resonance Energy Transfer (FRET) [81], enabling applications such as quantifying protein-protein interactions [82], monitoring biosensor activity [83], and ligand-receptor engagement *in vivo* [84]. However, FLI is not a direct imaging technique. To quantify lifetime or lifetime-derived parameters, an inverse solver is required for quantification and/or interpretation.

To date, image formation is the main utility of DL in FLI. Contributions include reconstructing quantitative lifetime image from raw FLI measurements, enabling enhanced multiplexed studies by leveraging both spectral and lifetime contrast simultaneously, and facilitating improved instrumentation with compressive measurement strategies.

**Lifetime quantification, representation, and retrieval.** Conventionally, lifetime quantification is obtained at each pixel via model-based inverse-solvers, such as least-square fitting and maximum-likelihood estimation [85], or the fit-free phasor method [86,87]. The former is time-consuming, inherently biased by user-dependent *a priori* settings, and requires operator expertise. The phasor method is the most widely-accepted alternative for lifetime representation [87]. However, accurate quantification using the phasor method requires careful calibration, and when considering tissues/turbid-media in FLI microscopy (FLIM)

applications, additional corrections are needed [87,88]. Therefore, it has largely remained qualitative in use.

Wu et al. [89] demonstrated a multilayer perceptron (MLP) for lifetime retrieval for ultrafast bi-exponential FLIM. The technique exhibited an 180-fold faster speed than conventional techniques, yet it was unable to recover the true lifetime-based values in many cases due to ambiguities caused by noise. Smith et al. [90] developed an improved 3D-CNN, FLI-Net, that can retrieve spatially independent bi-exponential lifetime parameter maps directly from the 3D  $(x, y, t)$  FLI data. By training with a model-based approach including representative noise and instrument response functions, FLI-Net was validated across a variety of biological applications. These include quantification of metabolic and FRET FLIM, as well as preclinical lifetime-based studies across the visible and near-infrared (NIR) spectra. Further, the approach was generalized across two data acquisition technologies—Time-correlated Single Photon Counting (TCSPC) and Intensified Gated CCDs (ICCD). FLI-Net has two advantages. First, it outperformed classical approaches in the presence of low photon counts, which is a common limitation in biological applications. Second, FLI-Net can output lifetime-based whole-body maps at 80 milliseconds in widefield pre-clinical studies, which highlights the potential of DL methods for fast and accurate lifetime-based studies. In combination with DL *in silico* training routines that can be crafted for many applications and technologies, DL is expected to contribute to the dissemination and translation of FLI methods as well as to impact the design and implementation of future-generation FLI instruments. An example FLI-Net output for metabolic FLI is shown in Figure 5.

**Emerging FLI applications using DL.** The technologies used in FLI have not fundamentally shifted over the last two decades. One bottleneck for translation is a lack of sensitive, widefield NIR detectors. Advances in computational optics have sparked development of new approaches using structured light [91], such as single-pixel methods [92]. These methods are useful when widefield detectors are lacking, such as in applications with low photon budget and when higher dimensional data are sought [92] (e.g., hyperspectral imaging [93]). However, these computational methods are based on more complex inverse models that require user expertise and input.

Yao et al. [94] developed a CNN, NetFLICS, capable of retrieving both intensity and lifetime images from single-pixel compressed sensing-based time-resolved input. NetFLICS generated superior quantitative results at low photon count levels, while being four orders of magnitude faster than existing approaches. Ochoa-Mendoza et al. [95] further developed the approach to increase its compression ratio to 99% and the reconstruction resolution to  $128 \times 128$  pixels. This dramatic improvement in compression ratio enables significantly faster imaging protocols and demonstrates how DL can impact instrumentation design to improve clinical utility and workflow [96].

Recent developments have made hyperspectral FLI imaging possible across microscopic [97] and macroscopic

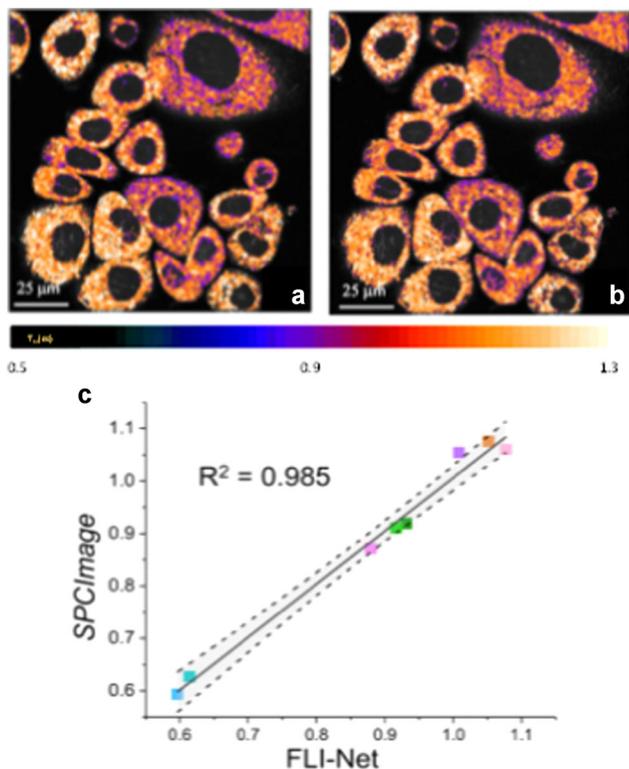


Fig. 5. Example of quantitative FLI metabolic imaging as reported by NADH  $\tau_m$  for a breast cancer cell line (AU565) as obtained (a) with SPCImage and (b) FLI-Net. (c) Linear regression with corresponding 95% confidence band (gray shading) of averaged NADH  $\tau_m$  values from 4 cell line data (adapted from [90]). FLI, fluorescence lifetime imaging.

settings [98]. Traditionally, combining spectral and lifetime contrast analytically is performed independently or sequentially using spectral decomposition and/or iterative fitting [99]. Smith et al. [100] proposed a DNN, UNMIX-ME, to unmix multiple fluorophore species simultaneously for both spectral and temporal information. UNMIX-ME takes a 4D voxel  $(x, y, t, \lambda)$  as the input and outputs spatial  $(x, y)$  maps of the relative contributions of distinct fluorophore species. UNMIX-ME demonstrated higher performance during tri- and quadri-abundance coefficient retrieval. This method is expected to find utility in applications such as autofluorescence imaging in which unmixing of metabolic and structural biomarkers is challenging.

Although FLI has shown promise for deep tissue imaging in clinical scenarios, FLI information is affected by tissue optical properties. Nonetheless, there are several applications that would benefit from optical property-corrected FLI without solving the full 3D inverse problem. For optical guided surgery, Smith et al. [101] proposed a DNN that outputs 2D maps of the optical properties, lifetime quantification, and the depth of fluorescence inclusion (topography). The DNN was trained using a model-based approach in which a data simulation

workflow incorporated “Monte Carlo eXtreme” [102] to account for light propagation through turbid media. The method was demonstrated experimentally, with real-time applicability over large FOVs. Both widefield time-resolved fluorescence imaging and spatial frequency domain imaging (SFDI) in its single snapshot implementation were performed with fast acquisition [91] and processing speeds [103]. Hence, their combination with DL-based image processing provides a possible future foundation for real-time intraoperative use.

While recent advances in FLI-based classification and segmentation are limited to using classical ML techniques [104–106], Sagar et al. [107] used MLPs paired with bi-exponential fitting for label-free detection of microglia. However, DL approaches often outperform such “shallow learning” classifiers. Although reports using DL for classification based on FLI data are currently absent from the literature, it is expected that DL will play a critical role in enhancing FLI classification and semantic segmentation tasks in the near future.

### In Vivo Microscopy (IVM)

**Overview.** IVM techniques enable real-time assessment of intact tissue at magnifications similar to that of conventional histopathology [108]. As high-resolution assessment of intact tissue is desirable for many biomedical imaging applications, a number of optical techniques and systems have been developed which have trade-offs in FOV, spatial resolution, achievable sampling rates, and practical feasibility for clinical deployment [108]. However, a commonality of IVM systems used for clinical imaging is the need for image analysis strategies to support intraoperative visualization and automated diagnostic assessment of the high-resolution image data. Currently, three of the major IVM techniques for which DL is being utilized are optical coherence tomography (OCT) [109], confocal laser endomicroscopy (CLE, Fig. 6) [110], and reflectance confocal microscopy (RCM) [111]. This section focuses on DL approaches for CLE and RCM. More specifically, endoscopic imaging using probe-based CLE (pCLE) and dermal imaging for RCM. OCT is discussed in a subsequent section.

**Automated diagnosis.** Automated diagnostic classification has been the earliest and most frequent application of DL within IVM. Most commonly, histopathology analysis of imaged specimens provides a ground truth categorization for assessing diagnostic accuracy. The limited size of pCLE and RCM data sets and logistical challenges in precisely correlating them with histopathology remain two ongoing challenges for training robust classifiers. To address these challenges, a variety of strategies have been applied which range from simple classification schemes (benign vs. malignant) using pre-trained CNNs [113] to more complicated tasks, such as cross-domain feature learning and multi-scale encoder-decoder networks [115,116]. The following section contrasts recent reports and methods utilizing DL

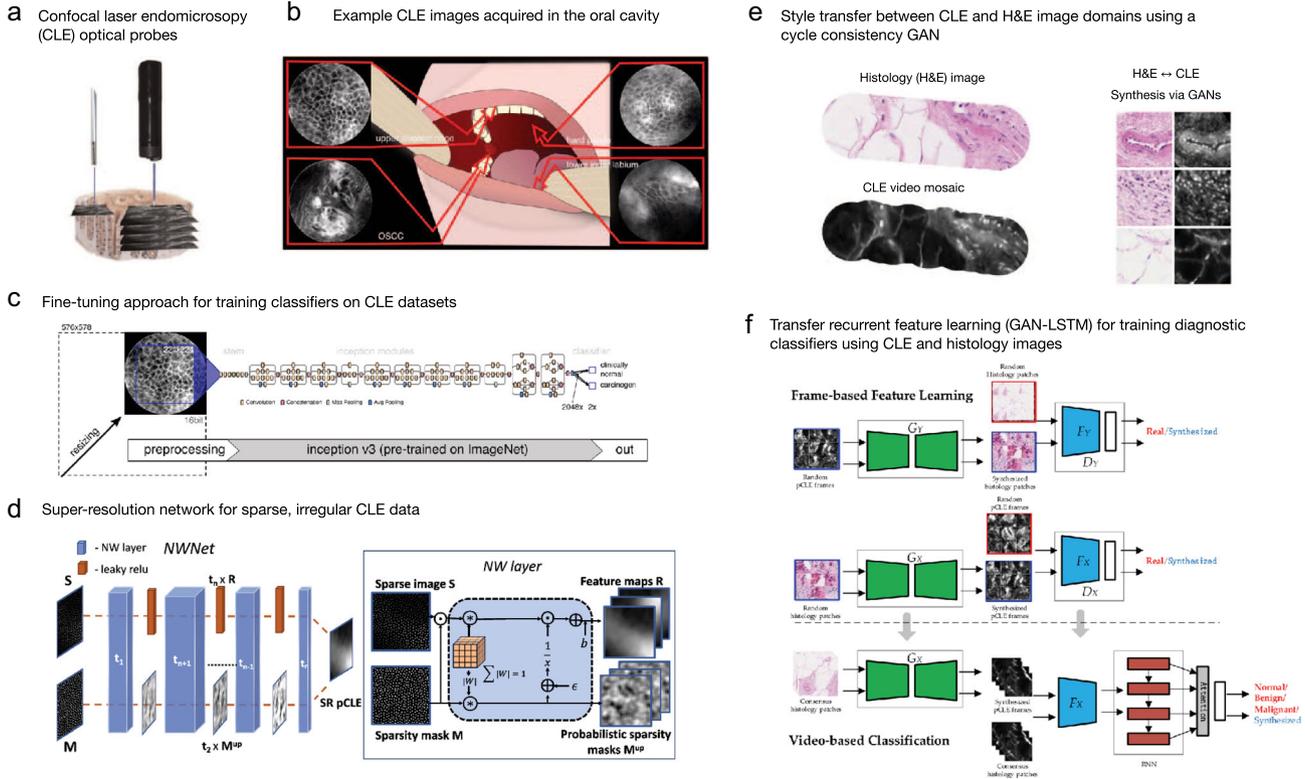


Fig. 6. DL approaches to support real-time, automated diagnostic assessment of tissues with confocal laser endomicroscopy. (a) Graphical rendering of two confocal laser endomicroscopy probes (left: Cellvizio, right: Pentax) (adapted from [112]). (b) Example CLE images obtained from four different regions of the oral cavity (adapted from [113]) (c) Fine-tuning of CNNs pre-trained using ImageNet is utilized in the majority of CLE papers reported since 2017 (adapted from [113]). (d) Super-resolution networks for probe-based CLE images incorporate novel layers to better account for the sparse, irregular structure of the images (adapted from [114]). (e) Example H&E stained histology images with corresponding CLE images. Adversarial training of GANs to transfer between these two modalities has been successful (adapted from [115]). (f) Transfer recurrent feature learning utilizes adversarially trained discriminators in conjunction with an LSTM for state-of-the-art video classification performance (adapted from [115]). CLE, confocal laser endomicroscopy; DL, deep learning; GAN, generative adversarial network.

for diagnostic image analysis of pCLE and RCM image data sets.

### **CNNs and transfer learning approaches:**

Early reports on DL-based image classification for CLE and RCM have demonstrated that transfer learning using pre-trained CNNs can outperform conventional image analysis approaches, especially when data is limited as is often the case for CLE and RCM [113,117–119].

Aubreville et al. published an early and impactful study comparing the performance of two CNN-based approaches to a textural feature-based classifier (random forest) on pCLE video sequences acquired during surgical resection of oral squamous carcinoma (Fig. 6b) [113]. Of their two CNN-based approaches, one was a LeNet-5 architecture and was trained to classify sub-image patches whereas the other utilized transfer learning of a pre-trained CNN (Fig. 6c) for whole image classification. Using leave-one-out cross validation on 7894 frames from 12 patients, the two CNN-based approaches both outperformed the textural classifier.

Transfer learning is one strategy to overcome limited data set sizes, which remains a common challenge for CLE and RCM. As larger CLE and RCM data sets are obtainable in the future, transfer learning is unlikely to be an optimal strategy for image classification; however, it can remain a useful benchmark for the difficulty of image classification tasks on novel, small-scale data sets moving forward. The subsequent sections introduce alternatives to transfer learning which utilize video data as well as cross-domain learning.

**Recurrent convolutional approaches:** CLE and RCM are typically used in video recording while the optical probe is physically or optically scanned to obtain images over a larger tissue area or at varying depths. Some reports have utilized recurrent convolutional networks to account for spatial and/or temporal context of image sequences [120–122]. The additional spatial/temporal modeling provided by recurrent networks is one promising approach to leverage video data. [120–122].

