Recent advances in deep-learning-enhanced photoacoustic imaging

Jinge Yang,[a,†] **Seongwook Choi** [a,†] **Jiwoong Kim,**[a] **Byullee Park,**[b,*] **and Chulhong Kim** [a,*]

[a]Pohang University of Science and Technology, School of Interdisciplinary Bioscience and Bioengineering, Graduate School of Artificial Intelligence, Medical Device Innovation Center, Department of Electrical Engineering, Convergence IT Engineering, and Mechanical Engineering, Pohang, Republic of Korea

[b]Sungkyunkwan University, Institute of Quantum Biophysics, Department of Biophysics, Suwon, Republic of Korea

Abstract. Photoacoustic imaging (PAI), recognized as a promising biomedical imaging modality for preclinical and clinical studies, uniquely combines the advantages of optical and ultrasound imaging. Despite PAI's great potential to provide valuable biological information, its wide application has been hindered by technical limitations, such as hardware restrictions or lack of the biometric information required for image reconstruction. We first analyze the limitations of PAI and categorize them by seven key challenges: limited detection, low-dosage light delivery, inaccurate quantification, limited numerical reconstruction, tissue heterogeneity, imperfect image segmentation/classification, and others. Then, because deep learning (DL) has increasingly demonstrated its ability to overcome the physical limitations of imaging modalities, we review DL studies from the past five years that address each of the seven challenges in PAI. Finally, we discuss the promise of future research directions in DL-enhanced PAI.

Keywords: photoacoustic imaging; deep learning; biomedical imaging.

1 Introduction

Photoacoustic imaging (PAI) is a noninvasive and radiation-free biomedical imaging modality that provides high spatial resolution, deep penetration, and great optical absorption contrast by synergistically combining optics and acoustics.[1] PAI is based on the photoacoustic (PA) effect, in which optical energy from a pulse laser is converted into acoustic energy waves by the light absorption characteristics of biomolecules.[2] The initial pressure of a generated PA wave can be calculated as

$$p_0 = \Gamma \mu_a \eta_{th} F, \qquad (1)$$

where p_0 denotes the initial pressure, Γ is the Gruneisen coefficient, μ_a is the optical absorption coefficient, η_{th} is the efficiency of heat conversion from the optical absorption, and F is the optical fluence. Because acoustic scattering in biological tissues is several orders of magnitude lower than light scattering, PAI can obtain biomolecule information based on the absorption contrast of light at a depth of several centimeters.[3,4] PAI also extracts the concentrations of intrinsic chromophores, such as oxyhemoglobin (HbO), deoxyhemoglobin (HbR), melanin, water, and lipids, using multispectral image processing.[5–13] In particular, oxygen saturation (sO_2), an important index for evaluating various diseases, is calculated through HbO and HbR values.[14] By exploiting spectral characteristics, PAI can analyze physiological functions such as sO_2, blood flow, and metabolic rates in preclinical and clinical research.[4,15–17] For example, the high sensitivity of PAI to hemoglobin has made it valuable in preclinical studies of angiogenic diseases, tumor hypoxia, and cerebral hemodynamics.[18–20] Further, the use of PAI is expanding into clinical research areas, such as thyroid and breast cancer screening, lymph node biopsy guidance, tissue examination, and melanoma staging.[21–24] Not limited to endogenous chromophores, the high molecular sensitivity of PAI enables molecular imaging when exogenous contrast agents are administered.[4] Biodistribution and pharmacokinetics in the

*Address all correspondence to Byullee Park, byullee@skku.edu; Chulhong Kim, chulhong@postech.edu

†These authors contributed equally to this work.

body can be imaged *in vivo* through exogenous contrast agents that generate PA signals.[25–27] In this way, PAI is being used in diagnosing cancer and brain diseases and monitoring their therapies, in studying the organ accumulation of substances, and in tracking the dissemination of drugs.[8,28–30]

Photoacoustic microscopy (PAM) and photoacoustic computed tomography (PACT) are the main PAI modalities. PAM is subdivided into two types, depending on which of the two co-aligned acoustic and optical components is more tightly focused.[31–36] Optical resolution PAM (OR-PAM), which implements focused optical illumination on the acoustic focal area, shows high spatial resolution (a few micrometers) and has been applied to investigate small biological structures.[37,38] Acoustic resolution PAM (AR-PAM) uses a less tightly focused optical beam than OR-PAM, but its acoustic focus is smaller than its laser focus.[39] AR-PAM achieves deeper light penetration (up to several centimeters, compared to the 1 mm depth in OR-PAM) despite its lower spatial resolution (tens/hundreds of micrometers), defined by the acoustic focus. PACT uses multiple detection positions to simultaneously reconstruct an image in 2D or 3D. It provides hundreds of images with micrometer-level spatial resolution at imaging depths ranging in the tens of millimeters. PACT uses a high-energy wide laser beam and an ultrasound (US) transducer array (e.g., linear, ring-shaped, arc-shaped, or hemispherical) to receive US waves generated by laser illumination.[3,21,40–42] Images are created by reconstruction algorithms,[43] such as delay-and-sum (DAS),[44,45] delay-multiply-and-sum (DMAS),[46] backprojection (BP),[47] Fourier beam forming,[48] time reversal (TR),[49] and model-based methods.[50,51]

PAI has gained widespread recognition as a promising biomedical imaging modality for preclinical and clinical studies. However, to fully realize PAI's great potential, the seven challenges listed in Table 1 and illustrated in Fig. 1 must be addressed to further enhance the image quality and expand PAI's applications:

(1) The first challenge, overcoming limited detection capability, arises because most PAI systems are still constrained by such factors as restricted bandwidth, a limited detection view, and sampling sparsity.[59]

(2) The second challenge is to compensate for low-dosage light delivery. The PACT systems based on LEDs or laser diodes are portable and cost-effective alternatives to bulky and expensive solid-state laser systems. However, their low-dosage light delivery provides only a low signal-to-noise ratio (SNR), which can affect image quality. In the case of OR-PAM, fast scanning with high repetition rates is necessary for certain applications such as recording brain-wide neuronal activities.[60] However, to ensure laser safety, low laser dosages are required, resulting in reduced SNRs and decreased image qualities.

(3) The third challenge is to improve the accuracy of quantitative PA imaging. Accurately determining physiological parameters remains a demanding task due to the complex and nonlinear nature of light absorption and scattering.[61]

(4) The fourth challenge is to optimize or replace current reconstruction methods, whose inherent limitations compromise their accuracy and effectiveness in generating high-quality images.

(5) The fifth challenge is to address the problems posed by tissue heterogeneity. Local variations in the acoustical properties of biological tissue can lead to inconsistencies in the reconstructed PA images, resulting in artifacts that degrade the accuracy of quantitative measurements derived from the images.[62]

(6) The sixth challenge is to improve the classification and segmentation accuracy of PA images. The limited availability of annotated PAI data sets has hindered the development of automated image classification and segmentation, resulting in either continued reliance on manual delineation by expert physicians or the adaptation of traditional methods from other imaging modalities.

(7) In addition to the six challenges mentioned earlier, there are still specific issues, such as motion artifacts, limited spatial resolution, electrical noise, image misalignment, accelerating superresolution imaging, and achieving digital histologic staining, which are also important for PAI studies. To ensure a comprehensive understanding of the challenges in PAI, these specific issues are categorized as a seventh challenge. These seven challenges are summarized in Table 1.

Overcoming these challenges is important because relying solely on hardware improvements will not be enough to resolve them. It will require significant investments of time and resources to find effective solutions. Deep learning (DL) plays a crucial role in advancing the field of medical and bioimaging by

Table 1 Summary of challenges facing PAI.