**Cross-domain learning:** A novel approach, termed “transfer recurrent feature learning,” was developed by Gu et al. which leveraged cross-domain feature learning for classification of pCLE videos obtained from 45 breast tissue specimens [115]. Although this method relied on data acquired *ex vivo*, the data itself is not qualitatively different from other pCLE data sets and still provides a proof-of-principle. Their model utilized a cycle-consistent GAN (CycleGAN) to first learn feature representations between H&E microscopy and pCLE images and to identify visually similar images (Fig. 6e). The optimized discriminator from the CycleGAN is then utilized in conjunction with a recurrent neural network to classify video sequences (Fig. 6f). The method outperformed other DL methods and achieved 84% accuracy in classifying normal, benign, and malignant tissues.

**Multiscale segmentation:** Kose et al. [116] developed a novel segmentation architecture, “multiscale encoder-decoder network” (MED-Net), which outperformed other state-of-the-art network architectures for RCM mosaic segmentation. In addition to improving accuracy, MED-Net produced more globally consistent, less fragmented pixel-level classifications. The architecture is composed of multiple, nested encoder-decoder networks and was inspired by how pathologists often examine images at multiple scales to holistically inform their image interpretation.

**Image quality assessment:** A remaining limitation of many studies was some level of manual or semi-automated pre-processing of pCLE and RCM images/videos to exclude low-quality and/or non-diagnostic image data. Building on the aforementioned reports for diagnostic classification, additional work utilized similar techniques for automated image quality assessment using transfer learning [123,124] as well as MED-Net [125].

**Super-resolution.** Several IVM techniques, including pCLE, utilize flexible fiber-bundles as contact probes to illuminate and collect light from localized tissue areas [126]. Such probes are needed for minimally invasive endoscopic procedures and can be guided manually or via robotics. The FOV of fiber-optic probes is typically  $<1 \text{ mm}^2$  and lateral resolution is limited by the inter-core spacing of individual optical fibers, which introduce a periodic image artifact (“honeycomb patterns”) from the individual fibers.

Shao et al. [127] developed a novel super-resolution approach which outperformed maximum a posteriori (MAP) estimation using a two-stage CNN model which first estimates the motion of the probe and then reconstructs a super-resolved image using the aligned video sequence. The training data was acquired using a dual camera system, one with and one without a fiber-bundle in the optical setup, to obtain paired data.

Others have taken a more computational approach to pCLE super-resolution by using synthetic data sets. For example, Ravi et al. [128] demonstrated super-resolution of pCLE images using unpaired image data via a CycleGAN, and Szcotka et al. [114] introduced a novel

Nadaraya-Watson layer to account for the irregular sparse artifacts introduced by the fiber-bundle (Fig. 6d).

**Future directions.** Beyond automatic diagnosis and super-resolution approaches in IVM, recent advances also highlight ways in which DL can enable novel instrumentation development and image reconstructions to enable new functionalities for compact microscopy systems. Such examples include multispectral endomicroscopy [129], more robust mosaicking for FOV expansion [130], and end-to-end image reconstruction using disordered fiber-optic probes [131,132]. We anticipate that similarly to *ex vivo* microscopy, in the coming years DL will be increasingly utilized to overcome physical constraints, augment contrast mechanisms, and enable new capabilities for IVM systems.

### Widefield Endoscopy

**Overview.** The largest application of optics in medical imaging, by U.S. market size, is widefield endoscopy [133]. In this modality, tissue is typically imaged on the  $>1 \text{ cm}$  scale, over a large working distance range, with epi-illumination and video imaging via a camera. Endoscopic and laparoscopic examinations are commonly used for screening, diagnostic, preventative, and emergency medicine. There has been extensive research in applying various DL tools for analyzing conventional endoscopy images for improving and automating image interpretation [134–137]. This section instead reviews recent DL research in image formation tasks in endoscopy, including denoising, resolution enhancement, 3D scene reconstruction, mapping of chromophore concentrations, and hyperspectral imaging.

**Denoising.** A hallmark of endoscopic applications is challenging geometrical constraints. Imaging through small lumens such as the gastrointestinal tract or -Ikeyholes- for minimally-invasive surgical applications requires optical systems with compact footprints—often on the order of 1-cm diameter. These miniaturized optical systems typically utilize small-aperture cameras with high pixel counts, wide FOVs, and even smaller illumination channels. Consequently, managing the photon budget is a significant challenge in endoscopy, and there have been several recent efforts to apply DL to aid in high-quality imaging in these low-light conditions. A low-light net (LLNET) with contrast-enhancement and denoising autoencoders has been introduced to adaptively brighten images [138]. This study simulated low-light images by darkening and adding noise and found that training on this data resulted in a learned model that could enhance natural low-light images. Other work has applied a U-Net for denoising on high-speed endoscopic images of the vocal folds, also by training on high-quality images that were darkened with added noise [139]. Brightness can also be increased via laser-illumination, which allows greater coupling efficiency than incoherent sources, but results in laser speckle noise in the image from coherent interference. Conditional GANs have been

applied to predict speckle-free images from laser-illumination endoscopy images by training on image pairs acquired of the same tissue with both coherent and incoherent illumination sources [140].

**Improving image quality.** In widefield endoscopy, wet tissue is often imaged in a perpendicular orientation to the optical axis, and the close positioning of the camera and light sources leads to strong specular reflections that mask underlying tissue features. GANs have been applied to reduce these specular reflections [141]. In this case, unpaired training data with and without specular reflections were used in a CycleGAN architecture with self-regularization to enforce similarity between the input specular and predicted specular-free images. Other work has found that specular reflection removal can be achieved in a simultaneous localization and mapping elastic fusion architecture enhanced by DL depth estimation [142]. Lastly, Ali et al. [143] introduced a DL framework that identifies a range of endoscopy artifacts (multi-class artifact detection), including specular reflection, blurring, bubbles, saturation, poor contrast, and miscellaneous artifacts using YOLOv3-spp with classes that were hand-labeled on endoscopy images. These artifacts were then removed and the image restored using GANs.

**Resolution enhancement.** For capsule endoscopy applications, where small detectors with low pixel counts are required, DL tools have been applied for super-resolution with the goal of obtaining conventional endoscopy-like images from a capsule endoscope [144]. In this study, a conditional GAN was implemented with spatial attention blocks, using a loss function that included contributions of pixel loss, content loss, and texture loss. The intuition behind the incorporation of spatial attention blocks is that this module guides the network to prioritize the estimation of the suspicious and diagnostically relevant regions. This study also performed ablation studies and found that the content and texture loss components are especially important for estimating high-spatial frequency patterns, which becomes more important for larger upsampling ratios. With this framework, the resolution of small bowel images was increased by up to 12 $\times$  with favorable quantitative metrics as well as qualitative assessment by gastroenterologists. Though this study demonstrated that the resolution of gastrointestinal images could be enhanced, it remains to be seen if preprocessing or enhancing these images provides any benefit to automated image analysis.

**3D imaging and mapping.** The 3D shape of the tissue being imaged via endoscopy is useful for improving navigation, lesion detection, and diagnosis, as well as obtaining meaningful quality metrics for the effectiveness of the procedure [145]. However, stereo and time-of-flight solutions are challenging and expensive to implement in an endoscopic form factor. Accordingly, there has been significant work in estimating the 3D shape of an

endoscopic scene from monocular images using conditional GANs trained with photo-realistic synthetic data [146,147]. Domain adaptation can be used to improve the generalizability of these models, either by making the synthetic data more realistic, or by making the real images look more like the synthetic data that the depth-estimator is trained on [148]. Researchers have also explored joint conditional random fields and CNNs in a hybrid graphical model to achieve state-of-the-art monocular depth estimation [149]. A U-Net style architecture has been implemented for simultaneously estimating depth, color, and oxygen saturation maps from a fiber-optic probe that sequentially acquired structured light and hyperspectral images [150]. Lastly, DL tools have been applied to improve simultaneous localization and mapping (SLAM) tasks in endoscopic applications, both by incorporating a monocular depth estimation prior into a SLAM algorithm for dense mapping of the gastrointestinal tract [142], and by developing a recurrent neural network to predict depth and pose in a SLAM pipeline [151].

**Widefield spectroscopy.** In addition to efforts to reconstruct high-quality color and 3D maps through an endoscope, DL is also being applied to estimate bulk tissue optical properties from wide FOV images. Optical property mapping can be useful for meeting clinical needs in wound monitoring, surgical guidance, minimally-invasive procedures, and endoscopy. A major challenge to estimating optical properties in turbid media is decoupling the effects of absorption, scattering, and the scattering phase function, which all influence the widefield image measured with flood illumination. SFDI can provide additional inputs to facilitate solving this inverse problem by measuring the attenuation of different spatial frequencies [152]. Researchers have demonstrated that this inverse model can be solved orders of magnitude faster than conventional methods with a 6-layer Perceptron [153]. Others have shown that tissue optical properties can be directly estimated from structured light images or widefield illumination images using content-aware conditional GANs [154]. In this application, the adversarial learning framework reduced errors in the optical property predictions by more than half when compared to the same network trained with an analytical loss function. Intuitively, the discriminator learns a more sophisticated and appropriate loss function in adversarial learning, allowing for the generation of more-realistic optical property maps. Moreover, this study found that the conditional GANs approach resulted in an increased performance benefit when data is tested from tissue types that were not spanned in the training set. The authors hypothesize that this observation comes from the discriminator preventing the generator from learning from and overfitting to the context of the input image. Optical properties can also be estimated more quickly using a lighter-weight twin U-Net architecture with a GPU-optimized look-up table [103]. Further, chromophores can be computed in real-time with

reduced error compared with an intermediate optical property inference by directly computing concentrations from structured illumination at multiple wavelengths using conditional GANs [155].

Going beyond conventional color imaging, researchers are also processing 1D hyperspectral measurement through an endoscope using shallow CNNs to classify pixels into the correct color profiles, illustrating the potential to classify tissue with complex absorbance spectra [156]. The spectral resolution can be increased in dual-modality color/hyperspectral systems from sparse spectral signals with CNNs [150]. To enable quantitative spectroscopy measurements in endoscopic imaging, it may be necessary to combine hyperspectral techniques with structured illumination and 3D mapping [103,150,154,157].

**Future directions.** Future research in endoscopy and DL will undoubtedly explore clinical applications. Imaging system for guiding surgery are already demonstrating clinical potential for ex-vivo tissue classification: a modified Inception-v4 CNNs was demonstrated to effectively classify squamous cell carcinoma versus normal tissue at the cancer margin from ex-vivo hyperspectral images with 91 spectral bands [158]. For *in vivo* applications, where generalizability may be essential and training data may be limited, future research in domain transfer [148] and semi-supervised learning [159] may become increasingly important. Moreover, for clinical validation, these solutions must be real-time, easy-to-use, and robust, highlighting the need for efficient architectures [103] and thoughtful user interface design [160].

### Optical Coherence Tomography (OCT)

**Overview.** OCT is a successful example of biophotonic technological translation into medicine [161,162]. Since its introduction in 1993, OCT has revolutionized the standard-of-care in ophthalmology around the world, and continued thriving in technical advances and other clinical applications, such as dermatology, neurology, cardiology, oncology, gastroenterology, gynecology, and urology [163–170].

**Image segmentation.** The most common use of OCT is to quantify structural metrics via image segmentation, such as retinal anatomical layer thickness, anatomical structures, and pathological features. Conventional image processing is challenging in the case of complex pathology where tissue structural alteration can be complex and may not be fully accounted for when designing a rigid algorithm. Image segmentation is the earliest application of DL explored in OCT applications. Several DNNs have been reported for OCT segmentation in conjunction with manual annotations (Fig. 7a), including U-Net [171–173], ResNet [174], and fully convolutional network (FCN) [175,176]. Successful implementation of DNNs have been broadly reported in different tissues beyond the eye [177–179]. In all areas of applications, the DNN showed superior segmentation accuracy over conventional techniques. For example, Devalla et al. [174] quantified

the accuracy of the proposed DRUNET (Dilated-Residual U-Net) for segmenting the retinal nerve fiber layer (RNFL), retinal Layers, the retinal pigment epithelium (RPE), and choroid on both healthy and glaucoma subjects, and showed that the DRUNET consistently outperformed alternative approaches on all the tissues measured by dice coefficient, sensitivity, and specificity. The errors of all the metrics between DRUNET and the observers were within 10% and the patch-based neural network always provided greater than 10% error irrespective of the observer chosen for validation. In addition, the DRUNET segmentation further allowed automatic extraction of six clinically relevant neural and connective tissue structural parameters, including the disc diameter, peripapillary RNFL thickness (p-RNFLT), peripapillary choroidal thickness (p-CT), minimum rim width (MRW), prelaminar thickness (PLT), and the prelaminar depth (PLD).

**Denosing and speckle removal.** OCT images suffer from speckle noise due to coherent light scattering, which leads to image quality degradation. There exist other sources of noise to further degrade the image quality when the signal level is low. Denoising and despeckling are important applications of DNNs, which are often trained with the averaged reduced-noise image as the “ground truth” in a U-Net and ResNet [182,183]. GAN has also been applied and provided improved visual perception than the DNNs trained with only the least-squares loss function [180] (Fig. 7b). For example, Dong et al. [180] showed that the GAN-based denoising network outperformed state-of-the-art image processing based (e.g., BM3D and MSBTD) and a few other DNNs (e.g., SRResNet and SRGAN) in terms of contrast- $\hat{R}$ to- $\hat{R}$ noise ratio (CNR) and peak signal- $\hat{R}$ to- $\hat{R}$ noise ratio (PSNR).

**Clinical diagnosis and classification.** In clinical applications using DL, a large body of literature over the past 3 years emerges particularly in ophthalmology. Most of the studies use a CNN to extract image features for diagnosis and classification [184]. A clear shift of attention recently is to interpret the DNN, for example using the attention map [181,185] (Fig. 7c). The purpose is to reveal the most important structural features that the DNN used for making the predictions. This addresses the major concern from the clinicians on the “black-box” nature of DL. Another emerging effort is to improve the generalization of a trained DNN to allow process images from different devices, with different image qualities and other possible variations. Transfer learning has been reported to refine pre-trained DNNs to other data set, with much reduced training and data burdens [186,187]. Domain adaptation is another method to generalize the DNN trained on images taken by one device to another [188,189]. We expect more innovations for addressing the generalization in clinical diagnosis and prediction.

**Emerging applications.** Beyond segmentation, denoising, and diagnosis/classification, there are several emerging DL applications for correlating the OCT

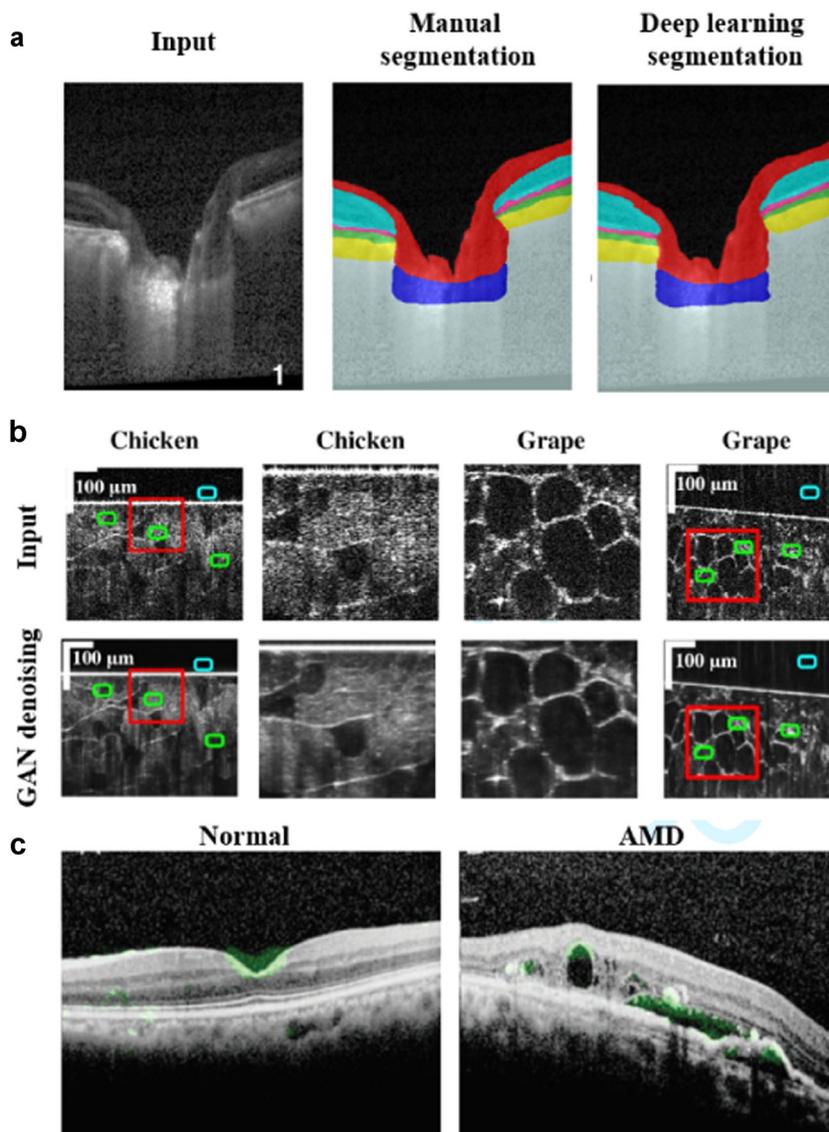


Fig. 7. (a) Example automatic retinal layer segmentation using DL compared to manual segmentation (reprinted from [174]). (b) GAN for denoising OCT images (adapted from [180]). (c) Attention map overlaid with retinal images indicated features that CNN used for diagnosing normal versus age-related macular degeneration (AMD) [181] (Reproduced from Rim et al. [181], with permission from BMJ Publishing Group Ltd.). DL, deep learning; GAN, generative adversarial network; OCT, optical coherence tomography.

measurements with vascular functions. OCT angiography (OCTA) and Doppler OCT (DOCT) are two advanced methods to measure label-free microangiography and blood flows. While normally requiring specific imaging protocols, the raw OCT measurements contain structural features that may be recognized by a CNN. Reports have shown that angiographic image and blood flows can be predicted by mere structural image input without specific OCTA or DOCT protocols [190–192]. For example, Braaf et al. [190] showed that DL enabled accurate quantification of blood flow from OCT intensity time-series measurements, and was robust to vessel angle, hematocrit levels, and measurement SNR.

This is appealing for generating not only anatomical features, but also functional readouts using the simplest OCT imaging protocols by any regular OCT device (Fig. 8a). Recent work also reports the use of a fully connected network and a CNN to extract the spectroscopic information in OCT to quantify the blood oxygen saturation ( $sO_2$ ) within microvasculature, as an important measure of the perfusion function [193] (Fig. 8b and c). The DL models in [193] demonstrated more than 60% error reduction for predicting  $sO_2$  as compared with the standard nonlinear least-squares fitting method. These advances present emerging directions of DL applied to OCT to extract functional metrics beyond structures.

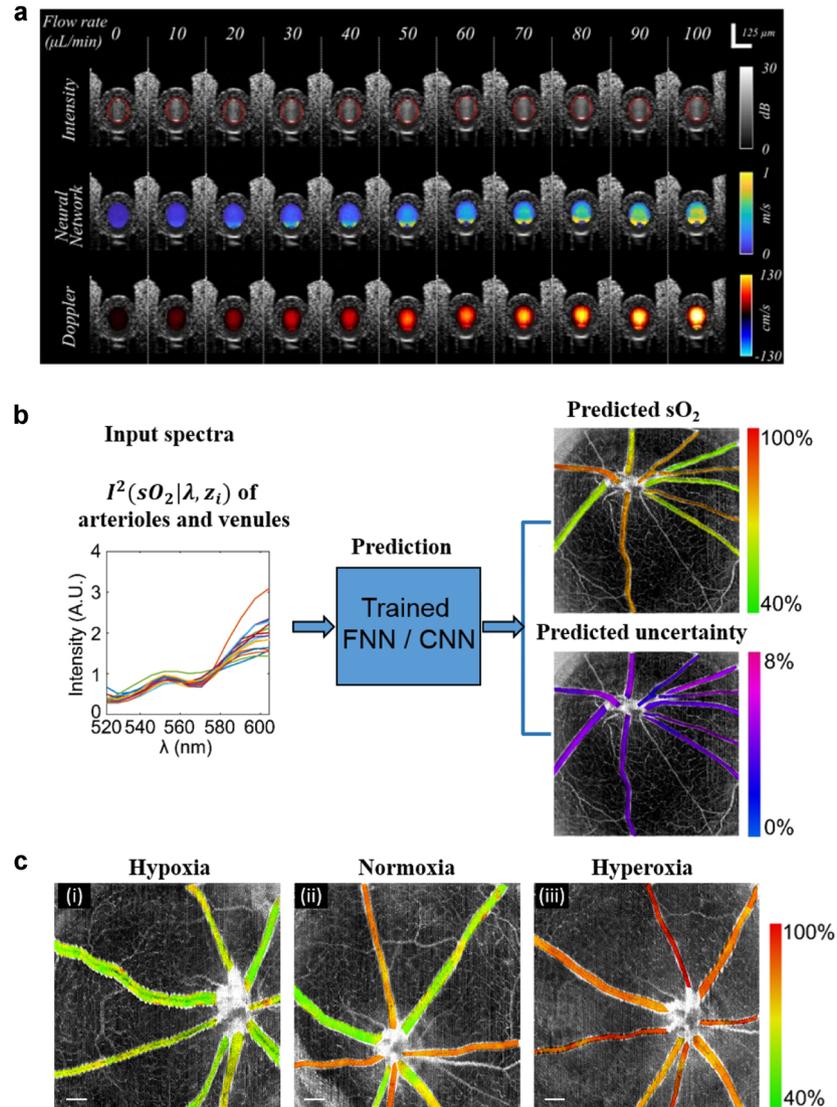


Fig. 8. (a) Examples of using DL to predict blood flow based on structural OCT image features (reprinted from [190]). (b) Example of deep spectral learning for label-free oximetry in visible light OCT (reprinted from [193]). (c) The predicted blood oxygen saturation and the tandem prediction uncertainty from rat retina *in vivo* in hypoxia, normoxia and hyperoxia (reprinted from [193]). DL, deep learning; OCT, optical coherence tomography.

## Photoacoustic Imaging and Sensing

**Overview.** Photoacoustic imaging relies on optical transmission, followed by sensing of the resulting acoustic response [194,195]. This response may then be used to guide surgeries and interventions [196,197] (among other possible uses [198]). In order to guide these surgeries and interventions, image maps corresponding to structures of high optical absorption must be formed, which is a rapidly increasing area of interest for the application of DL to photoacoustic imaging and sensing. This section focuses on many of the first reports of DL for photoacoustic source localization, image formation, and artifact removal. Techniques applied after an image has been formed (e.g., segmentation, spectral unmixing, and quantitative

imaging) are also discussed, followed by a summary of emerging applications based on these DL implementations.

**Source localization.** Localizing sources correctly and removing confusing artifacts from raw sensor data (also known as channel data) are two important precursors to accurate image formation. Three key papers discuss the possibility of using DL to improve source localization. Reiter and Bell [199] introduced the concept of source localization from photoacoustic channel data, relying on training data derived from simulations based on the physics of wave propagation. Allman et al. [200] built on this initial success to differentiate true photoacoustic sources from reflection artifacts based on wavefront shape

appearances in raw channel data. Waves propagating spherically outward from a photoacoustic source are expected to have a unique shape based on distance from the detector, while artifacts are not expected to preserve this shape-to-depth relationship [200]. A CNN (VGG-16) was trained to demonstrate this concept, with initial results shown in Figure 9. Johnstonbaugh et al. [201] expanded this concept by developing an encoder-decoder CNN with custom modules to accurately identify the origin of photoacoustic wavefronts inside an optically scattering deep-tissue medium. In the latter two papers [200,201], images were created from the accurate localization of photoacoustic sources.

**Image formation.** Beyond source localization, DL may also be used to form photoacoustic images directly from raw channel data with real-time speed [202,203]. This section summarizes the application of DL to four technical challenges surrounding image formation: (i) challenges surrounding the limited view of transducer arrays [204–206] (in direct comparison to what is considered the “full view” provided by ring arrays), (ii) sparse sampling of photoacoustic channel data [205,207–209], (iii) accurately estimating and compensating for the fluence differences surrounding a photoacoustic target of interest [210], and (iv) addressing the traditional limited bandwidth issues associated with array detectors [211].

**Limited view:** Surgical applications often preclude the ability to completely surround a structure of interest.

Historically, ring arrays were introduced for small animal imaging [212]. While these ring array geometries can also be used for *in vivo* breast cancer detection [213] or osteoarthritis detection in human finger joints [214], a full ring geometry is often not practical for many surgical applications [197]. The absence of full ring arrays often leads to what is known as “limited view” artifacts, which can appear as distortions of the true shape of circular targets or loss in the appearance of the lines in vessel targets.

DL has been implemented to address these artifacts and restore our ability to interpret the true structure of photoacoustic targets. For example, Hauptmann et al. [204] considered backprojection followed by a learned denoiser and a learned iterative reconstruction, concluding that the learned iterative reconstruction approach sufficiently balanced speed and image quality, as demonstrated in Figure 10. To achieve this balance, a physical model of wave propagation was incorporated during the gradient of the data fit and an iterative algorithm consisting of several CNNs was learned. The network was demonstrated for a planar array geometry. Tong et al. [205] learned a feature projection, inspired by the AUTOMAP network [215], with the novelty of incorporating the photoacoustic forward model and universal backprojection model in the network design. The network was demonstrated for a partial ring array.

**Sparse sampling:** In tandem with limited view constraints, it is not always possible to sufficiently sample

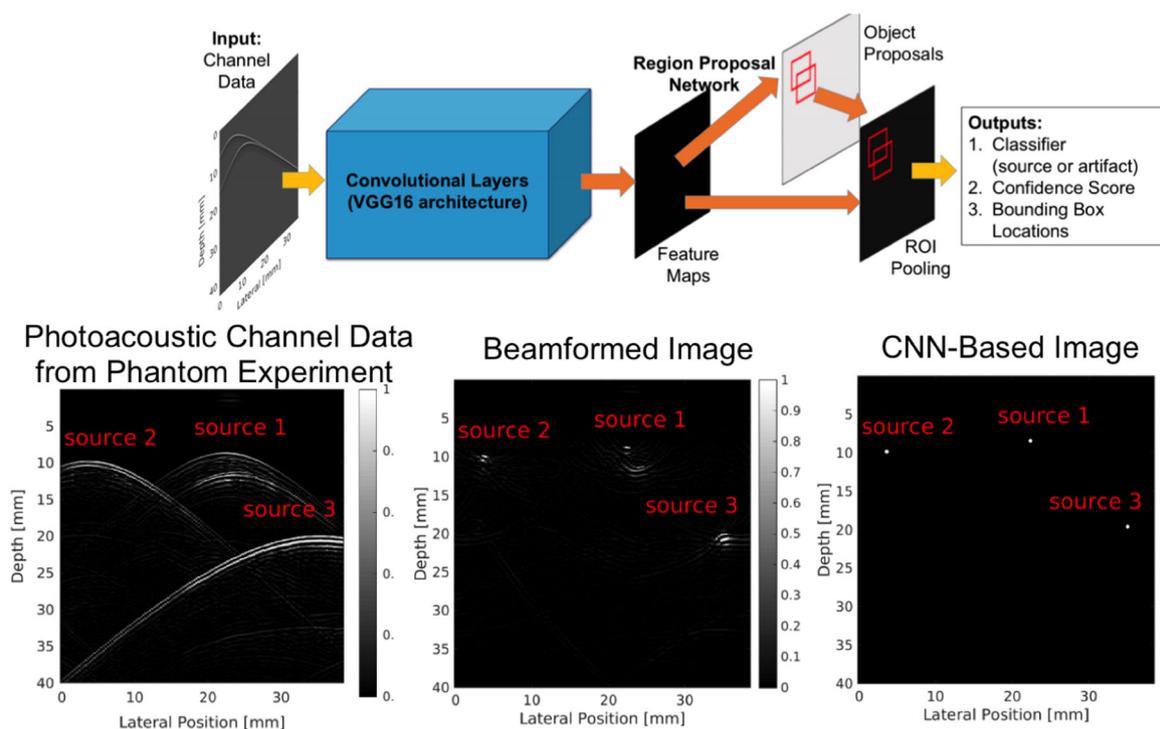


Fig. 9. Example of point source detection as a precursor to photoacoustic image formation after identifying true sources and removing reflection artifacts, modified from [200]. (©2018 IEEE. Adapted, with permission, from Allman et al. [200]).

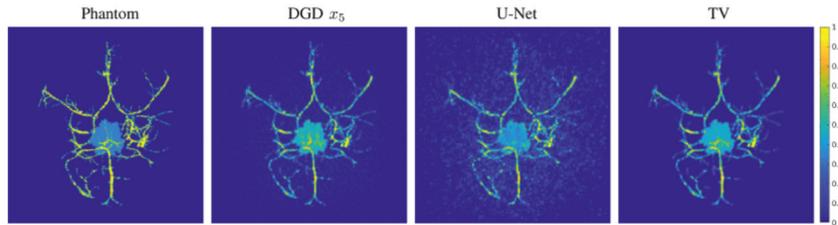


Fig. 10. Example of blood vessel and tumor phantom results with multiple DL approaches. (Reprinted from [204].) DL, deep learning.

an entire region of interest when designing photoacoustic detectors, resulting in sparse sampling of photoacoustic responses. This challenge may also be seen as an extension of limited view challenges, considering that some of the desired viewing angles or spatial locations are missing (i.e., limited) due to sparse sampling, which often results in streak artifacts in photoacoustic images [216]. Therefore, networks that address limited view challenges can simultaneously address sparse sampling challenges [205,217].