Section	Title	Challenges to be solved
3.1	Overcoming limited detection capabilities	Restricted bandwidth, limited detection view, sampling sparsity
3.2	Compensating for low-dosage light delivery	Low SNR in the low-dosage light-delivery system
3.3	Improving the accuracy of quantitative PA imaging	Inaccuracy in quantitative estimates (sO_2, optical absorption coefficient)
3.4	Optimizing or replacing conventional reconstruction algorithms	Limitations in conventional reconstruction algorithms
3.5	Addressing tissue heterogeneity	Acoustic reflection and imaging artifacts led by tissue heterogeneity
3.6	Improving the accuracy of image classification and segmentation	Inaccuracy and rough classification and segmentation of PA image
3.7	Overcoming other specified issues	Motion artifacts, limited spatial resolution, electrical noise and interference, image misalignment, accelerating superresolution imaging, achieving digital histologic staining
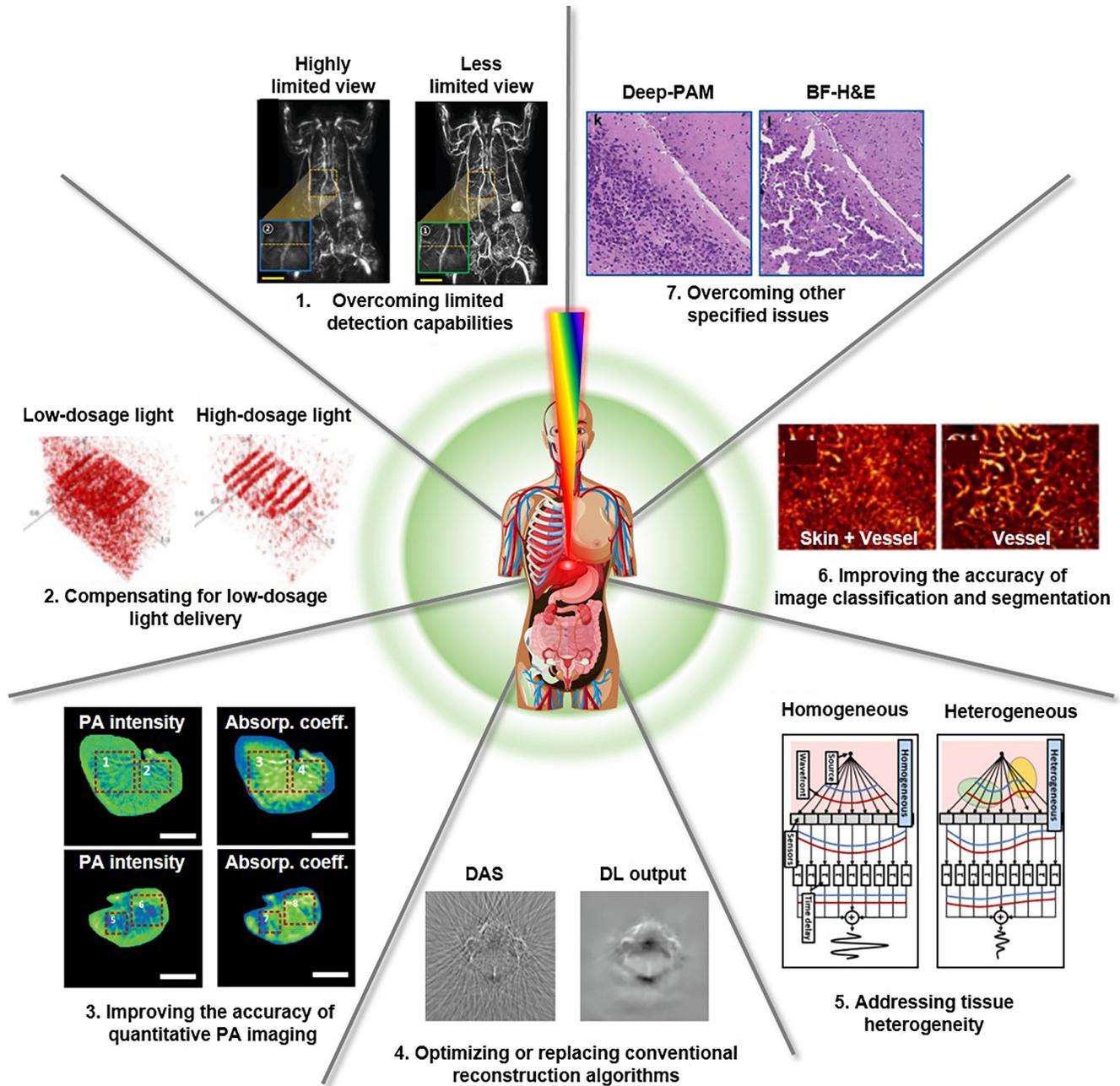
Fig. 1 Representations of seven major challenges in PAI, and DL-related methods to overcome them. DAS, delay-and-sum; DL, deep learning; BF-H&E, bright-field hematoxylin and eosin staining. The images are adapted with permission from Ref. 52, © 2021 Wiley-VCH GmbH; Ref. 53, © 2020 Optica; Ref. 54, © 2022 Optica; Ref. 55, © 2020 Elsevier GmbH; Ref. 56, CC-BY; Ref. 57, © 2021 Elsevier GmbH; and Ref. 58, © 2021 Elsevier GmbH.

not only addressing the inherent limitations of imaging systems but also by driving substantial improvements in classification and segmentation performance. In recent years, DL has gained significant traction in PAI research, leading to remarkable breakthroughs and achievements. This comprehensive review article provides an in-depth analysis of diverse methodologies and outcomes showcasing the utilization of DL techniques to effectively address the seven challenges encountered in PAI, as previously outlined.

2 Principles of DL Methods

DL is a subset of machine-learning algorithms that encompasses supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is a modeling technique that establishes a correlation between input data and their corresponding ground truth (GT). This approach is commonly utilized in DL-enhanced medical imaging, where high- and low-quality images can be paired. On the other hand, unsupervised learning
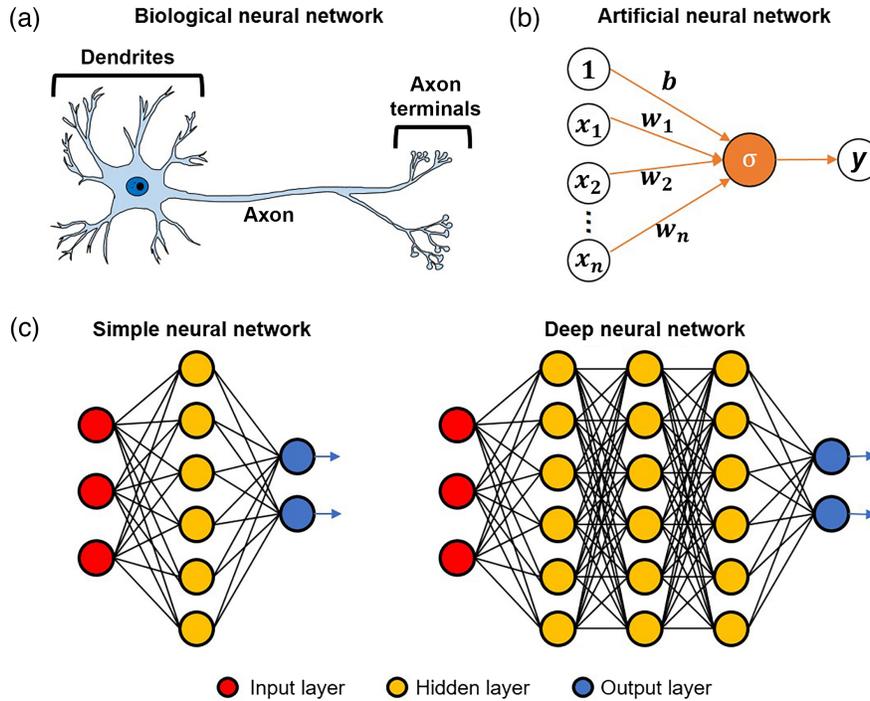
Fig. 2 The concept of (a) a biological neural network and (b) an ANN derived from (a). (c) Schematics of a simple neural network and a DNN.

identifies specific patterns hidden within data, without the use of labeled examples or *a priori* known answers. Lastly, in reinforcement learning, an algorithm maximizes the final reward by learning through rewards obtained as a result of performing specific actions in a particular environment. A notable example of such a learning algorithm is AlphaGo, the first computer program to beat a human champion Go player.[63] Next, we explain the basic structure of DL network and the basic operating principle of DL training. In addition, we introduce the representative DL architectures that are most widely used in the field of image and video: convolution neural network (CNN), U-shaped neural network (U-Net), and generative adversarial network (GAN) architecture.

2.1 Artificial Neural Networks

Artificial neural networks (ANNs) draw inspiration from biological neural networks, wherein different stimuli enter neurons through dendrites and are transmitted to other cells through axons once a threshold of activation is achieved [Fig. 2(a)]. In ANNs, an artificial neuron is a mathematical function conceived as a biological neuron. This function multiplies the various inputs by each weight, sums them, and adds the deviation.

This sum is then sent to a specific activation function and produces an output [Fig. 2(b)]. This artificial neuron is expressed as

$$y = \sigma \cdot \left([x_1 \ x_2 \ \dots \ x_n] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} + b \right) , \quad (2)$$

where σ is the activation function, x is the input, w is the weight, b is the variance, and y is the output. The calculated output acquires nonlinearity through such activation functions (σ) as sigmoid, hyperbolic tangent, and rectified linear unit.[64] Without such an activation function, it would be pointless to build a deep model, since a linear transformation would occur regardless of how many hidden layers are present.