Antholzer et al. [207] performed image reconstruction to address sparse sampling with a CNN, modeling a filtered backprojection algorithm [218] as a linear pre-processing step (i.e., the first layer), followed by the U-Net architecture to remove undersampling artifacts (i.e., the remaining layers). Guan et al. [208] proposed pixel-wise DL (Pixel-DL) for limited-view and sparse PAT image reconstruction. The raw sensor data was first interpolated to window information of interest, then provided as an input to a CNN for image reconstruction. In contrast to previously discussed model-based approaches [204,207], this approach does not learn prior constraints from training data and instead the CNN uses more information directly from the CNN and sensor data to reconstruct an image. This utilization of sensor data directly shares similarity with source localization methods [196,200,201].

The majority of methods discussed up until this point have used simulations in the training process for photoacoustic image formation. Davoudi et al. [209] take a different approach to address sparse sampling challenges by using whole-body *in vivo* mouse data acquired with a high-end, high-channel count system. This approach also differs from previously discussed approaches by operating solely in the image domain (i.e., rather than converting sensor or channel data to image data).

**Fluence correction:** The previous sections address challenges related to sensor spacing and sensor geometries. However, challenges introduced by the laser and light delivery system limitations may also be addressed with DL. For example, Hariri et al. [210] used a multi-level wavelet-CNN to denoise photoacoustic images acquired with low input energies, by mapping these low fluence illumination source images to a corresponding high fluence excitation map.

**Limited transducer bandwidth:** The bandwidth of a photoacoustic detector determines the spatial frequencies that can be resolved. Awasthi et al. [211]

developed a network with the goal of resolving higher spatial frequencies than those present in the ultrasound transducer. Improvements were observable as better boundary distinctions in the presented photoacoustic data. Similarly, Gutta et al. [219] used a DNN to predict missing spatial frequencies.

**Segmentation.** After photoacoustic image formation is completed, an additional area of interest has been segmentation of various structures of interest, which can be performed with assistance from DL. Moustakidis et al. [220] investigated the feasibility of DL to segment and identify skin layers by using pretrained models (i.e., ResNet50 [221] and AlexNet [222]) to extract features from images and by training CNN models to classify skin layers directly using images, skipping the processing, transformation, feature extraction, and feature selection steps. These DL methods were compared to other ML techniques. Boink et al. [223] explored simultaneous photoacoustic image reconstruction and segmentation for blood vessel networks. Training was based on the learned primal-dual algorithm [224] for CNNs, including spatially varying fluence rates with a weighting between imaging reconstruction quality and segmentation quality.

**Spectral unmixing and quantitative imaging.** Photoacoustic data and images may also be used to determine or characterize the content of identified regions of interest based on data obtained from a series of optical wavelength excitations. These tasks can be completed with assistance from DL.

Cai et al. [225] proposed a DL framework for quantitative photoacoustic imaging, starting with the raw sensor data received after multiple wavelength transmissions, using a residual learning mechanism adopted to the U-Net to quantify chromophore concentration and oxygen saturation.

**Emerging applications.** Demonstrated applications for image formation with DL has spanned multiple spatial scales, with applications that include cellular-level imaging (e.g., microscopy [226], label-free histology), molecular imaging (e.g., low concentrations of contrast agents *in vivo* [210]), small animal imaging [209], clinical and diagnostic imaging, and surgical guidance [196]. In addition to applications for image formation, other practical applications in photoacoustic imaging and sensing include neuroimaging [208,227], dermatology

(e.g., clinical evaluation, monitoring, and diagnosis of diseases linked to skin inflammation, diabetes, and skin cancer [220]), real-time monitoring of contrast agent concentrations, microvasculature, and oxygen saturation during surgery [202,225], and localization of biopsy needle tips [228], cardiac catheter tips [228–230] or prostate brachytherapy seeds [200].

### Diffuse Tomography

**Overview.** Diffuse optical tomography (DOT), fluorescence diffuse optical tomography (fDOT, also known as fluorescence molecular tomography [FMT]), and bioluminescence tomography (BLT) are non-invasive and non-ionizing 3D diffuse optical imaging techniques [231]. They are all based on acquiring optical data from spatially resolved surface measurements and performing similar mathematical computational tasks that involve the modeling of light propagation according to the tissue attenuation properties. In DOT, the main biomarkers are related to the functional status of tissues reflected by the total blood content (HbT) and relative oxygen saturation (StO<sub>2</sub>) that can be derived from the reconstructed absorption maps [232]. DOT has found applications in numerous clinical scenarios including optical mammography [233,234], muscle physiology [235], brain functional imaging [236], and peripheral vascular diseases monitoring. In fDOT, the inverse problem aims to retrieve the effective quantum yield distribution (related to concentration) of an exogenous contrast agent [237,238] or reporter gene in animal models [239] while illuminated by excitation light. In BLT, the goal is to retrieve the location and strength of an embedded bioluminescent source.

The highly scattering biological tissues lead to ill-posed nonlinear inverse problems that are highly sensitive to model mismatch and noise amplification. Therefore, tomographic reconstruction in DOT/fDOT/BLT is often performed via iterative approaches [240] coupled with regularization. Moreover, the model is commonly linearized using the Rytov (DOT) or Born (DOT/fDOT) methods [241]. Additional constraints, such as preconditioning [242] and *a priori* information are implemented [243–246]. Further experimental constraints in DOT/fDOT are also incorporated using spectral and temporal information [247]. Despite this progress, the implementation and optimization of a regularized inverse problem is complex and requires vast computational resources. Recently, DL methods have been developed for DOT/fDOT/BLT to either aid or fully replace the classical inverse solver. These developments have focused on two main approaches, including (i) learned denoisers and (ii) end-to-end solvers.

**Learned denoisers.** Denoisers can enhance the final reconstruction by correcting for errors from model mismatch and noise amplification. Long [248] proposed a 3D CNN for enhancing the spatial accuracy of mesoscopic FMT outputs. The spatial output of a Tikhonov regularized inverse solver was translated into

a binary segmentation problem to reduce the regularization-based reconstruction error. The network was trained with 600 random ellipsoids and spheres as it only aimed to reconstruct simple geometries *in silico*. Final results displayed improved “intersection over union” values with respect to the ground truth. Since denoising approaches still involve inverting the forward model, it can still lead to large model mismatch. Hence, there has been great interest in end-to-end solutions that directly map the raw measurements to the 3D object without any user input.

**End-to-end solvers.** Several DNNs have been proposed to provide end-to-end inversion. Gao et al. [249] proposed a MLP for BLT inversion for tumor cells, in which the boundary measurements were inputted to the first layer that has a similar number of surface nodes as a standardized mesh built using MRI and CT images of the mouse brain, and output the photon distribution of the bioluminescent target. Similarly, Guo et al. [250] proposed -IJ3D-En-Decoder- $\bar{I}$ , a DNN for FMT with the encoder-decoder structure that inputs photon densities and outputs the spatial distribution of the fluorophores. It was trained with simulated FMT samples. Key features of the measurements were extracted in the encoder section and the transition of boundary photon densities to fluorophore densities was accomplished in the middle section with a fully connected layer. Finally, a 3D-Decoder outputted the reconstruction with better accuracy than L1-regularized inversion method in both simulated and phantom experiments.

Huang et al. [251] proposed a similar CNN approach. After feature encoding, a -IJGated Recurrent unit (GRU)- $\bar{I}$  combines all the output features in a single vector, and the MLP (composed of two hidden layers with dropout and ReLu activations) outputs the fluorophores’ location. Simulated samples of a mouse model (with five organs and one fluorescent tumor target) were used. *In silico* results displayed comparable performance to an L1 inversion method. It was also validated with single-embeddings *in silico* by outputting only the positions since the network does not support 3D rendering. Yoo et al. [252] proposed an encoder-decoder DNN to invert the Lippmann-Schwinger integral photon equation for DOT using the deep convolutional framelet model [253] and learn the nonlinear scattering model through training with diffusion-equation based simulated data. Voxel domain features were learned through a fully connected layer, 3D convolutional layers and a filtering convolution. The method was tested in biomimetic phantoms and live animals with absorption-only contrast. Figure 11 shows an example reconstruction for an *in vivo* tumor in a mouse inside water/milk mixture media.

**Summary and future challenges.** DL has been demonstrated for improving (f)DOT image formation for solving complex ill-posed inverse problems. The DL models are often trained with simulated data, and in a few cases, validated in simple experimental settings.

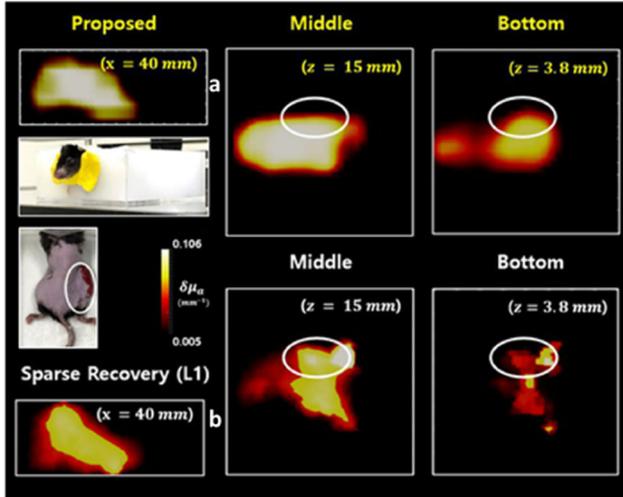


Fig. 11. Reconstruction for a mouse with tumor (right thigh) where higher absorption values are resolved (slices at  $z = 15$  and  $3.8$  mm) for the tumor area with the DNN in (a) compared with the L1-based inversion in (b). (Adapted with permission from the authors from [252]).

With efficient and accurate light propagation platforms such as MMC/MCX [254,255], model-based training could become more efficient. Still, it is not obvious that such DL approaches will lead to universal solutions in DOT/FMT since many optical properties of tissues are still unknown and/or heterogeneous. Hence, further studies should aim to validate the universality of the architectures across different tissue conditions.

### Functional Optical Brain Imaging

**Overview.** Functional optical brain imaging provides the opportunity to correlate neurobiological biomarkers with human behaviors, which impacts numerous fields, such as basic neuroscience, clinical diagnostics, brain computer interface (BCI), and social sciences. The two main established human functional optical brain imaging approaches are functional near-infrared spectroscopy (fNIRS) and diffuse correlation spectroscopy (DCS), both of which report brain activations via monitoring changes in optical signals as light reflected back to the detector while traveling through cortical areas.

Classical neuroimaging studies are based on statistical analysis of biomarkers from a large group of subjects with different statuses (resting/active, stimuli/non-stimuli, disease/disease-free, etc.). However, the derivation of the biomarkers of interests are associated with data processing workflows that can be complex and computationally intensive. While numerous applications in neuroimaging inherently focus on classification of subjects based on spatiotemporal features, DL methods have two outstanding benefits. First, there is the potential to extract meaningful features from high-dimensional noisy data without expert knowledge required for the input/output mapping. Second, DL methods enable statistical inference at the single subject level which is critical for

clinical practice. Hence, there has been a recent surge in DL solutions for functional optical brain imaging.

**Classification based on cortical activations.** Most DL applications to functional optical brain imaging have focused on classification tasks based on fNIRS. Hiroyasu et al. [256] reported a DNN to perform gender classification on subjects performing a numerical memory task while subjecting to a white-noise sound environment to elicit gender-based differences in cortical activations. Using time series data of oxygenated hemoglobin of the inferior frontal gyrus on the left side of the head captured by 4 fNIRS channels, they reported a 81% accuracy in gender classification. The learned classifier identified the inferior frontal gyrus and premotor areas provide the highest discrimination accuracy. Mirbagheri et al. [257] developed a DNN to predict stress using fNIRS data collected on the prefrontal cortex regions, demonstrating 88% accuracy when following the Montreal Imaging Stress Task (MIST) protocols [258].

DL has also been used on fNIRS data for diagnostic and therapeutic applications [259]. Rosas-Romero et al. [260] developed a CNN to predict epileptic seizure (Fig. 12) and reported accuracy ranging between 97% and 100%, sensitivity between 95% and 100% and specificity between 98% and 100% using both oxy- and deoxy-hemoglobin time series as the input. Electroencephalography (EGG) data were acquired simultaneously, but fNIRS predictive features outperformed EGG predictive features. Another use of fNIRS is in psychological studies. Bandara et al. [261] reported a CNN with Long Short Term Memory (LSTM) to analyze spatiotemporal oxy- and deoxy- hemodynamics data from the prefrontal cortex for classifying human emotions and achieved 77% accuracy using both oxy- and deoxy-hemoglobin data and 1-second time steps. These results demonstrate that spatiotemporal features are desired for fNIRS-based classification tasks, and the DL methods excel in feature extraction in such high dimensional data sets. However, all the reported studies followed well-defined protocols that are prevalent in neuroimaging studies but are not always conducive for real-world applications.