ANNs typically have an input layer, a hidden layer, and an output layer, each comprising multiple units. The number of hidden layers between the input and output determines whether the ANN is a simple neural network or a deep neural network (DNN) [Fig. 2(c)]. Formulaically, the two networks in Fig. 2(c) are described as

$$y = w_2(\sigma(w_1 \times x + b_1)) + b_2 , \quad (3)$$

and

$$y = w_4(\sigma(w_3(\sigma(w_2(\sigma(w_1 \times x + b_1)) + b_2)) + b_3)) + b_4. \quad (4)$$

2.2 Backpropagation

Backpropagation serves as the fundamental principle for training DL models. In order to grasp the concept of backpropagation, it is necessary to first understand forward propagation. Forward propagation involves sequentially passing an input value through multiple hidden layers to generate an output. For instance, given an input x, a weight w, a variance b, and an activation function σ, the DNNs' forward propagation on Fig. 2(c) is represented as

$$y^* = w_4(\sigma(w_3(\sigma(w_2(\sigma(w_1 \times x + b_1)) + b_2)) + b_3)) + b_4, \quad (5)$$

where y^* represents the value predicted by DNNs. The process of finding the optimized w and b variables that minimize the loss, which is the difference between the predicted result and the actual y, is called training. The loss is calculated for all training data sets. As loss-function metrics, image data sets such as PA images commonly employ the structural similarity index measure (SSIM) and peak-signal-to-noise ratio (PSNR). Backpropagation refers to the process of transmitting the loss back to the input stage using the chain rule.[65,66] This process determines the weight w that yields the minimal loss function value. Most DL models adopt the gradient descent technique throughout this procedure.[67] By calculating the gradient at a specific w and continually updating w, the loss function's minima can be estimated using the following equation:

$$w_{t+1} = w_t - \text{gradient} \times \text{learning rate}. \qquad (6)$$

The learning rate, a hyperparameter that determines the variable's update amount in proportion to the calculated slope, is set before the learning process and remains unchanged. The number of hidden layers and their dimensions are also hyperparameters. In the following section, we introduce representative ANN architectures commonly used in biomedical imaging, including PAI.

2.3 CNN

CNNs, the most basic DL architecture, have received significant attention in the field of PAI due to their extensive use in image processing and computer vision. CNNs were developed to extract features or patterns in local areas of an image. A convolution operation is a mathematical process that measures the similarity between two functions. The convolution operation in image processing is the process of calculating how well a subsection of an image matches a filter (also referred to as a kernel) and is used for things such as edge filtering. In CNN, the network is trained by learning this filter, which is used to extract image features, as a weight. The encoder–decoder CNN architecture is a relatively simple network structure capable of performing image-to-image translation tasks [Fig. 3(a)].[68] Initially, the input image undergoes a series of downsampling and convolution operations. Throughout this process, the dimensions of the image progressively decrease while the number of image channels, representing an additional dimension, increases.

This results in a bottleneck in the representation, which is subsequently reversed through a sequence of upsampling and convolutional operations. The bottleneck enforces the network to encode the image into a compact set of abstracted variables (also referred to as latent variables) along the channel dimension. The predicted image is synthesized by decoding these variables in the second half of the network.

2.4 U-Net

A U-Net [Fig. 3(b)] is a CNN-based model that was originally proposed for image segmentation in the biomedical field.[69] The U-Net is composed of two symmetric networks: a network for obtaining overall context information of an image and a second network for accurate localization. The left part of the U-Net is the encoding process, which encodes the input image to obtain overall context information. The right part of U-Net is the decoding process, which decodes the encoded context information to generate a segmented image. The feature maps obtained during the encoding process are concatenated with up-convolved feature maps at each expanding step in the decoding process, using skip connections.[70] This enables the decoder to make more accurate predictions by directly conveying important information in the image. As a result, the U-Net architecture has shown excellent performance in several biomedical image segmentation tasks, even when trained on a very small amount of data, due to data augmentation techniques.

2.5 GAN

A GAN is a type of generative model that learns data through a competition between a generator network and a discriminator network [Fig. 3(c)].[71,72] The generator network generates the fake data, and the discriminator network tries to distinguish the real data from the fake data. To deceive the discriminator, the generator aims to generate data that look as realistic as possible, while the discriminator attempts to distinguish the real data from the realistic fake data. Through this competition, both networks learn and improve iteratively, resulting in a generator that can generate increasingly realistic data. As a result, GANs have been successful in generating synthetic data that are very similar to real data, making them useful for applications such as data augmentation and image synthesis. These three representative networks are summarized in Table 2.
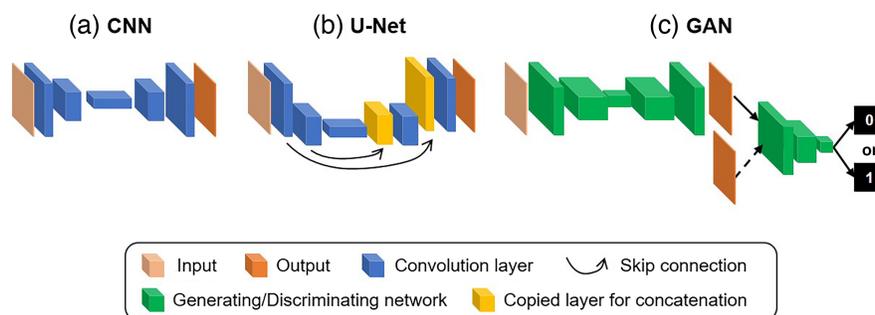
Fig. 3 Three typical neural network architectures for biomedical imaging. (a) CNN, (b) U-Net, and (c) GAN.

Table 2 Three representative networks.

Network	Key feature	Use case
CNN	Performs convolution operation for feature extraction.	Image enhancement
	Exhibits outstanding performance in feature extraction.	Image classification and object detection
	Captures spatial information of input data efficiently.	Image segmentation
U-Net	Comprises an encoder–decoder structure.	Image enhancement
	Utilizes skip connections to leverage high-resolution feature maps.	
	Demonstrates strong performance even with small data sets.	Image segmentation
	Excels in segmentation tasks.	
GAN	Consists of a generator network and a discriminator network.	Image generation
	Generates data that closely resembles real input data (generator).	
	Discriminates between generated data and real data (discriminator).	Image style transfer
	Engages in competitive training between the generator and discriminator.	Image/data augmentation
	Applies for generating new data.	

3 Challenges in PAI and Solutions through DL

3.1 Overcoming Limited Detection Capabilities

In PAI, optimal image quality requires a broadband US transducer and dense spatial sampling to enclose the target.[43,61,73] However, real-world scenarios introduce limitations, such as limited bandwidth, limited view, and data sparsity. DL methods have been used as postprocessing techniques to overcome these limitations and enhance the PA signals or images, reducing artifacts. This section provides an overview of studies utilizing DL methods as PAI postprocessing methods. The DL-based image reconstruction studies are discussed separately in Sec. 3.4.

3.1.1 Limited bandwidth

The bandwidth of US transducer arrays is limited compared to the natural broadband PA signal (from tens of kilohertz to a hundred megahertz).[74] Although optical detectors of PA waves have expanded the detection bandwidth, manufacturing high-density optical detector arrays and adopting them to PACT remains a technical challenge.[75]

To solve the limited-bandwidth problem, Gutte et al. proposed a DNN with five fully connected layers to enhance the PA bandwidth [Fig. 4(a)].[76] The network takes a limited-bandwidth signal as input and outputs an enhanced bandwidth signal, which is then used for PA image reconstruction using DAS. To train the network, the authors generated numerical phantoms using the k-Wave toolbox[79] to create pairs of full-bandwidth and limited-bandwidth PA signals. The synthesized results from the numerical phantoms demonstrated an enhanced bandwidth that is like that of images obtained from the full-bandwidth signal [Fig. 4(a)].

3.1.2 Limited view

PA image quality is reduced by the scant information provided by the limited coverage angle of the PA signals detected by the US transducer.[3] This problem is commonly encountered in PACT systems, particularly in linear US array-based systems and is referred to as the "limited view" problem.[75,80] Researchers

have addressed this problem to some extent by developing reconstruction methods with iterative methods.[80] Recent studies based on the fluctuation of the PA signal of blood flow or microbubbles[81,82] show another effective solution, but they need a number of single images to reconstruct one fluctuation image, which compromises the temporal resolution.

Deng et al.[83] developed DL methods using U-Net and principal component analysis processed very deep convolutional networks (PCA-VGG)[84] while Zhang et al.[85] designed a dual domain U-Net (DuDoUnet) incorporating reconstructed images and frequency domain information. In addition to utilizing the U-Net architecture, researchers have also explored the use of GAN networks, which have garnered attention due to their ability to preserve high-frequency features and prevent over-smoothing in images. Lu et al. proposed a GAN-based network called the limited view GAN (LV-GAN).[77] Figure 4(b) shows the architecture of LV-GAN, which consists of two networks: the generator network responsible for generating high-quality PA images from limited-view images, and the discriminator network designed to distinguish the generated images from the GT. To ensure accurate and generalizable results, the LV-GAN was trained using both simulated data generated by the k-Wave toolbox and experimental data obtained from a custom-made PA system. The results presented in Fig. 4(b), using *ex vivo* data, demonstrate the ability of LV-GAN to successfully reconstruct high-quality PA images in limited-view scenarios. The quantitative analysis further confirms that LV-GAN outperforms the U-Net framework, achieving the highest retrieval accuracy.