Another thrust in fNIRS study is to evaluate mental workload from human computer interaction (HCI) in scenarios, such as driving, air traffic control, and surgery. Benerradi et al. [262] reported a CNN for classifying mental workload using fNIRS data from HCI tasks and achieved an accuracy of 72.77% for 2 classes and 49.53% for 3 classes. The CNN was benchmarked against logistic regression and SVM, but no particular improvements were noted. Gao et al. [263] reported a BRAIN-Net to predict surgical skill levels within the Fundamental of Laparoscopic Surgery (FLS) program environment, demonstrating a ROC-AUC of 0.91 in predicting the FLS Score using fNIRS data collected on the prefrontal cortex of medical students performing the FLS pattern cutting task. BRAIN-Net outperformed classical ML techniques, including Kernel Partial Least Squares (KPLS), nonlinear

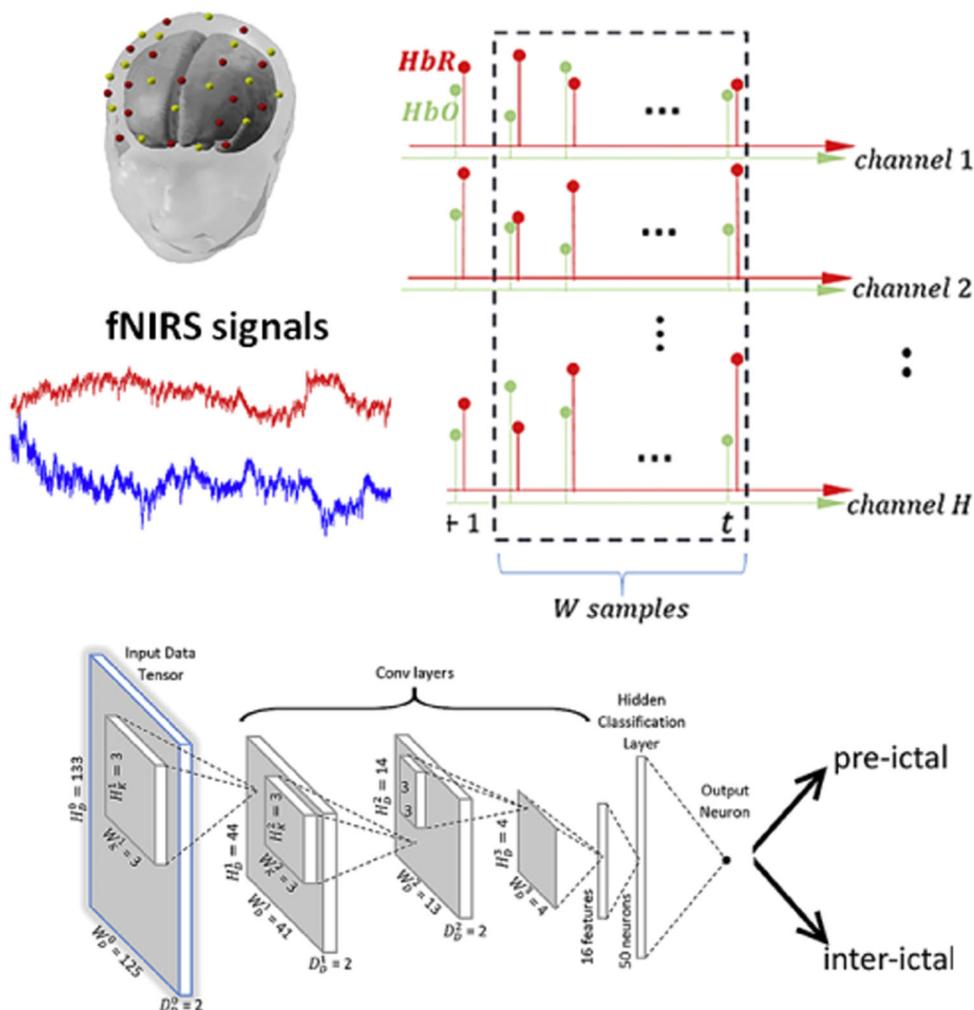


Fig. 12. Hemodynamic time series for prediction of epileptic seizure using a CNN (with permission from Romero et al. [260]) (Copyright (2020), with permission from Elsevier). fNIRS, functional near-infrared spectroscopy.

SVM, and Random Forest, when the data was larger than 600 samples. These results demonstrated the potential of DL for behavior prediction as reported by well-established metrics with freely mobile and unconstrained subjects performing challenging bimanual tasks. Hence, DL-enabled fNIRS methods have the potential for impacting real-world applications. In this regard, one of the most exciting applications of neuroimaging is BCI.

**Brain computer interface.** DL is expected to advance BCI [264]. To date, DL methods for BCI have mainly focused on EGG and to a lesser extent to Magnetic resonance imaging (MRI) or Electromyography (EMG). About 70% of the current work use CNN as discriminative models, 20% use Recurrent neural network (RNN) [106], while generative models (e.g., GAN or VAE) are rarely employed. Impressive results have been reported for real time control of a robot arm using DL-based BCI [265]. Following these trends, a few studies have been reported on DL-enabled fNIRS BCI. Hennrich et al. [266] reported

a DNN that offered similar accuracy as compared to conventional methods in mental task classification. Dargazany et al. [267] implemented a CNN to recognize activity response in fNIRS data for five different activities and reported a 77-80% accuracy in classifying these tasks. Trakoolwilaiwan et al. [268] developed a CNN to classify between rest, right- and left-hand motor execution tasks and achieved classification accuracy within 82-99% depending on the specific subject, which was 6.49% more accurate than SVM and 3.33% more accurate than ANN. As BCI is a challenging task due to noisy data, one current research direction is the implementation of multimodal systems, especially EGG-fNIRS systems, for improved performance. Saadati et al. [269] reported a DNN for processing multimodal input from the variation of oxy- and deoxy-hemoglobin from fNIRS and the event-related desynchronization (ERD) from EGG, achieving the highest accuracy when compared to methods using a single biomarker and 92% accuracy for the word generation task compared to 86% for SVM.

**Denosing and fast data processing.** Data preprocessing in optical neuroimaging is critical and includes dynamic range correction, transforming light attenuation to chromophore concentration, regressing shallow hemodynamic response to increase the sensitivity to cortical tissues, identifying and removing noise, especially motion artifacts. These steps typically require user inputs and are computationally intensive. Gao et al. [270] demonstrated a DNN for suppressing motion artifacts in raw fNIRS signals and identified 100% of the motion artifacts almost in real time. Poon et al. [271] reported a DNN in DCS that was 23× faster in estimating the tissue blood flow index compared to the traditional nonlinear fitting method. Hence, DL methodologies may facilitate the adoption of DCS for neuroimaging studies by enabling real-time and accurate tissue blood flow quantification in deep tissues.

**Future directions and associated challenges.** DL methods herald the potential for subject-specific classification on the fly, leading to fast and direct feedback based on real-time monitoring of brain functions. It also has potential for neuro-feedback in numerous therapeutic scenarios or cognitive/skill learning programs. In addition, DL has been largely adopted in brain connectivity studies [272], which has become prevalent for deciphering the brain circuitry [273] and diagnostic purposes [274]. Similar to MRI [275], DL is expected to play a critical role in next generation functional brain connectivity studies [276]. Still, numerous challenges lie ahead to implement full end-to-end solutions in data processing and classification.

One main challenge is the size of the population needed for generating the data sets. As we are still far from being able to model the complexity of brain functions and dynamics, this challenge is complicated by the need to train and validate neuroimaging DL approaches with experimental data. In numerous fNIRS and DCS studies, subject recruitment is limited and no public database is readily available. Such limitations have been recognized in all existing work. For such emerging methodologies, great care should be directed to appropriate cross-validation of the DL methods. Hence, validation methods such as k-fold and/or leave-one-out (one refers to one subject out, one trial out, or one day out, etc.) are essential to convey confidence of the usefulness of the methodology [277].

In addition, numerous applications of optical neuroimaging involve environments and tasks that cannot be fully controlled and/or restricted. Thus, brain cortical activations and connectivity can reflect response to complex stimuli in which “ground truth” can be challenging to establish. For example, it would be ideal to use a standardized and accredited metric (i.e., the FLS score) in various neuro-based applications. However, such objective standards do not exist and labeling of the data can be problematic. These challenges also limit the potential of DL for discovery and mapping of the brain circuitry. If DL were to become preponderant in functional connectivity

studies, it also faces the current challenge of being primarily employed in the brain at rest, which does not offer insight into active states of interest.

## CHALLENGES AND OPPORTUNITIES ACROSS MULTIPLE IMAGING DOMAINS

### Challenges

**Data availability and bias.** Most DL models for biomedical optics rely on “supervised learning” that are trained on domain- and/or task-specific data sets, which need to be carefully curated to ensure high-quality predictions. As a result, there are several inherent challenges in the data generation process that need to be addressed, including data availability and data bias [278]. For many applications, it is often difficult and costly to acquire a large-scale data set. Novel techniques that can better leverage small-scale data set while still providing high-quality models are needed, such as unsupervised, semi-supervised, and self-supervised learning, transfer learning, and domain adaptation. In addition to the overall scale of the data set, the data may also be skewed or biased [279] because it may be difficult to acquire data with a balanced distribution for each subgroup, such as gender, ethnicity, etc. DNNs trained on biased data set can result in erroneous predictions in particular for under-represented populations and diseases. These obstacles may be mitigated to some extent with careful planning and data collection. However, there is a need to also identify and reduce data biases in the modeling step, such as data augmentation and balanced loss function design.

**Interpretability.** A common challenge of DL models is that they are generally “black-boxes” and their predictions typically cannot be precisely explained. This is particularly problematic in health applications. To address this issue, “interpretable/explainable” DL techniques [280,281] are needed. To this end, there are two general approaches that are actively being researched in the field [282]. The first is to develop an interpretable computational structure instead of DNNs [283,284], so that the predictions are made based on the crafted logic in the DL model. The second approach is to provide *post hoc* model prediction interpretation, such as attention mechanism [42,285] and uncertainty quantification [20,286,287], while keeping the same DNN structure.

**Prospective and real-world validation.** In general, there is a need for prospective evaluations of DL-based systems in real clinical settings. The performance of DL models are commonly evaluated *post hoc* using metrics often not directly translatable to improving patient care. To critically evaluate the performance and move to clinical impact, these gaps must be bridged. First and foremost, large-scale prospective testing is needed, ideally with multiple sites, users, and instruments. Second, it is also important to develop quantitative metrics to relate those commonly used in DL model development to those most pivotal in improving the management of disease.

## Opportunities

**Exploiting multimodal data.** DNNs provide powerful frameworks for integrating multimodal and multi-dimensional data [15]. Biomedical optics systems often acquire measurements that augment traditional visualization or span a wide range of resolutions, imaging speeds, and sources of contrast. A fundamental barrier to the clinical translation of these technologies is that their benefit must outweigh the cost of additional training and time required to interpret and monitor these data. DL models can efficiently analyze these data together and transform them to actionable representations, reducing these training barriers while increasing the diagnostic power of multimodal imaging.

**Lowering costs.** First, as shown in many examples herein, DL can enable new imaging capabilities that improve resolution, acquisition speed, FOV, and DOF often with minimal hardware modifications. This means that high-quality measurements can increasingly be made using relatively simple and lower cost systems. Second, DL technologies can enable more efficient workflows in healthcare and research, such as digital staining/labeling of tissues to reduce the cost and time associated with sample preparation.

**Deskilling procedures.** Automated data processing and interpretation by DL may reduce the level of skill needed to obtain measurements and provide a diagnosis. A major benefit of DL-based processing is that it is “end-to-end.” This means that once the DNN is trained, it enables automated reconstruction without any additional manual parameter tuning, potentially making it more generalizable and robust than classical approaches. This advantage must be balanced with great care and heightened responsibility to ensure ethical usage and unbiased outputs of these end-to-end DNN algorithms.

**Increasing access to high-quality health care.** The ability of DL to lower cost and training requirements for diagnostic technologies holds tremendous potential for increasing access to high-quality health care in low-resource settings.

## SUMMARY AND OUTLOOK

DL-based techniques have shown promise in addressing various technical challenges for developing novel biomedical optics systems, such as overcoming physical trade-offs, as well as enabling novel capabilities beyond existing solutions. Successful examples are available across multiple imaging domains, including microscopy, fluorescence lifetime imaging, *in vivo* microscopy, wide-field endoscopy, optical coherence tomography, photoacoustic imaging, diffuse tomography, and functional optical brain imaging. Techniques are vast and varied, ranging from providing microscopic sub-cellular information to localizing image sources and offering macroscopic biomarkers. With the advances of DL techniques in many different biomedical optics domains, there are

also some outstanding challenges that must be addressed in order to fully realize the impact of these techniques. As we are rapidly seeing across multiple biomedical optics modalities, DL techniques have promising potential to lower system costs, reduce required skill levels to carry out measurements, and ultimately increase the quality, affordability, and accessibility of health care.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation 1813848, 1846784 (L.T.); the National Institutes of Health R21EB024700 (N.J.D.), R21EB025621 (M.A.L.B), R01NS108464 (J.Y.), R01EB019443, R01CA207725, R01CA237267, and R01CA250636 (X.I., J.T.S., M.O.); MTEC:20-05-IMPROVE-004 and W912CG2120001 (X.I.).

## REFERENCES

1. Yun SH, Kwok SJJ. Light in diagnosis, therapy and surgery. *Nat Biomed Eng* 2017;1:1–16.
2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–444.
3. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning*. Cambridge: MIT Press; 2016.
4. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88. arXiv: 1702.05747.
5. Nichols JA, HerbertChan HW, Baker MAB. Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophys Rev* 2019;11:111–118.
6. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science* 2015;349:255–260.
7. NVIDIA Blog. The difference between AI, machine learning, and deep learning?
8. Karhunen J, Raiko T, Cho K. Chapter 7—Unsupervised deep learning: A short review. In: Bingham E, Kaski S, Laaksonen J, Lampinen J, editors. *Advances in Independent Component Analysis and Learning Machines*. Cambridge, MA: Academic Press; 2015. pp 125–142.
9. Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020.
10. Zhu X, Goldberg AB. *Introduction to Semi-Supervised Learning Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool Publishers. 2009. 3:pp 1–130.
11. Masci J, Meier U, Cireş D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. 2011. pp. 52–59.
12. Falk T, Mai D, Bensch R, et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat Methods* 2019; 16(1):67–70.
13. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. arXiv:1406.2661 [cs, stat]. arXiv: 1406.2661. 2014.
14. Kingma DP, Adam BJ. A method for stochastic optimization. arXiv:1412.6980 [cs]. arXiv: 1412.6980. 2017.
15. Ramachandram D, Taylor GW. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process Mag* 2017;34:96–108.
16. Abadi M, Agarwal A, Barham P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467. 2016.
17. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*. 2019. pp 8026–8037.