The combination of a postprocessing method with direct processing using PA signals is considered as another approach to reduce artifacts in limited-view scenarios. Lan et al.[78] designed a new network architecture, called Y-net, which reconstructs PA images by optimizing both raw data and reconstructed images from the traditional method [Fig. 4(c)]. This network has two inputs, one from raw PA data and the other from the traditional reconstruction. It combines two encoders, each corresponding to one of the input paths, with a shared decoder path. The training data were generated by the k-Wave toolbox with a linear array setup. The public vascular data set[86] was used to generate PA signals. They compared the proposed
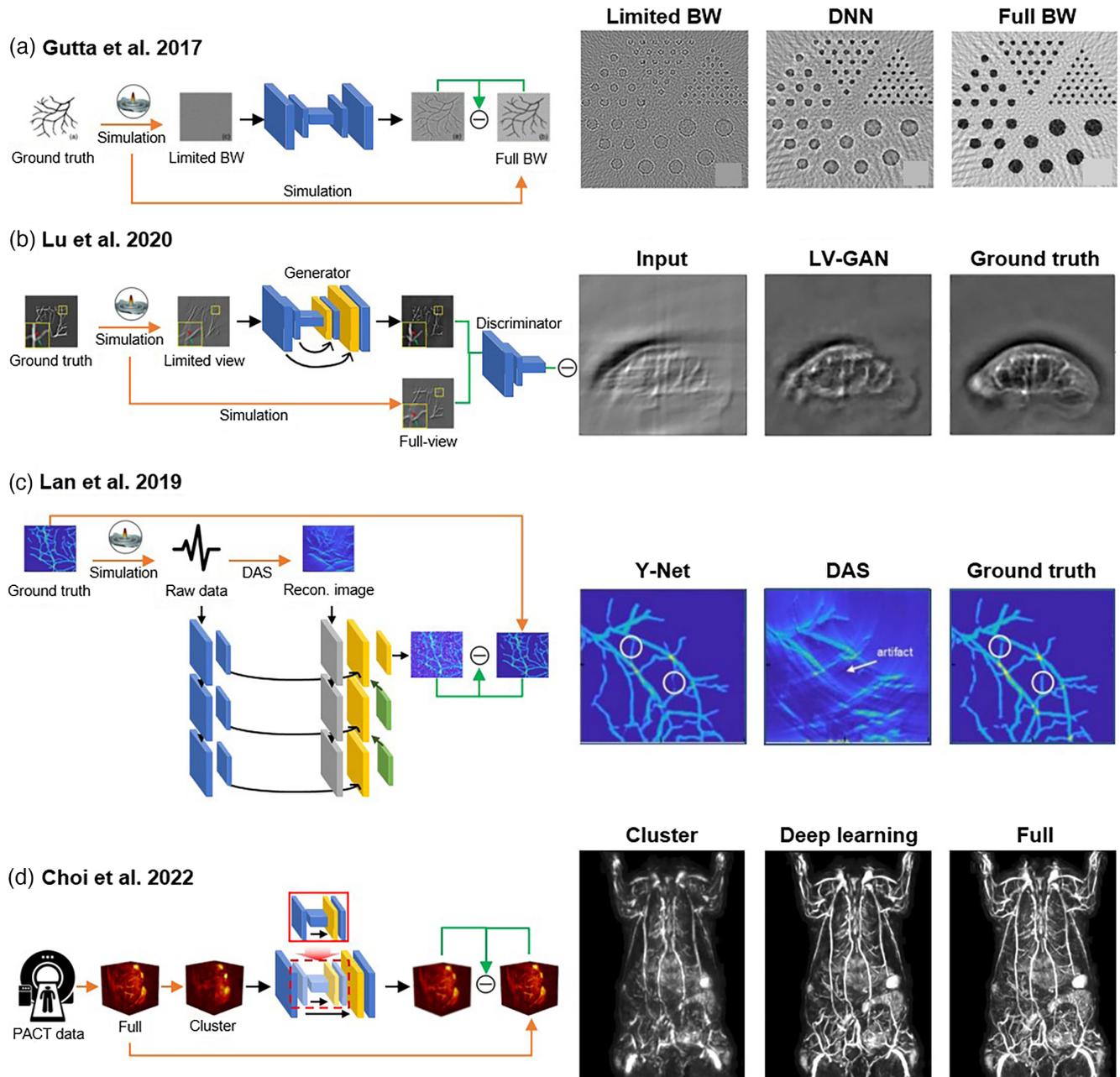
Fig. 4 Representative studies using DL methods to overcome limited-detection capabilities. (a) A DNN with five fully connected layers enhances bandwidth. (b) LV-GAN for addressing the limited-view problem. (c) A Y-Net generates the PA images by optimizing both raw data and reconstructed images from the traditional method. (d) A 3D progressive U-Net (3D-pUnet) to diminish the effects of limited-view artifacts and sparsity arising from cluster view detection. The images are adapted with permission from Ref. 76, © 2017 SPIE; Ref. 77, © 2020 Wiley-VCH GmbH; Ref. 78, © 2020 Elsevier GmbH; and Ref. 52, © 2021 Wiley-VCH GmbH. BW, bandwidth; DNN, deep neural network; DAS, delay-and-sum; cluster, cluster view detection; full, full view detection.

method with conventional reconstruction methods [e.g., DAS and time reversal (TR)] and other DL methods such as U-Net. In *in vitro* and *in vivo* experiments, the proposed method showed superior performance to the other methods, with the best spatial resolution.

3.1.3 Sparsity

To achieve the best image quality, the interval between two adjacent positions of the transducer or array elements must be less than half of the lowest detectable acoustic wavelength, according to the Nyquist sampling criterion.[87] In sparse sampling, the

actual detector density is lower than this requirement, introducing streak-shaped artifacts in images.[74] Sparse sampling can also result from a trade-off between image quality and temporal resolution, which is sometimes driven by system cost and hardware limitations.[88]

To remove artifacts caused by data sparsity, Guan et al.[89] added additional dense connectivity into the contracting and expanding paths of a U-Net. Farnia et al.[90] combined a TR method with a U-Net by inserting it in the first layer. Guo et al.[91] built a network containing a signal-processing method and an attention-steered network (AS-Net). Lan et al.[92] proposed a knowledge infusion GAN (Ki-GAN) architecture that combines DAS and PA signals for reconstruction from sparsely sampled data. DiSpirito et al.[93] compared various CNN architectures for PAM image recovery from undersampled data of *in vivo* mouse brains.[94] They chose a fully dense U-Net (FD U-Net) with a dense block, allowing PAM image reconstruction using just 2% of the original pixels. Later, they proposed a new method based on a deep image prior (DIP) method[95] to solve this problem without pretraining or GT data.

3.1.4 Combinational limited-detection problems

Previous studies have primarily tackled individual issues in isolation, neglecting the simultaneous occurrence of multiple limited-detection challenges in PA systems.[74] However, researchers have recently focused on utilizing a single NN to address two or three limited-detection problems concurrently, leading to promising advancements in this area.

For linear array, Godefroy et al.[96] incorporated dropout layers[97] into a modified U-Net and further built a Bayesian NN to improve the PA image quality. Vu et al.[98] built a Wasserstein GAN (WGAN-GP) that combined a U-Net and a deep convolutional GAN (DCGAN).[99] The network reduced limited-view and limited-bandwidth artifacts in PACT images. For a ring-shaped array, Zhang et al.[100] developed a 10-layer CNN, termed a ring-array DL network (RADL-net), to eliminate limited-view and under-sampling artifacts in photoacoustic tomography (PAT, also known as PACT) images. Davoudi et al.[101] proposed a U-Net network to improve the image quality from sparsely sampled data from a full-ring transducer array. They later updated their U-Net architecture[102] to operate on both images and PA signals. Awasthi et al.[103] proposed a U-Net architecture to achieve superresolution, denoising, and bandwidth enhancements. They replaced the softmax activation function in the final two layers of the U-Net for segmentation with an exponential linear unit.[104] Schwab et al. proposed a network that combined the BP with dynamic aperture length (DAL) correction, which they called DALnet[105] to address the limited-view and under-sampling issues in the 3D imaging PACT system.

One of the notable achievements in applying DL to the 3D-PACT system was made by Choi et al.[52] They introduced a 3D progressive U-Net (3D-pUnet) as a solution to address limited-view artifacts and sparsity caused by clustered-sampling detection, as shown in Fig. 4(d). The design of their network was inspired by the progressive growth GAN,[106] which utilizes a progressively increasing procedure to optimize a U-Net. In their 3D-pUnet, subnetworks were trained sequentially using downsampled data from the original high-resolution volume data, gradually transferring knowledge obtained from each progressive step.

The training data set consisted of *in vivo* experimental data from rats, and the results demonstrated superior performance compared with the conventional 3D-U-Net method. Interestingly, they demonstrated that the 3D-pUnet trained cluster-sampled data set also works in sparsely sampled data sets. The proposed approach was also applied to predict dynamic contrast-enhanced images and functional neuroimaging in rats, achieving increased imaging speed while preserving high image quality. In addition, they demonstrated the ability to accurately measure physiological phenomena and enhance structural information in untrained subjects, including tumor-bearing mice and humans.

All the research reviewed in this section is summarized in Table 3.

3.2 Compensating for Low-Dosage Light Delivery

Pulsed laser sources, such as an optical parametric oscillator laser system with a Nd:YAG pumped laser, are commonly used in PACT systems to achieve deep penetration with a high SNR, but those laser systems are bulky and expensive.[2,107] In recent years, researchers have explored compact and less expensive alternatives, such as pulsed-laser diodes[107] and light-emitting diodes (LEDs).[108] While these alternatives have shown promising results, their low pulse energy results in a low SNR, requiring frame averaging to increase image quality. Unfortunately, this method comes at a cost, as it reduces imaging speed. Furthermore, in dynamic imaging, frame averaging can cause blurring or ghosting due to the movement of the object being imaged. To address these problems, DL methods can be applied to enhance image quality in situations where the light intensity is low.

One of the representative works for LED-based systems was achieved by Hariri et al.[53] They proposed a multilevel wavelet-convolutional NN (MWCNN) that could map the low-fluence PA images to high-fluence PA images from an Nd:YAG laser system. This approach helps to eliminate the background noise while preserving the structures of the target, as shown in Fig. 5(a). Phantom and *in vivo* studies were conducted to assess the performance of their model. The MWCNN demonstrated a significant improvement in contrast-to-noise ratio (CNR) with up to a 4.3-fold enhancement in the phantom study and a 1.76-fold enhancement in the *in vivo* study. These results highlight the practicality of the proposed method in real-world scenarios. Singh et al.[111] and Anas et al.[112] proposed a U-Net and a deep CNN-based approach to improve the image quality with a similar system setup. Anas et al. later introduced a recurrent neural network (RNN)[113] to further improve the system's performance.[114]

To enhance the image quality in a pulsed-laser-diode PA system, Rajendran et al.[109] proposed a hybrid dense U-Net (HD-UNet) [Fig. 5(b)]. To train the network, they generated simulated data using the k-Wave toolbox, and evaluated the model with both single- and multi-US transducer (1-UST and multi-UST-PLD) PACT systems, using both phantom and *in vivo* images. Compared with their previous system, the HD-UNet improved the imaging speed by approximately 6 times in the 1-UST system and 2 times in the multi-UST-PLD system. To address the challenges of balancing laser dosage, imaging speed, and image quality in OR-PAM, Zhao et al.[110] proposed a multitask residual dense network (MT-RDN) that performs image denoising, superresolution, and vascular enhancement [Fig. 5(c)]. The network comprises three subnetworks, each using an independent RDN framework and assigned a supervised learning task. The first subnetwork

Table 3 Summary of overcoming the limited detection capabilities with DL approaches.

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded)		Test data set	Specified task	Representative evaluation results
			Source	Data amount			
Gutte et al.[76]	FC-DNN	CNN	Simulation of the breast phantom	286,300 slices (from 2863 volumes)	Simulation/in vitro phantom	Reduce limited-bandwidth artifacts	CNR (versus DAS) 0.01 → 2.54; PC 0.22 → 0.75
Deng et al.[83]	U-Net and VGG	U-Net	In vivo mouse liver	50	Numerical simulation data/in vitro phantom/in vivo data	Reduce limited-view artifacts from the circular US array	SSIM (versus DAS) 0.39 → 0.91; PSNR 7.54 → 24.34
Zhang et al.[85]	DuDoUnet	U-Net	k-Wave simulation	1500	k-Wave simulation	Reduce limited-view artifacts from the linear US array	SSIM (versus U-Net) 0.909 → 0.935; PSNR 19.4 → 20.8
Lu et al.[77]	LV-GAN	GAN	k-Wave simulation of absorbers and vessels/in vitro phantom of microsphere and vessel structure	793 pairs (absorbers)/1600 pairs (vessels)/30 pairs (microsphere) 22 pairs (vessel structures)	k-Wave simulation of absorbers and vessels/in vitro phantom (microsphere and vessel structure)	Reduce limited-view artifacts from the circular US array	SSIM (versus DAS) 0.135 → 0.871; PSNR 9.41 → 30.38; CNR 22.72 → 43.41
Lan et al.[78]	Y-Net	—	k-Wave simulation of segmented blood vessels from DRIVE data set	4700	k-Wave simulation/in vitro phantom/in vivo human palm	Reduce limited-view artifacts from the linear US array	SSIM (versus DAS) 0.203 → 0.911; PSNR 17.36 → 25.54; SNR 1.74 → 9.92
Guan et al.[89]	FD-UNet	U-Net	k-Wave simulation:/k-Wave simulation of realistic vasculature phantom from micro-CT images of mouse brain	1000 simulation/1000 (realistic vasculature)	k-Wave simulation:/k-Wave simulation of realistic vasculature phantom (micro-CT images of the mouse brain)	Reduce artifacts from sparse data in the circular US array	SSIM (versus DAS) 0.75 → 0.87; PSNR 32.48 → 44.84
Famia et al.[90]	U-Net	U-Net	k-Wave simulation from the DRIVE data set	3200	k-Wave simulation from DRIVE data set/in vivo mouse brain	Reduce artifacts from sparse data in the circular US array	SSIM (versus DAS) 0.81 → 0.97; PSNR 29.1 → 35.3; SNR 11.8 → 14.6; EPI 0.68 → 0.90
Guo et al.[91]	AS-Net	Non	k-Wave simulation of human fundus culi vessel/in vivo fish/in vivo mouse	3600/1744/1046	k-Wave simulation of human fundus culi vessel/in vivo fish/in vivo mouse	Reduce artifacts from sparse data and speed up reconstruction from the circular US array	SSIM (versus DAS) 0.113 → 0.985; PSNR 8.64 → 19.52

Table 3 (*Continued*).

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded) Source	Data amount	Test data set	Specified task	Representative evaluation results
Lan et al.[92]	Ki-GAN	GAN	k-Wave simulation of retinal vessels from public data set	4300	k-Wave simulation of retinal vessels from public data set	Remove artifacts from sparse data from the circular US array	SSIM (versus DAS) 0.215 → 0.928; PSNR 15.61 → 25.51; SNR 1.63 → 11.52
DiSpirito et al.[93]	FD U-Net	U-Net	*In vivo* mouse brain	304	*In vivo* mouse brain	Improve the image quality of undersampled PAM images	SSIM (versus zero fill) 0.510 → 0.961; PSNR 16.94 → 34.04; MS-SSIM 0.585 → 0.990; MAE 0.0701 → 0.0084; MSE 0.0027 → 0.00044
Vu et al.[95]	DIP	CNN	*In vivo* blood vessels	—	*In vivo* blood vessels/non-vascular data	Improve the image quality of undersampled PAM images	SSIM (versus bilinear) 0.851 → 0.928; PSNR 25.6 → 31.0
Godefroy et al.[96]	U-Net/Bayesian NN	U-Net	Pairs of PAI and photographs of leaves/Corresponded numerical simulation	500	PAI and photographs of leaves/numerical simulation	Reduce limited-view and limited-bandwidth artifacts from the linear US array	NCC (versus DAS) 0.31 → 0.89; SSIM 0.29 → 0.87
Vu et al.[98]	WGAN-GP	GAN	k-Wave simulation: disk phantom and TPM vascular data	4000 (disk)/7200 (vascular)	k-Wave simulation: disk phantom and TPM vascular data/tube phantom/*in vivo* mouse skin	Reduce limited-view and limited-bandwidth artifacts from the linear US array	SSIM (versus U-Net) 0.62 → 0.65; PSNR 25.7 → 26.5
Zhang et al.[100]	RADL-net	CNN	k-Wave simulation	161,000 (including augmentation and cropping from 126 vascular images)	k-Wave simulation/vascular structure phantom/*in vivo* mouse brain	Reduce limited-view and sparsity artifacts from the ring-shaped US array	SSIM (versus DAS) 0.11 → 0.93; PSNR 17.5 → 23.3
Davoudi et al.[101]	U-Net	U-Net	Simulation: planar parabolic absorber and mouse/*in vitro* circular phantom/*in vitro* vessel-structure phantom/*in vivo* mouse	Not mentioned/28/33/420	Simulation: planar parabolic absorber and mouse/*in vitro* circular phantom/*in vitro* vessel-structure phantom/*in vivo* mouse	Reduce limited-view and sparsity artifacts from the circular US array	SSIM (versus input) 0.281 → 0.845

Table 3 (*Continued*).