18. Weigert M, Schmidt U, Boothe T, et al. Content-aware image restoration: Pushing the limits of fluorescence microscopy. *Nat Methods* 2018;15:1090–1097.
19. Wang H, Rivenson Y, Jin Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat Methods* 2019;16:103–110.
20. Xue Y, Cheng S, Tian L. Reliable deep-learning-based phase imaging with uncertainty quantification. *Optica* 2019;6: 618–629.
21. Wu Y, Rivenson Y, Wang H, et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nat Methods* 2019;16(12):1323–1331.
22. Rivenson Y, Wang H, Wei Z, et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nature Biomed Eng* 2019;3:466–477.
23. Christiansen EM, Yang SJ, Ando DM, et al. In silico labeling: Predicting fluorescent labels in unlabeled images. *Cell* 2018;173:792–803.
24. Jin L, Liu B, Zhao F, et al. Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nat Commun* 2020;11:1934.
25. Lee S, Negishi M, Urakubo H, Kasai H, Ishii S. Mu-net: Multi-scale U-net for two-photon microscopy image denoising and restoration. *Neural Netw* 2020;125:92–103.
26. Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. *Optica* 2017;4: 1437–1443.
27. Liu T, Haan K, Rivenson Y, et al. Deep learning-based super-resolution in coherent imaging systems. *Sci Rep* 2019;9(1):3926.
28. Jin L, Tang Y, Wu Y, et al. Deep learning extended depth-of-field microscope for fast and slide-free histology. *Proc Natl Acad Sci.* 2020;117(52):33051–33060.
29. Nguyen T, Xue Y, LY, Tian L, Nehmetallah G. Deep learning approach for Fourier Ptychography microscopy. *Opt Express* 2018;26:26470–26484.
30. Kellman M, Bostan E, Chen M, Waller L. Data-driven design for Fourier Ptychographic microscopy. arXiv: 1904.04175 [cs, eess]. arXiv: 1904.04175 2019.
31. Ouyang W, Aristov A, Lelek M, Hao X, Zimmer C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnol* 2018;36:460–468.
32. Nehme E, Weiss LE, Michaeli T, Shechtman Y. Deep-STORM: Super-resolution single-molecule microscopy by deep learning. *Optica* 2018;5:458.
33. Nehme E, Freedman D, Gordon R, et al. DeepSTORM3D: Dense 3D localization microscopy and PSF design by deep learning. *Nat Methods* 2020;17:734–740.
34. Zhang Y, Haan K, Rivenson Y, LJ, Delis A, Ozcan A. Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light Sci Appl* 2020;9(1):78.
35. Rivenson Y, Liu T, Wei Z, Zhang Y, Haan K, Ozcan A. PhaseStain: The digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci Appl* 2019;8:23.
36. Nygate YN, Levi M, Mirsky SK, et al. Holographic virtual staining of individual biological cells. *Proc Natl Acad Sci.* 2020;117(17):9223–9231.
37. Borhani N, Bower AJ, Boppert SA, Psaltis D. Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy. *Biomed Opt Express* 2019;10:1339.
38. Rana A, Lowe A, Lithgow M, et al. Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. *JAMA Network Open* 2020;3:e205111.
39. Rivenson Y, Haan K, Wallace WD, Ozcan A. Emerging advances to transform histopathology using virtual staining. *BME Frontiers* 2020.
40. Ounkomol C, Seshamani S, Maleckar MM, Collman F, Johnson GR. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat Methods* 2018;15:917.
41. Guo S-M, Yeh L-H, Folkesson J, et al. Revealing architectural order with quantitative label-free imaging and deep learning. *eLife* 2020;9:e55502.
42. Cheng S, Fu S, Kim YM, et al. Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy. *Sci Adv* 2021;7:eabe0.
43. Kandel ME, He Yuchen R, Lee Young J, et al. Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments. *Nat Commun* 2020;11:6256.
44. Eulenberg P, Köhler N, Blasi T, et al. Reconstructing cell cycle and disease progression using deep learning. *Nat Commun* 2017;8(1):1–6.
45. Doan M, Case M, Masic D, et al. Label-free leukemia monitoring by computer vision. *Cytometry Part A* 2020;97:407–414.
46. You S, Sun Y, Yang L, et al. Real-time intraoperative diagnosis by deep neural network driven multiphoton virtual histology. *NPJ Precis Oncol* 2019;3(1):33.
47. Matsumoto T, Niioka H, Kumamoto Y, et al. Deep-UV excitation fluorescence microscopy for detection of lymph node metastasis using deep neural network. *Sci Rep* 2019;9(1): 16912.
48. Zhang JK, R He Yuchen, N Sobh, Popescu G. Label-free cancer screening using deep learning and spatial light interference microscopy (SLIM). *APL Photonics* 2020;5(4):040805.
49. Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat Machine Intell* 2019;1:538–544.
50. Lippeveld M, Knill C, Ladlow E, et al. Classification of human white blood cells using machine learning for stain-free imaging flow cytometry. *Cytometry Part A* 2020;97: 308–319.
51. Kandel ME, Rubessa M, He Yuchen R. Reproductive outcomes predicted by phase imaging with computational specificity of spermatozoon ultrastructure. *Proc Natl Acad Sci U S A* 2020;117:18302–18309.
52. Buggenthin F, Buettner F, Hoppe PS, et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nat Methods* 2017;14(4):403–406.
53. Kusumoto D, Lachmann M, Kunihiro T, et al. Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells. *Stem Cell Rep* 2018;10:1687–1695.
54. Waisman A, LaGreca A, Möbbs AM, et al. Deep learning neural networks highly predict very early onset of pluripotent stem cell differentiation. *Stem Cell Rep* 2019;12:845–859.
55. Kobayashi H, Lei C, Wu Y, et al. Intelligent whole-blood imaging flow cytometry for simple, rapid, and cost-effective drug-susceptibility testing of leukemia. *Lab on a Chip* 2019;19:2688–2698.
56. Kellman MR, Bostan E, Repina NA, Waller L. Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. *IEEE Trans Comput Imag* 2019;5:344–353.
57. Hershko E, Weiss LE, Michaeli T, Shechtman Y. Multicolor localization microscopy and point-spread-function engineering by deep learning. *Opt Express* 2019;27: 6158–6183.
58. Štefko M, Ottino B, Douglass KM, Manley S. Autonomous illumination control for localization microscopy. *Opt Express* 2018;26:30882–30900.
59. Yang SJ, Berndt Marc, Michael AD, et al. Assessing microscope image focus quality with deep learning. *BMC Bioinform* 2018;19:77.
60. Jiang S, Liao J, Bian Z, Guo K, Zhang Y, Zheng G. Transform- and multi-domain deep learning for single-frame rapid autofocusing in whole slide imaging. *Biomed Opt Express* 2018;9(4):1601.
61. Pinkard H, Phillips Z, Babakhani A, Fletcher DA, Waller L. Deep learning for single-shot autofocus microscopy. *Optica* 2019;6:794–797.
62. Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2Noise: Learning image restoration without clean data. arXiv:1803.04189 [cs, stat]. arXiv: 1803.04189 2018.
63. Krull A, Buchholz T-O, Jug F. Noise2void—Learning denoising from single noisy images. arXiv:1811.10980 [cs]. arXiv: 1811.10980 2018.

64. Batson J, Royer L. Noise2Self: Blind denoising by self-supervision. arXiv:1901.11365 [cs, stat]. arXiv: 1901.11365 2019.
65. Broaddus C, Krull A, Weigert M, Schmidt U, Myers G. Removing structured noise with self-supervised blind-spot networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). 2020. pp 159–163.
66. Tahir W, Kura S, Zhu J, et al. Anatomical modeling of brain vasculature in two-photon microscopy by generalizable deep learning. *BME Frontiers* 2021;2021:8620932.
67. Gur S, Wolf L, Golgher L, Blinder P. Unsupervised microvascular image segmentation using an active contours mimicking neural network. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE. 2019. pp 10721–10730.
68. Bostan E, Heckel R, Chen M, Kellman M, Waller L. Deep phase decoder: Self-calibrating phase microscopy with an untrained deep neural network. *Optica* 2020;7:559–562.
69. Wu Z, Sun Y, Matlock A, Liu J, Tian L, Kamilov US. SIMBA: Scalable Inversion in optical tomography using deep denoising priors. *IEEE J Sel Topics in Signal Process* 2020;14(6):1163.
70. Lakowicz JR, Szmajcinski H, Nowaczyk K, Johnson ML. Fluorescence lifetime imaging of free and protein-bound NADH. *Proc Natl Acad Sci U S A* 1992;89:1271–1275.
71. Das B, Shi L, Budansky Y, Rodriguez-Contreras A, Alfano R. Alzheimer mouse brain tissue measured by time resolved fluorescence spectroscopy using single- and multi-photon excitation of label free native molecules. *J Biophotonics* 2018;11:e201600318.
72. Rudkouskaya A, Sinsuebphon N, Ochoa M, et al. Multiplexed non-invasive tumor imaging of glucose metabolism and receptor-ligand engagement using dark quencher FRET acceptor. *Theranostics* 2020;10:10309–10325.
73. Marcu L. Fluorescence lifetime techniques in medical applications. *Ann Biomed Eng* 2012;40:304–331.
74. Becker W. Fluorescence lifetime imaging—Techniques and applications. *J Microsc* 2012;247:119–136.
75. Datta R, Heaster TM, Sharick JT, Gillette AA, Skala MC. Fluorescence lifetime imaging microscopy: Fundamentals and advances in instrumentation, analysis, and applications. *J Biomed Opt* 2020;25:1.
76. Suhling K, Hirvonen LM, Levitt JA, et al. Fluorescence lifetime imaging (Flim): Basic concepts and recent applications. In: *Advanced Time-Correlated Single Photon Counting Applications*. pp 119–188. Cham: Springer; 2015.
77. Wang M, Tang F, Pan X, et al. Rapid diagnosis and intraoperative margin assessment of human lung cancer with fluorescence lifetime imaging microscopy. *BBA Clin* 2017;8:7–13.
78. Georgakoudi I, Quinn KP. Optical imaging using endogenous contrast to assess metabolic state. *Ann Rev Biomed Eng* 2012;14:351–367.
79. Datta R, Alfonso-García A, Cinco R, Gratton E. Fluorescence lifetime imaging of endogenous biomarker of oxidative stress. *Sci Rep* 2015;5:9848.
80. Lin H-J, Herman P, Lakowicz JR. Fluorescence lifetime-resolved pH imaging of living cells. *Cytometry* 2003; 52A:77–89.
81. Rajoria S, Zhao L, Intes X, Barroso M. FLIM-FRET for cancer applications. *Curr Mol Imag* 2015;3:144–161.
82. Sun Y, Day RN, Periasamy A. Investigating protein-protein interactions in living cells using fluorescence lifetime imaging microscopy. *Nat Protoc* 2011;6:1324–1340.
83. Zadrán S, Standley S, Wong K, Otiniano E, Amighi A, Baudry M. Fluorescence resonance energy transfer (FRET)-based biosensors: Visualizing cellular dynamics and bioenergetics. *Appl Microbiol Biotechnol* 2012;96:895–902.
84. Rudkouskaya A, Sinsuebphon N, Ward J, Tubbesing K, Intes X, Barroso M. Quantitative imaging of receptor-ligand engagement in intact live animals. *J Control Release* 2018;286:451–459.
85. Thaler C, Koushik SV, Blank PS, Vogel SS. Quantitative multiphoton spectral imaging and its use for measuring resonance energy transfer. *Biophys J* 2005;89:2736–2749.
86. Ranjit S, Malacrida L, Jameson DM, Gratton E. Fit-free analysis of fluorescence lifetime imaging data using the phasor approach. *Nat Protoc* 2018;13:1979–2004.
87. Chen S, Sinsuebphon N, Rudkouskaya A, Barroso M, Intes X, Michalet X. In vitro and in vivo phasor analysis of stoichiometry and pharmacokinetics using short-lifetime near-infrared dyes and time-gated imaging. *J Biophotonics* 2019;12:e201800185.
88. Ankri R, Basu A, Ulku AC, et al. Single-photon, time-gated, phasor-based fluorescence lifetime imaging through highly scattering medium. *ACS Photonics* 2020;7:68–79.
89. Wu G, Nowotny T, Zhang Y, Yu HQ, Li DDU. Artificial neural network approaches for fluorescence lifetime imaging techniques. *Opt Lett* 2016;41:2561.
90. Smith JT, Yao R, Sinsuebphon N, et al. Fast fit-free analysis of fluorescence lifetime imaging via deep learning. *Proc Natl Acad Sci U S A* 2019;116:24019–24030.
91. Angelo JP, ChenSez-Jade K, Ochoa M, Sunar U, Gioux S, Intes X. Review of structured light in diffuse optical imaging. *J Biomed Opt* 2018;24:1.
92. Edgar MP, Gibson GM, Padgett MJ. Principles and prospects for single-pixel imaging. *Nat Photonics* 2019; 13:13–20.
93. Pian Q, Yao R, Sinsuebphon N, Intes X. Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging. *Nat Photonics* 2017;11:411–414.
94. Yao R, Ochoa M, Yan P, Intes X. Net-FLICS: Fast quantitative wide-field fluorescence lifetime imaging with compressed sensing—A deep learning approach. *Light Sci Appl* 2019;8:26.
95. Ochoa MM, Rudkouskaya A, Yao R, Yan P, Barroso M, Intes X. High compression deep learning based single-pixel hyperspectral macroscopic fluorescence lifetime imaging in vivo. *Biomed Opt Express* 2020;11(10):5401.
96. Rahman A, Jahangir C, Lynch SM, et al. Advances in tissue-based imaging: Impact on oncology research and clinical practice. *Exp Rev of Mol Diagn* 2020;20:1027–1037.
97. Haraguchi T, Shimi T, Koujin T, Hashiguchi N, Hiraoka Y. Spectral imaging fluorescence microscopy. *Genes Cells*. 2002;7:881–887.
98. Pian Q, Yao R, Zhao L, Intes X. Hyperspectral time-resolved wide-field fluorescence molecular tomography based on structured light and single-pixel detection. *Opt Lett* 2015;40:431.
99. Becker W, Bergmann A, Biskup C. Multispectral fluorescence lifetime imaging by TCSPC. *Microsc Res Tech* 2007;70:403–409.
100. Smith JT, Ochoa M, Intes X. UNMIX-ME: Spectral and lifetime fluorescence unmixing via deep learning. *Biomed Opt Express* 2020;11:3857.
101. Smith JT, Aguénon E, Gioux S, Intes X. Macroscopic fluorescence lifetime topography enhanced via spatial frequency domain imaging. *Opt Lett* 2020;45:4232.
102. Yao R, Intes X, Fang Q. Direct approach to compute Jacobians for diffuse optical tomography using perturbation Monte Carlo-based photon “replay”. *Biomed Opt Express* 2018;9:4588.
103. Aguénon E, Smith JT, Al-Taher M, et al. Real-time, wide-field and high-quality single snapshot imaging of optical properties with profile correction using deep learning. *Biomed Opt Express* 2020;11:5701–5716.
104. Mannam V, Zhang Y, Yuan X, Ravasio C, Howard S. Machine learning for faster and smarter fluorescence lifetime imaging microscopy. *J Phys Photonics* 2020;2:042005.
105. Walsh AJ, Mueller KP, Tweed K, et al. Classification of T-cell activation via autofluorescence lifetime imaging. *Nat Biomed Eng* 2021;5(1):77.
106. Zhang Y, Hato T, Dagher PC, et al. Automatic segmentation of intravital fluorescence microscopy images by K-means clustering of FLIM phasors. *Opt Lett* 2019;44:3928.
107. Sagar MAK, Cheng KP, Ouellette JN, Williams JC, Watters JJ, Eliceiri KW. Machine learning methods for fluorescence lifetime imaging (FLIM) based label-free detection of microglia. *Front Neurosci*. 2020;14:931.
108. Wells WA, Thrall M, Sorokina A, et al. In vivo and ex vivo microscopy: Moving toward the integration of optical imaging technologies into pathology practice. *Arch Pathol Lab Med* 2019;143:288–298.
109. Adhi M, Duker JS. Optical coherence tomography-current and future applications. *Curr Opin Ophthalmol*. 2013;24:213.