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded)		Test data set	Specified task	Representative evaluation results
			Source	Data amount			
Davoudi et al.[102]	U-Net	U-Net	*In vivo* human finger from seven healthy volunteers	4109 (including validation)	*In vivo* human finger	Reduce the limited-view and sparsity artifacts from the US circular array	SSIM (versus U-Net) 0.845 → 0.944 PSNR 14.3 → 19.0 MSE 0.04 → 0.014 NRMSE 0.818 → 0.355
Awasthi et al.[103]	Hybrid end-to-end U-Net	U-Net	k-Wave simulation from breast sinogram images	1000	k-Wave simulation of the numerical phantom, blood vessel, and breast/ horsehair phantoms/ *in vivo* rat brain	Super-resolution, denoising, and bandwidth enhancement of the PA signal from the circular US array	PC (versus DAS) 0.307 → 0.730 SSIM 0.272 → 0.703 RMSE 0.107 → 0.0617
Schwab et al.[105]	DALnet	CNN	Numerical simulation of 200 projection images from 3D lung blood vessel data	3000 (after cropping)	Numerical simulation/ *in vivo* human finger	Reduce limited-view, sparsity and limited bandwidth artifacts	SSIM (versus input) 0.305 → 0.726 Correlation 0.382 → 0.933
Choi et al.[52]	3D-pUnet	U-Net	*In vivo* rat	1089	*In vivo* rat/*in vivo* mouse/*in vivo* human	Reduce limited-view and sparsity artifacts	MS-SSIM (versus input) 0.83 → 0.94 PSNR 32.0 → 34.8 RMSE 0.025 → 0.019

CNR, contrast-to-noise ratio; PC, Pearson's correlation coefficient; SSIM, structural similarity index; MS-SSIM, multiscale SSIM; MAE, mean absolute value; MSE, mean squared error; NCC, normalized 2D cross-correlation; TPM, three photon microscopy; sSSIM, shifted structured similarity index; NRMSE, normalized root mean squared error.
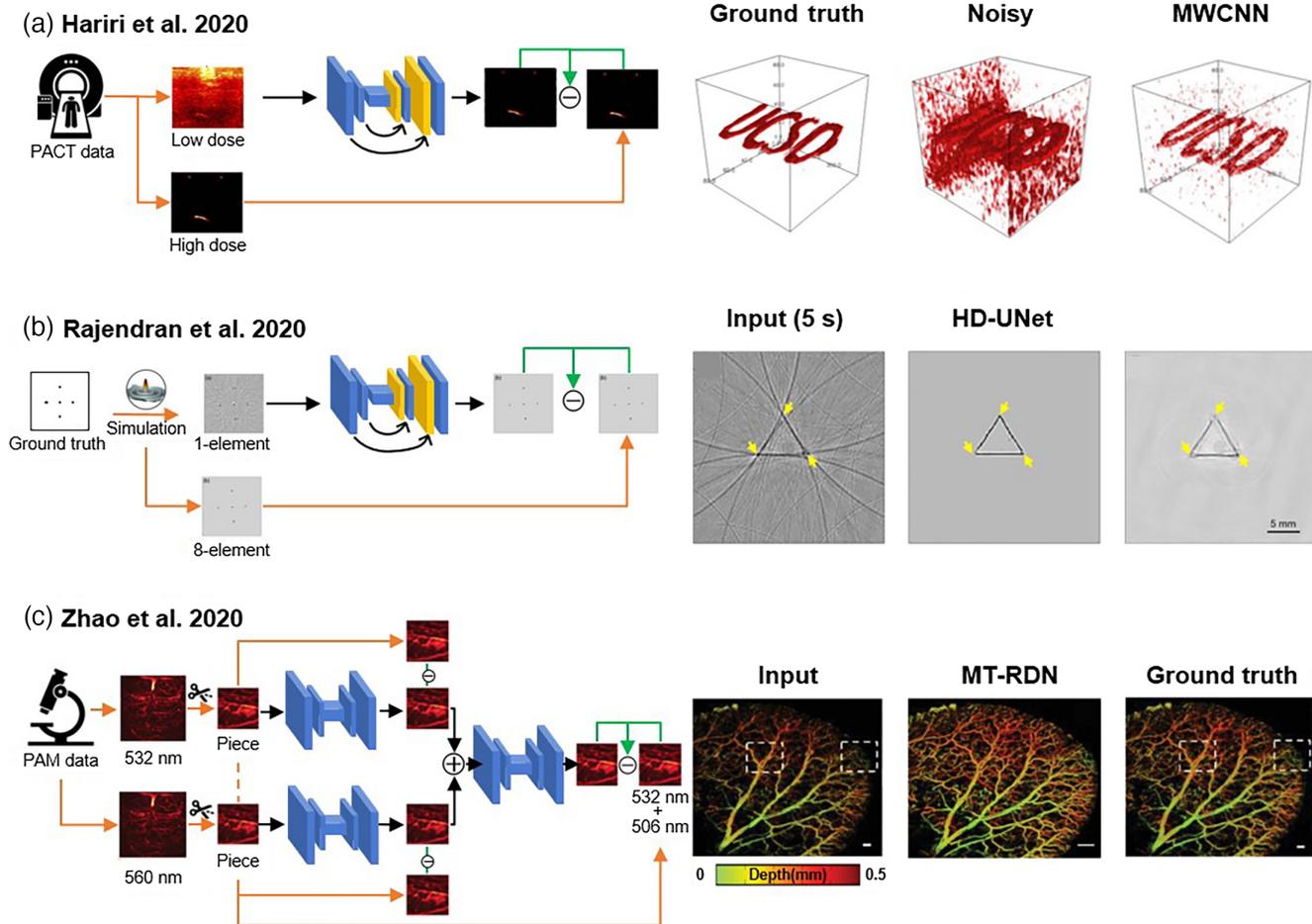
Fig. 5 Representative DL approaches compensate for low laser dosage. (a) An MWCNN that generates high-quality PA images from low-fluence PA images. (b) An HD-UNet that enhances the image quality in a pulsed-laser diode PACT system. (c) An MT-RDN that performs image denoising, superresolution, and vascular enhancement. The images are adapted with permission from Ref. 53, © 2020 Optica; Ref. 109, © 2022 SPIE; and Ref. 110, © 2020 Wiley-VCH GmbH. MWCNN, multi-level wavelet-convolutional neural network; HD-UNet, hybrid dense U-Net; MT-RDN, multitask residual dense network.

processes the data of input 1 (i.e., 532 nm data) to obtain output 1, and the second subnetwork processes the data of input 2 (i.e., 560 nm data) to obtain output 2. These outputs are then combined and processed by subnetwork 3, and the differences between the outputs and the GT are compared.

To train the network, input images were undersampled at half-per-pulse laser energy of the GT, while the GT images were sampled at the full ANSI per-pulse fluence limit. To evaluate the performance of the proposed method, U-Net and RDN were used. The MT-RDN method achieved a 16-fold reduction in laser dosage at 2 times data undersampling and a 32-fold reduction in dosage at 4 times undersampling compared to the GT images.

All the research reviewed in this section is summarized in Table 4.

3.3 Improving the Accuracy of Quantitative PAI

Quantitative photoacoustic imaging (qPAI) quantifies molecular concentrations in biological tissue using multiwavelength PA images, enabling the estimation of various endogenous and exogenous contrast agents and physiological parameters, such as sO_2.[61] However, qPAI presents significant challenges due to the wavelength-dependent nature of light absorption and scattering, leading to varying levels of light attenuation across different wavelengths.[2,115] Thus, it is hard to accurately determine the fluence distribution, which is nonlinear and complex in biological tissues. Early research in qPAI assumed constant optical properties of biological tissue and uniform parameters such as the scattering coefficient throughout the imaging field.[61] However, recent studies have shown that these assumptions lead to errors, especially in deep-tissue imaging.[116] Model-based iterative optimization methods have been developed to address this issue and provide more accurate solutions.[117] But these methods are time-consuming and sensitive to quantification errors.[118] A new approach called eigenspectral multispectral optoacoustic tomography (eMSOT) has been proposed to improve qPAI accuracy.[116] eMSOT formulates light fluence in tissues as an affine function of reference base spectra, leading to improved accuracy in qPAI.

Table 4 Summary of studies on compensating for low-dosage light delivery.

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded)		Test data set	Specified task	Representative evaluation results
			Source	Data amount			
Hariri et al.[53]	MWCNN	U-Net	Agarose hydrogel phantom: LED-based PA image and Nd:YAG-based PA image	229	Agarose hydrogel phantom of LED-based PA image/in vivo mouse	Denoise PA images from low-dosage system	SSIM (versus input) 0.63 → 0.93 PSNR 15.58 → 53.88
Singh et al.[111]	U-Net	U-Net	LED-based and Nd:YAG-based tube phantom	150	LED-based phantom using ICG and MB	Reduce the frame averaging	SNR 14 → 20
Anas et al.[112]	—	CNN	In vitro phantom	4536	In vivo fingers	Improve the quality of PA images	SSIM (versus average) 0.654 → 0.885 PSNR 28.3 → 36.0
Anas et al.[114]	—	CNN and LSTM	In vitro wire phantom/in vitro nanoparticle phantom	352,000	In vitro phantom/in vivo human fingers	Improve the quality of PA images	SSIM (versus input) 0.86→0.96 PSNR 32.3 → 37.8
Rajendran et al.[109]	HD-Unet	U-Net	k-Wave simulation	450	In vitro phantom/in vivo rat	Improve the frame rate	SSIM (versus U-Net) 0.92 → 0.98 PSNR 28.6 → 32.9 MAE 0.025 → 0.017
Zhao et al.[110]	MT-RDN	—	In vivo mouse brain and ear	6696	In vivo mouse brain and ear	Improve the quality from low dosage laser and downsampled data	SSIM (versus input) 0.64 → 0.79 PSNR 21.9 → 25.6

ICG, indocyanine green; MB, methylene blue; LSTM, long short-term memory.

However, it requires *ad hoc* inversion and has limitations in scale invariance.

Researchers have pursued multiple avenues to extract fluence distribution information from multiwavelength PA images using DL architectures. Cai et al.[119] introduced ResU-Net, which adds a residual learning mechanism to the U-Net. Chang et al.[120] developed DR2U-Net, a fine-tuned deep residual recurrent U-Net. Luke et al.[121] combined two U-Nets to create a new network called O-Net, which segments blood vessels and estimates sO$_2$. A novel DL architecture that contains an encoder, decoder, and aggregator was introduced by Yang et al.[122] termed called EDA-Net. The encoder and decoder paths both feature a dense block, while the aggregator path incorporates an aggregation block. Gröhl et al.[123] designed a nine-layer fully connected NN that directly estimates sO$_2$ from PA images. All showed much high accuracy in estimating sO$_2$ distributions or other molecular concentrations compared with linear unmixing.

One of the representative results is from Ref. 124. Researchers built two separated convolutional encoder–decoder type networks with skip connections, termed EDS to solve this problem in 3D conditions [Fig. 6(a)]. One network was trained to output images of sO$_2$ from 3D-image data and the other network was trained to segment vessels. By leveraging the spatial information present in the 3D images, the 3D fully convolutional networks could produce precise sO$_2$ maps. Besides getting more accurate sO$_2$ results, these networks were able to handle limited-detection capabilities, such as limited-view artifacts, and showed promise for producing accurate estimates *in vivo*.

Researchers have also employed DL methods to recover the absorption coefficient from reconstructed PA images. Chen et al.[126] proposed a U-Net-based DL network to recover the optical absorption coefficient and Grohl et al.[127] adapted a U-Net to compute error estimates for optical parameter estimations. A notable contribution was made by Li et al. in a recent study.[54] They addressed the challenge of insufficient data-label pairs in qPAI by introducing two DNNs, depicted in Fig. 6(b). First, they introduced a simulation-to-experiment end-to-end data translation network (SEED-Net) that provides GT images for experimental images through unsupervised data translation from a simulation data set. They then designed a dual-path network based on U-Net (QPAT-Net) to reconstruct images of the absorption coefficient for deep tissues. The QPAT-Net outperformed the previous QPAT method[128] in simulation, *ex vivo*, and *in vivo*, with more accurate absorption information and relatively few errors.

Another seminal study was done by Zou et al.[125] They developed the US-enhanced U-Net model (US-Unet), which combines information from US images and PA images to reconstruct the optical absorption distribution [Fig. 6(c)]. They implemented a pretrained ResNet-18 to extract features from US images of ovarian lesions.

This feature information was incorporated into a U-Net structure designed to reconstruct the optical absorption coefficient. The U-Net was trained on simulation data and subsequently tested on a phantom, blood tubes, and clinical data from 35 patients. The US-Unet outperformed both the U-Net model without US features and the standard DAS method in phantom and clinical studies, demonstrating its potential for improving accuracy in clinical PAI applications.

Compensating for the distribution of light fluence can improve the accuracy of qPAI.[129–131] To this end, Madasamy et al.[132] compared the compensation performance of different DL models. The models tested included U-Net,[69] FD U-Net,[89] Y-Net, FD Y-Net,[78] deep residual U-Net (deep ResU-Net),[133] and GAN.[134] Results showed the robustness of all DL models to noise and their effectiveness; FD U-Net showed the best performance. qPAI requires an unmixing process, which can be achieved through linear or model-based methods.[115] Durairaj et al.[135] proposed an unsupervised learning approach using an initialization network and an unmixing network. Olefir et al.[136] introduced DL-eMSOT, combining eMSOT with a bidirectional RNN and CNN blocks for accurate sO$_2$ estimation and faster calculations.

All the research reviewed in this section is summarized in Table 5.

3.4 Optimizing or Replacing Conventional Reconstruction Algorithms

In PACT, the acoustic inverse problem involves reconstructing the PA initial pressure from raw data. Several reconstruction methods have been developed, including BP,[47] FB,[48] DAS,[44] DMAS,[46] TR,[49] and model-based methods.[51] However, each method has limitations, and either to enhance existing reconstruction techniques or to directly reconstruct PA images using NNs, researchers have turned to DL methods.

Various DL methods have been developed to convert PA raw data into images. One such method, called Pixel-DL, proposed by Guan et al.,[137] uses pixel-wise interpolation followed by an FD U-Net for limited-view and sparse PAT image reconstruction [Fig. 7(a)]. The Pixel-DL model was trained and tested using simulated PA data from synthetic, mouse brain, lung, and fundus vasculature phantoms. It achieved comparable or better performance than iterative methods and consistently outperformed other CNN-based approaches for correcting artifacts.

To direct reconstruct PA images, Waibel et al.[139] introduced a modified U-Net that includes additional convolutional layers in each skip connection. Antholzer et al.[140] proposed a direct reconstruction process, based on a U-Net and a simple CNN, that can resolve limited-view and sparse-sampling issues. Lan et al.[141] proposed a modified U-Net, termed DU-Net, to reconstruct PA images using multifrequency US-sensor raw data. A noteworthy study of this topic is an end-to-end reconstruction network developed by Feng et al.,[138] termed Res-U-Net [Fig. 7(b)]. They integrated residual blocks into the contracting and symmetrically expanding path of U-Net and added a skip connection between the input of raw data and the output of images. The training, validation, and test data sets were synthesized using the k-Wave toolbox. In digital phantom experiments, the Res-UNet showed performance superior to other reconstruction methods [Fig. 7(b)].