110. Goetz M. Confocal laser endomicroscopy: Applications in clinical and translational science—A comprehensive review. *Int Scholarly Research Notices* 2012;2012:387145.
111. Rajadhyaksha M, Marghoob A, Rossi A, Halpern AC, Nehal KS. Reflectance confocal microscopy of skin in vivo: From bench to bedside. *Lasers Surg Med* 2017;49:7–19.
112. Goetz M. Confocal laser endomicroscopy: Current indications and future perspectives in gastrointestinal diseases. *Endoscopy* 2012;24:67–74.
113. Aubreville M, Knipfer C, Oetter N, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. *Sci Rep* 2017;7:1–10.
114. Szczotka AB, Shakir DI, Ravi D, Clarkson MJ, Pereira SP, Vercauteren T. Learning from irregularly sampled data for endomicroscopy super-resolution: A comparative study of sparse and dense approaches. *Int J Comput Assist Radiol Surg* 2020;15:1167–1175.
115. Gu Y, Vyas K, Yang J, Yang G-Z. Transfer recurrent feature learning for endomicroscopy image recognition. *IEEE Trans Med Imag* 2018;38:791–801.
116. Kose K, Bozkurt A, Alessi-Fox C, et al. Segmentation of cellular patterns in confocal images of melanocytic lesions in vivo via a multiscale encoder-decoder network (med-net). *Med Image Anal* 2021;7:101841.
117. Gessert N, Bengs M, Wittig L, et al. Deep transfer learning methods for colon cancer classification in confocal laser microscopy images. *Int J Comput Assist Radiol Surg* 2019;14:1837–1845.
118. Cendre R, Mansouri A, Benezeth Y, Marzani F, Perrot JL, Cinotti E. Two schemes for automated diagnosis of lentigo on confocal microscopy images. In: 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). IEEE. 2019. pp 143–147.
119. Wodzinski M, Skalski A, Witkowski A, Pellacani G, Ludzik J. Convolutional neural network approach to classify skin lesions using reflectance confocal microscopy. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. 2019. pp 4754–4757.
120. Bozkurt A, Gale T, Kose K, et al. Delineation of skin strata in reflectance confocal microscopy images with recurrent convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017. pp 25–33.
121. Li Y, Charalampaki P, Liu Y, Yang G-Z, Giannarou S. Context aware decision support in neurosurgical oncology based on an efficient classification of endomicroscopic data. *Int J Comput Assist Radiol Surg* 2018;13:1187–1199.
122. Lucas M, EIML Liem, Savci-Heijink CD, et al. Toward automated in vivo bladder tumor stratification using confocal laser endomicroscopy. *J Endourol* 2019;33:930–937.
123. Aubreville M, Stoeve M, Oetter N, et al. Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images. *Int J Comput Assist Radiol Surg* 2019;14:31–42.
124. Wodzinski M, Pajak M, Skalski A, Witkowski A, Pellacani G, Ludzik J. Automatic quality assessment of reflectance confocal microscopy mosaics using attention-based deep neural network. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE. 2020. pp 1824–1827.
125. Kose K, Bozkurt A, Alessi-Fox C, et al. Utilizing machine learning for image quality assessment for reflectance confocal microscopy. *J Invest Dermatol* 2020;140:1214–1222.
126. Wallace MB, Fockens P. Probe-based confocal laser endomicroscopy. *Gastroenterology* 2009;136:1509–1513.
127. Shao J, Zhang J, Liang R, Barnard K. Fiber bundle imaging resolution enhancement using deep learning. *Opt Express* 2019;27:15880–15890.
128. Ravi D, Szczotka AB, Pereira SP, Vercauteren T. Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy. *Med Image Anal* 2019;53:123–131.
129. Meng Z, Qiao M, Ma J, Yu Z, Xu K, Yuan X. Snapshot multispectral endomicroscopy. *Optics Lett* 2020;45:3897–3900.
130. Bano S, Vasconcelos F, Tella-Amo M, et al. Deep learning-based fetoscopic mosaicking for field-of-view expansion. *Int J Comput Assist Radiol Surg* 2020;15:1807–1816.
131. Rahmani B, Loterie D, Konstantinou G, Psaltis D, Moser C. Multimode optical fiber transmission with a deep learning network. *Light Sci Appl* 2018;7:1–11.
132. Zhao J, Sun Y, Zhu Z, et al. Deep learning imaging through fully-flexible glass-air disordered fiber. *ACS Photonics* 2018; 5: 3930–3935.
133. Pogue BW. Optics of medical imaging. SPIE Professional. 2018.
134. Wang P, Xiao X, Brown JRG, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* 2018;2(10): 741–748.
135. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94–100.
136. Poon CCY, Jiang Y, Zhang R, et al. AI-doscopist: A real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices. *NPJ Digit Med* 2020;3(1):1–8.
137. LeBerre C, Sandborn WJ, Aridhi S, et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020;158:76–94.
138. Lore KG, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit* 2017;61:650–662.
139. Gómez P, Semmler M, Schützenberger A, Bohr C, Döllinger M. Low-light image enhancement of high-speed endoscopic videos using a convolutional neural network. *Med Biol Eng Comput* 2019;57:1451–1463.
140. Bobrow TL, Mahmood F, Inserni M, Durr NJ. DeepLSR: A deep learning approach for laser speckle reduction. *Biomed Opt Express* 2019;10:2869–2882.
141. Funke I, Bodenstedt S, Riediger C, Weitz J, Speidel S. Generative adversarial networks for specular highlight removal in endoscopic images. In: Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling. International Society for Optics and Photonics. Vol 10576. 2018. p 1057604.
142. Chen RJ, Bobrow TL, Athey T, Mahmood F, Durr NJ. SLAM Endoscopy enhanced by adversarial depth prediction. arXiv:1907.00283 [cs, eess]. arXiv: 1907.00283. 2019.
143. Ali S, Zhou F, Bailey A, et al. A deep learning framework for quality assessment and restoration in video endoscopy. arXiv preprint arXiv:1904.07073. 2019.
144. Almalioglu Y, Ozyoruk KB, Gokce A, et al. EndoL2H: Deep super-resolution for capsule endoscopy. arXiv:2002.05459 [cs, eess]. arXiv: 2002.05459. 2020.
145. Durr NJ, González G, Parot V. 3D imaging techniques for improved colonoscopy. *Exp Rev Med Devices* 2014;11: 105–107.
146. Mahmood F, Chen R, Sudarsky S, Yu D, Durr NJ. Deep learning with cinematic rendering: Fine-tuning deep neural networks using photorealistic medical images. *Phys Med Biol* 2018;63:185012.
147. Chen R, Mahmood F, Yuille A, Durr NJ. Rethinking monocular depth estimation with adversarial training. arXiv:1808.07528 [cs]. arXiv: 1808.07528. 2019.
148. Mahmood F, Chen R, Durr NJ. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE Trans Med Imag* 2018;37:2572–2581.
149. Mahmood F, Durr NJ. Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. *Med Image Anal* 2018;48:230–243.
150. Lin J, Clancy NT, Qi J, et al. Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks. *Med Image Anal* 2018;48:162–176.
151. Ma R, Wang R, Pizer S, et al., editors. Real-Time 3D Reconstruction of Colonoscopic Surfaces for Determining Missing Regions in Medical Image Computing and

- Computer Assisted Intervention—MICCAI 2019. Cham: Springer International Publishing; 2019. pp 573–582.
152. Cuccia DJ, Bevilacqua FP, Durkin AJ, Ayers FR, Tromberg BJ. Quantitation and mapping of tissue optical properties using modulated imaging. *J Biomed Opt* 2009;14:024012.
  153. Zhao Y, Deng Y, Bao F, Peterson H, Istfan R, Roblyer D. Deep learning model for ultrafast multifrequency optical property extractions for spatial frequency domain imaging. *Opt Lett* 2018;43:5669–5672.
  154. Chen MT, Mahmood F, Sweer JA, Durr NJ. GANPOP: Generative adversarial network prediction of optical properties from single snapshot wide-field images. *IEEE Trans Med Imag* 2020;39(6):1988–1999.
  155. Chen MT, Durr NJ. Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning. *J Biomed Opt* 2020;25:112907.
  156. Grigoriou A, Yoon J, Bohndiek SE. Deep learning applied to hyperspectral endoscopy for online spectral classification. *Sci Rep* 2020;10(1):3947.
  157. Chen MT, Papadakis M, Durr NJ. Speckle illumination spatial frequency domain imaging for projector-free optical property mapping. arXiv:2006.03661 [physics]. arXiv:2006.03661. 2020.
  158. Halicek M, Dormer JD, Little JV, et al. Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning. *Cancers* 2019;11(9):1367.
  159. Golhar M, Bobrow TL, Khoshknab MP, Jit S, Ngamrueng-phong S, Durr NJ. Improving colonoscopy lesion classification using semi-supervised deep learning. *IEEE Access* 2021;9:631–640.
  160. Thienphrapa P, Bydlon T, Chen A, et al. Interactive endoscopy: A next-generation, streamlined user interface for lung surgery navigation. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*. Cham: Springer. 2019. pp 83–91.
  161. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science* 1991;254:1178–1181.
  162. Zysk AM, Nguyen FT, Oldenburg AL, Marks DL, Stephen Md, Boppart A. Optical coherence tomography: A review of clinical development from bench to bedside. *J Biomed Opt* 2007;12:051403.
  163. Olsen J, Holmes J, JemecGregor BE. Advances in optical coherence tomography in dermatology—A review. *J Biomed Opt* 2018;23:040901.
  164. Wang J, Xu Y, Boppart SA. Review of optical coherence tomography in oncology. *J Biomed Opt* 2017;22:121711.
  165. Fujimoto J, Huang D. Foreword: 25 years of optical coherence tomography. *Invest Ophthalmol Vis Sci* 2016;57:OCTi–OCTii.
  166. Tsai T-H, Leggett CL, Trindade AJ, et al. Optical coherence tomography in gastroenterology: A review and future outlook. *J Biomed Opt* 2017;22:121716.
  167. Prati F, Regar E, Mintz GS, et al. Expert review document on methodology, terminology, and clinical applications of optical coherence tomography: Physical principles, methodology of image acquisition, and clinical application for assessment of coronary arteries and atherosclerosis. *Eur Heart J* 2010;31:401–415.
  168. Petzold A, Boer JF, Schippling S, et al. Optical coherence tomography in multiple sclerosis: A systematic review and meta-analysis. *Lancet Neurol* 2010;9:921–932.
  169. McLean JP, Fang S, Gallos G, Myers KM, Hendon CP. Three-dimensional collagen fiber mapping and tractography of human uterine tissue using OCT. *Biomed Opt Express* 2020;11:5518–5541.
  170. Lurie KL, Smith GT, Khan SA, Liao JC, Bowden AK. Three-dimensional, distendable bladder phantom for optical coherence tomography and white light cystoscopy. *J Biomed Opt* 2014;19:036009.
  171. Venhuizen FG, Ginneken B, Liefers B, et al. Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. *Biomed Opt Express* 2017;8:3292–3316.
  172. Shah A, Zhou L, Abrámoff MD, Wu X. Multiple surface segmentation using convolution neural nets: Application to retinal layer segmentation in OCT images. *Biomed Opt Express* 2018;9:4509–4526.
  173. Lu D, Heisler M, Lee S, et al. Deep-learning based multi-class retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. *Med Image Anal* 2019;54:100–110.
  174. Devalla SK, Renukanand PK, Sreedhar BK, et al. DRUNET: A dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomed Opt Express* 2018;9:3244–3265.
  175. Chen J, Yang L, Zhang Y, Alber M, Chen DZ. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. *Adv Neural Inf Process Syst* 2016;29:3036–3044.
  176. Milletari F, Navab N, Ahmadi S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. 2016. pp 565–571.
  177. Li D, Wu J, He Y, et al. Parallel deep neural networks for endoscopic OCT image segmentation. *Biomed Opt Express* 2019;10:1126–1135.
  178. Li L, Jia T. Optical coherence tomography vulnerable plaque segmentation based on deep residual U-Net. *Rev Cardiovasc Med* 2019;20:171–177.
  179. Stefan S, Lee J, Lee J. Deep learning toolbox for automated enhancement, segmentation, and graphing of cortical optical coherence tomography microangiograms. *Biomed Opt Express* 2020;11:7325–7342.
  180. Dong Z, Liu G, Ni G, Jerwick J, Duan L, Zhou C. Optical coherence tomography image denoising using a generative adversarial network with speckle modulation. *J Biophotonics* 2020;13:e201960135.
  181. Rim TH, Lee AY, Ting DS, et al. Detection of features associated with neovascular age-related macular degeneration in ethnically distinct data sets by an optical coherence tomography: Trained deep learning algorithm. *Br J Ophthalmol* 2020.
  182. Qiu B, Huang Z, Liu X, et al. Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function. *Biomed Opt Express* 2020;11:817–830.
  183. Mao Z, Miki A, Mei S, et al. Deep learning based noise reduction method for automatic 3D segmentation of the anterior of lamina cribrosa in optical coherence tomography volumetric scans. *Biomed Opt Express* 2019;10:5832–5851.
  184. Fauw JD, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine* 2018;24(9):1342–1350.
  185. Fang L, Wang C, Li S, Rabbani H, Chen X, Liu Z. Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. *IEEE Trans Med Imag* 2019;38:1959–1970.
  186. Asaoka R, Murata H, Hirasawa K, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol* 2019;198:136–145.
  187. Loo J, Fang L, Cunefare D, Jaffe GJ, Farsiu S. Deep longitudinal transfer learning-based automatic segmentation of photoreceptor ellipsoid zone defects on optical coherence tomography images of macular telangiectasia type 2. *Biomed Opt Express* 2018;9:2681–2698.
  188. He Y, Carass A, Liu Y, Saidha S, Calabresi PA, Prince JL. Adversarial domain adaptation for multi-device retinal OCT segmentation. In: *Medical Imaging 2020: Image Processing*. International Society for Optics and Photonics: Vol 11313. 2020. p 1131309.
  189. Yang S, Zhou X, Wang J, et al. Unsupervised domain adaptation for cross-device OCT lesion detection via learning adaptive features. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 2020. pp 1570–1573.
  190. Braaf B, Donner S, Uribe-Patarroyo N, Bouma BE, Vakoc BJ. A neural network approach to quantify blood flow from retinal OCT intensity time-series measurements. *Sci Rep* 2020;10(1):9611.

191. Lee CS, Tying AJ, Wu Y, et al. Generating retinal flow maps from structural optical coherence tomography with artificial intelligence. *Sci Rep* 2019;9(1):5694.
192. Liu Xi, Huang Z, Wang Z, et al. A deep learning based pipeline for optical coherence tomography angiography. *J Biophotonics* 2019;12:e201900008.
193. Liu R, Cheng S, Tian L, Yi J. Deep spectral learning for label-free optical imaging oximetry with uncertainty quantification. *Light Sci Appl* 2019;8(1):102.
194. Xu M, Wang LV. Photoacoustic imaging in biomedicine. *Rev Sci Instrum* 2006;77(4):041101.
195. Beard P. Biomedical photoacoustic imaging. *Interface Focus* 2011;1:602–631.
196. Bell MAL. Deep learning the sound of light to guide surgeries. In: *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XVII*. International Society for Optics and Photonics: Vol 10868. 2019. p. 108680G.
197. Lediju BMA. Photoacoustic imaging for surgical guidance: Principles, applications, and outlook. *J Appl Phys* 2020;128(6):060904.
198. Hauptmann A, Cox BT. Deep learning in photoacoustic tomography: Current approaches and future directions. *J Biomed Opt* 2020;25:112903.
199. Reiter A, Bell MAL. A machine learning approach to identifying point source locations in photoacoustic data. In: *Photons Plus Ultrasound: Imaging and Sensing 2017*. International Society for Optics and Photonics: Vol 10064. 2017. p 100643J.
200. Allman D, Reiter A, Bell MAL. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE Trans Med Imag* 2018;37:1464–1477.
201. Johnstonbaugh K, Agrawal S, Durairaj DA, et al. A deep learning approach to photoacoustic wavefront localization in deep-tissue medium. *IEEE Trans Ultrason Ferroelectr Freq Control* 2020;67(12):2649–2659.
202. Kim MW, Jeng G-S, Pelivanov I, O'Donnell M. Deep-learning image reconstruction for real-time photoacoustic system. *IEEE Trans Med Imag* 2020;39(11):3379–3390.
203. Wu D, Kim K, Li Q. Computationally efficient deep neural network for computed tomography image reconstruction. *Med Phys* 2019;46:4763–4776.
204. Hauptmann A, Lucka F, Betcke M, et al. Model-based learning for accelerated, limited-view 3-d photoacoustic tomography. *IEEE Trans Med Imag* 2018;37:1382–1393.
205. Tong T, Huang W, Wang K, et al. Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data. *Photoacoustics* 2020;19(5):100190.
206. Waibel D, Gröhl J, Isensee F, Kirchner T, Maier-Hein K, Maier-Hein L. Reconstruction of initial pressure from limited view photoacoustic images using deep learning. In *Photons Plus Ultrasound: Imaging and Sensing 2018*. International Society for Optics and Photonics: Vol 10494. 2018. p 104942S.
207. Antholzer S, Haltmeier M, Schwab J. Deep learning for photoacoustic tomography from sparse data. *Inverse Probl Sci Eng* 2019;27:987–1005.
208. Guan S, Khan AA, Sikdar S, Chitnis PV. Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning. *Sci Rep* 2020;10:1–12.
209. Davoudi N, Deán-Ben X, Razansky D. Deep learning photoacoustic tomography with sparse data. *Nat Mach Intell* 2019;1:453–460.
210. Hariri A, Alipour K, Mantri Y, Schulze JP, Jokerst JV. Deep learning improves contrast in low-fluence photoacoustic imaging. *Biomed Opt Express* 2020;11:3360–3373.
211. Awasthi N, Jain G, Kalva SK, Pramanik M, Yalavarthy PK. Deep neural network based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography. *IEEE TTrans Ultrason Ferroelectr Freq Control* 2020;67(12):2660–2673.
212. Gamelin J, Maurudis A, Aguirre A, et al. A real-time photoacoustic tomography system for small animals. *Opt Express* 2009;17:10489–10498.
213. Lin L, Hu P, Shi J, et al. Single-breath-hold photoacoustic computed tomography of the breast. *Nat Commun* 2018;9:1–9.
214. Xi L, Jiang H. High resolution three-dimensional photoacoustic imaging of human finger joints in vivo. *Appl Phys Lett* 2015;107:063701.
215. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature* 2018;555:487–492.
216. Xu Y, Wang LV, Ambartsoumian G, Kuchment P. Reconstructions in limited-view thermoacoustic tomography. *Med Phys* 2004;31:724–733.
217. Vu T, Li M, Humayun H, Zhou Y, Yao J. A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer. *Exp Biol Med* 2020;245:597–605.
218. Finch D, Haltmeier M. Rakesh. Inversion of spherical means and the wave equation in even dimensions. *SIAM J Appl Math* 2007;68:392–412.
219. Gutta S, Kadimesetty VS, Kalva SK, Pramanik M, Ganapathy S, Yalavarthy PK. Deep neural network-based bandwidth enhancement of photoacoustic data. *J Biomed Opt* 2017;22:116001.
220. Moustakidis S, Omar M, Aguirre J, Mohajerani P, Ntziachristos V. Fully automated identification of skin morphology in raster-scan optoacoustic mesoscopy using artificial intelligence. *Med Phys* 2019;46:4046–4056.
221. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. pp 770–778.
222. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. 2012. pp 1097–1105.
223. Boink YE, Manohar S, Brune C. A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation. *IEEE Trans Med Imag* 2019;39:129–139.
224. Adler J, Öktem O. Learned primal-dual reconstruction. *IEEE Trans Med Imag* 2018;37:1322–1332.
225. Cai C, Deng K, Ma C, Luo J. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Opt Lett* 2018;43:2752–2755.
226. Chen X, Qi W, Xi L. Deep-learning-based motion-correction algorithm in optical resolution photoacoustic microscopy. *Vis Comput Ind Biomed Art* 2019;2:12.
227. Manwar R, Li X, Mahmoodkalayeh S, Asano E, Zhu D, Avannai K. Deep learning protocol for improved photoacoustic brain imaging. *J Biophotonics* 2020;13:e202000212.
228. Allman D, Assis F, Chrispin J, Bell MAL. Deep neural networks to remove photoacoustic reflection artifacts in ex vivo and in vivo tissue. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2018. pp 1–4.
229. Allman D, Assis F, Chrispin J, Bell MAL. A deep learning-based approach to identify in vivo catheter tips during photoacoustic-guided cardiac interventions. In: *Photons Plus Ultrasound: Imaging and Sensing 2019*. International Society for Optics and Photonics: Vol 10878. 2019. p 108785E.
230. Allman D, Assis F, Chrispin J, Bell MAL. Deep learning to detect catheter tips in vivo during photoacoustic-guided catheter interventions: Invited presentation. In: *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*. IEEE. 2019. pp 1–3.
231. Wang Y, Chang J-H, Aronson R, Barbour RL, Graber HL, Lubowsky J. Imaging of scattering media by diffusion tomography: an iterative perturbation approach. In: *Mang TS, editor. Proc. SPIE 1641, Physiological Monitoring and Early Detection Diagnostic Methods*. 1992. pp 58–71.
232. Boas DA, Brooks DH, Miller EL, et al. Imaging the body with diffuse optical tomography. *IEEE Signal Process Mag* 2001;18:57–75.
233. Gibson AP, Hebden JC, Arridge SR. Recent advances in diffuse optical imaging. *Phys Med Biol* 2005;50:R1–R43.
234. Intes X. Time-domain optical mammography softScan. *Acad Radiol* 2005;12:934–947.
235. Ferrari M. Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *J Biomed Opt* 2007;12:062104.

236. Eggebrecht AT, Ferradal SL, Robichaux-Viehoever A, et al. Mapping distributed brain function and networks with diffuse optical tomography. *Nat Photonics* 2014;8:448–454.
237. Corlu A, Choe R, Durduran T, et al. Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans. *Opt Express* 2007;15:6696.
238. Darne C, Lu Y, Sevick-Muraca EM. Small animal fluorescence and bioluminescence tomography: A review of approaches, algorithms and technology update. *Phys Med Biol* 2014;59:R1–R64.
239. Rice WL, Shcherbakova DM, Verkhusha VV, Kumar ATN. In vivo tomographic imaging of deep-seated cancer using fluorescence lifetime contrast. *Cancer Res* 2015;75:1236–1243.
240. Kak AC, Slaney M, Wang Ge. Principles of computerized tomographic imaging. *Med Phys* 2002;29:107.
241. O'Leary MA, Boas DA, Chance B, Yodh AG. Experimental images of heterogeneous turbid media by frequency-domain diffusing-photon tomography. *Opt Lett* 1995;20:426.
242. Yao R, Pian Q, Intes X. Wide-field fluorescence molecular tomography with compressive sensing based preconditioning. *Biomed Opt Express* 2015;6:4887.
243. Zhang X, Badea CT, Johnson GA. Three-dimensional reconstruction in free-space whole-body fluorescence tomography of mice using optically reconstructed surface and atlas anatomy. *J Biomed Optics* 2009;14:064010.
244. Guven M, Yazici B, Intes X, Chance B. Diffuse optical tomography with a priori anatomical information. *Phys Med Biol* 2005;50:2837–2858.
245. Li A, Miller EL, Kilmer ME, et al. Tomographic optical breast imaging guided by three-dimensional mammography. *Appl Opt* 2003;42:5181.
246. Schulz RB, Ale A, Sarantopoulos A, et al. Hybrid system for simultaneous fluorescence and X-ray computed tomography. *IEEE Trans Med Imag* 2010;29:465–473.
247. Venugopal V, Chen J, Barroso M, Intes X. Quantitative tomographic imaging of intermolecular FRET in small animals. *Biomed Opt Express* 2012;3(12):3161.
248. Long F. Deep learning-based mesoscopic fluorescence molecular tomography: An in silico study. *J Med Imag* 2018;5:1.
249. Gao Y, Wang K, An Y, Jiang S, Meng H, Tian J. Nonmodel-based bioluminescence tomography using a machine-learning reconstruction strategy. *Optica* 2018;5:1451.
250. Guo L, Liu F, Cai C, Liu J, Zhang G. 3D deep encoder-decoder network for fluorescence molecular tomography. *Opt Lett* 2019;44:1892.
251. Huang C, Meng H, Gao Y, Jiang S, Wang K, Tian J. Fast and robust reconstruction method for fluorescence molecular tomography based on deep neural network. In: Farkas DL, Leary JF, Tarnok A, editors. *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XVII*. SPIE. 2019. p 55.
252. Yoo J, Sabir S, Heo D, et al. Deep learning diffuse optical tomography. *IEEE Trans Med Imag* 2020;39:877–887.
253. Ye JC, Han Y, Cha E. Deep convolutional framelets: A general deep learning framework for inverse problems. *SIAM J Imag Sci* 2018;11:991–1048.
254. Fang Q. Mesh-based Monte Carlo method using fast ray-tracing in Plücker coordinates. *Biomed Opt Express* 2010;1:165.
255. Yao R, Intes X, Fang Q. Generalized mesh-based Monte Carlo for wide-field illumination and detection via mesh retessellation. *Biomed Opt Express* 2016;7:171.
256. Hiroyasu T, Hanawa K, Yamamoto U. Gender classification of subjects from cerebral blood flow changes using Deep Learning. In 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM). IEEE. 2014. pp 229–233.
257. Mirbagheri M, Jodeiri A, Hakimi N, Zakeri V, Setarehdan SK. Accurate stress assessment based on functional near infrared spectroscopy using deep learning approach. In: 2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME). IEEE. 2019. pp 4–10.
258. Dedovic K, D'Aguiar C, Pruessner JC. What stress does to your brain: A review of neuroimaging studies. *Can J Psychiatry* 2009;54:6–15.
259. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci Biobehav Rev* 2017;74:58–75.
260. Rosas-Romero R, Guevara E, Peng K, et al. Prediction of epileptic seizures with convolutional neural networks and functional near-infrared spectroscopy signals. *Comput Biol Med* 2019;111:103355.
261. Bandara D, Hirshfield L, Velipasalar S. Classification of affect using deep learning on brain blood flow data. *J Near Infrared Spec* 2019;27:206–219.
262. Benerradi Johann A, Maior H, Marinescu A, Clos J, Max L W. Exploring machine learning approaches for classifying mental workload using fNIRS data from HCI tasks. In: *Proceedings of the Halfway to the Future Symposium 2019*. New York, NY: ACM. 2019. pp 1–11.
263. Gao Y, Kruger U, Intes X, Schwaitzberg S, De S. A machine learning approach to predict surgical learning curves. *Surgery*. 2020;167:321–327.
264. Cecotti H, Graesser A. Convolutional neural network with embedded Fourier transform for EEG classification. In: 2008 19th International Conference on Pattern Recognition. IEEE. 2008. pp 1–4.
265. Tayeb Z, Fedjaev J, Ghaboosi N, et al. Validating deep neural networks for online decoding of motor imagery movements from EEG signals. *Sensors* 2019;19:210.
266. Henrich J, Herff C, Heger D, Schultz T. Investigating deep learning for fNIRS based BCI. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. 2015. pp 2844–2847.
267. Dargazany AR, Abtahi M, Mankodiya K. An end-to-end (deep) neural network applied to raw EEG, fNIRS and body motion data for data fusion and BCI classification task without any pre-/post-processing. *arXiv preprint arXiv:1907.09523*. 2019.
268. Trakoolwilaiwan T, Behboodi B, Lee J, Kim K, Choi J-W. Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain-computer interface: three-class classification of rest, right-, and left-hand motor execution. *Neurophotonics* 2017;5(1):011008.
269. Saadati M, Nelson J, Ayaz H. Multimodal fNIRS-EEG classification using deep learning algorithms for brain-computer interfaces purposes. In: *International Conference on Applied Human Factors and Ergonomics*. Cham: Springer; 2019.
270. Gao Y, Cavuoto L, Yan P, et al. A deep learning approach to remove motion artifacts in fNIRS data analysis. In: *Biophotonics Congress: Biomedical Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN)*. Washington, DC: BM2C.7OSA. 2020.
271. Poon C-S, Long F, Sunar U. Deep learning model for ultrafast quantification of blood flow in diffuse correlation spectroscopy. *bioRxiv*. 2020.
272. Suk H-I, Wee C-Y, Lee S-W, Shen D. State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *NeuroImage* 2016;129:292–307.
273. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van EDC, Raichle ME. From the cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A* 2005;102:9673–9678.
274. Zhou J, Greicius MD, Gennatas ED, et al. Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain*. 2010;133:1352–1367.
275. Kawahara J, Brown CJ, Miller SP, et al. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*. 2017;146:1038–1049.
276. Behboodi B, Lim S-H, Luna M, Jeon H-A, Ji-Woong C. Artificial and convolutional neural networks for assessing functional connectivity in resting-state functional near infrared spectroscopy. *J Near Infrared Spectrosc* 2019;27:191–205.
277. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv* 2010;4:40–79.
278. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.

279. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. arXiv:1908.09635 [cs]. arXiv: 1908.09635. 2019.
280. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv.* 2018; 51: 1–42.
281. Samek W, Montavon G, Lapuschkin S, Anders CJ, Müller K-R. Toward interpretable machine learning: Transparent deep neural networks and beyond. arXiv:2003.07631 [cs, stat]. arXiv: 2003.07631. 2020.
282. Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. arXiv:2005.13799 [cs, eess]. arXiv: 2005.13799. 2020.
283. Biffi C, Cerrolaza JJ, Tarroni G, et al. Explainable anatomical shape analysis through deep hierarchical generative models. arXiv:1907.00058 [cs, eess]. arXiv: 1907.00058. 2020.
284. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 2019;1:206–215.
285. Hägele M, Seegerer P, Lapuschkin S, et al. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. *Sci Rep* 2020;10(1):6423.
286. Li X, Zhou Y, Dvornek NC, Gu Y, Ventola P, DuncanJames S. Efficient Shapley explanation for features importance estimation under uncertainty. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., editors. International Conference on Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science. Cham: Springer International Publishing. 2020. pp 792–801.
287. Reinhold JC, He Y, Han S, et al. Finding novelty with uncertainty. arXiv:2002.04626 [cs, eess]. arXiv: 2002.04626. 2002.