Another representative work was done by Tong et al.[55] They proposed a novel two-step reconstruction process with a feature projection network (FPnet) and a U-Net [Fig. 7(c)]. The FPnet converts PA signals to images and contains several convolutional layers to extract features. There is one max pooling layer for downsampling and one full connection layer for domain transformation. The U-Net performs postprocessing to improve image quality. The resulting network, trained using numerical simulations and *in vivo* experimental data, outperformed other approaches to handle limited-view and sparsely
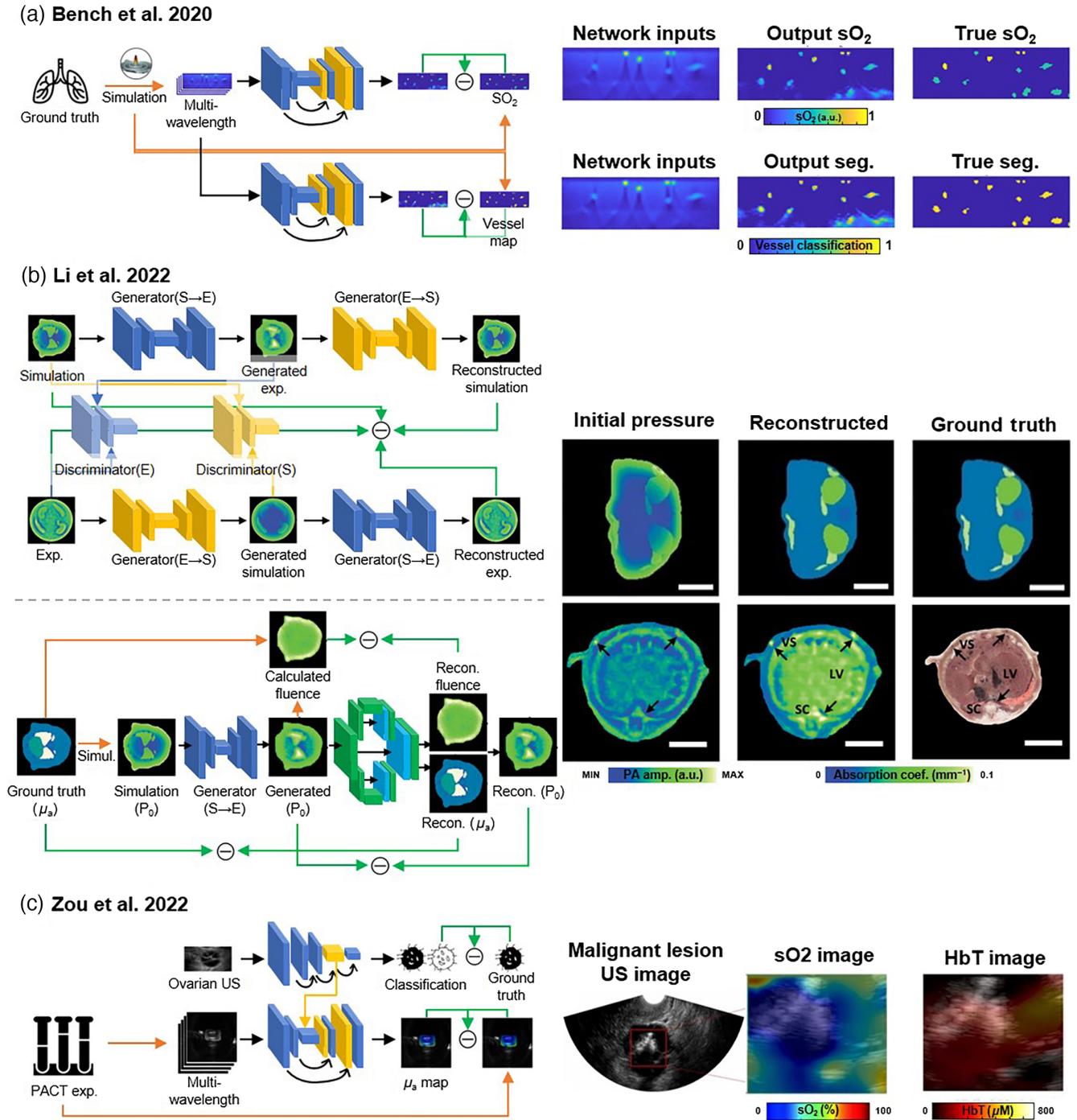
Fig. 6 Representative studies to improve the accuracy of quantitative PAI by DL. (a) Convolutional encoder–decoder type network with skip connections (EDS) to produce accurate estimates of sO_2 in a 3D data set. (b) Dual-path network based on U-Net (QPAT-Net) to reconstruct images of the absorption coefficient for deep tissues. (c) US-enhanced U-Net model (US-UNet) to reconstruct the optical absorption distribution. The images are adapted with permission from Ref. 124, © 2020 SPIE; Ref. 54, © 2022 Optica; and Ref. 125, © 2022 Elsevier GmbH.

sampled experimental data, exhibiting superior performance on *in vivo* experiments [Fig. 7(c)].

In addition, Yang et al.[142] introduced recurrent inference machines (RIM), an iterative PAT reconstruction method using

convolution layers. Kim et al.[143] employed upgUNET, a U-Net model with 3D transformed arrays for image reconstruction. Hauptmann et al.[144] proposed DGD, a deep gradient descent algorithm, outperforming U-Net and other model-based

Table 5 Summary of studies to improve the accuracy of quantitative PAI.

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded) Source	Data amount	Test data set	Specified task	Representative evaluation results
Cai et al.[119]	ResU-net	U-Net	Numerical simulation	2048	Numerical simulation	Extract information from multispectral PA images	Relative errors (versus linear unmixing) 36.9% → 0.76%
Chang. et al.[120]	DR2U-net	U-Net	Monte Carlo simulation of simulated tissue structure	2560	Monte Carlo simulation of simulated tissue structure	Extract fluence distribution from optical absorption images	Relative Errors (versus linear unmixing) 48.76% → 1.27%
Luke et al.[121]	O-Net:	U-Net	Monte Carlo simulation of epidermis, dermis, and breast tissue	1600 pairs (one pair has two-wavelength PA data)	Monte Carlo simulation of epidermis, dermis, and breast tissue	Estimate the oxygen saturation and segment	Relative errors (versus linear unmixing) 43.7% → 5.15%
Yang et al.[122]	EDA-net	—	Monte Carlo and k-Wave simulation from female breast phantom	4888	Monte Carlo and k-Wave simulation based on clinically obtained female breast phantom	Extract the information from the multi-wavelength PA images	Relative errors (versus linear unmixing) 41.32% → 4.78%
Gröhl et al.[123]	Nine-layer fully connected NN	CNN	Monte Carlo simulation of in silico vessel phantoms	776	In vivo porcine brain and human forearm	Obtain quantitative estimates for blood oxygenation	No statistical results
Bench et al.[124]	EDS	U-Net	k-Wave simulation of human lung from lung CT scans/k-Wave simulation of three-layer skin model	—	k-Wave simulation	Produce 3D maps of vascular sO_2 and vessel positions	Mean difference (versus linear unmixing) 6.6% → 0.3%
Chen et al.[126]	U-Net	U-Net	Monte Carlo simulation	2880	In vitro phantom	Recover the optical absorption coefficient	Relative error less than 10%
Gröhl et al.[127]	U-Net	U-Net	Monte Carlo and k-Wave simulations of in silico tissue	3600	Monte Carlo and k-Wave simulation of in silico	Improve optical absorption coefficient estimation	Estimation error (versus linear unmixing) 58.3% → 3.1%
Li et al.[54]	Two GANs: SEED-Net and QOAT-Net	GAN	Numerical simulation of phantom, mouse, and human brain/experimental data of phantom, ex vivo, and in vivo mouse	3040, 2560, and 2560/2916, 3200, and 3800	Ex vivo porcine tissue, mouse liver, and kidney/In vivo mouse	Improve optical absorption coefficient estimation	Relative errors (versus linear unmixing) 8.00% → 4.82% Relative errors 8.00% → 4.82%
Zou et al.[125]	US-Unet	U-Net	Monte Carlo and k-Wave simulation/in vitro phantom	2000/480	In vitro blood tube/in vivo clinical data set	Improve optical absorption coefficient estimation	Accuracy (versus linear unmixing) 0.71 → 0.89

Table 5 (*Continued*).

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded)		Test data set	Specified task	Representative evaluation results
			Source	Data amount			
Madasamy et al.[132]	Network comparing: U-Net, FD U-Net, Y-Net, FD Y-Net, Deep ResU-Net, and GAN	—	2D numerical simulation of retinal fundus (from Kaggle and RFMID)/3D numerical simulation of breast phantom	1858 (before augmentation)/5 3D volumes (12,288 slices after augmentation)	2D numerical blood vessel/3D numerical breast phantom	Fluence correction	PSNR (versus linear unmixing) 37.9 → 45.8 SSIM 0.80 → 0.96
Durairaj et al.[135]	Two networks: initialization network and unmixing network	—	NIRFAST and k-Wave simulation	Not mentioned	NIRFAST and k-Wave simulation	Unmix the spectral information	Regardless of prior spectral information
Olefir et al.[136]	DL-eMSOT: bi-directional RNN with two LSTMs	—	Monte Carlo simulation	10,944	*In vitro* phantom/*in vivo* mouse	Replace inverse problem of eMSOT	Mean error (versus eMSOT) 4.9% → 1.4% Median error 3.5% → 0.9% Standard deviation 4.8% → 1.5%
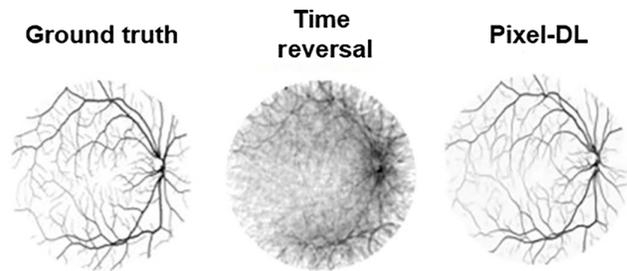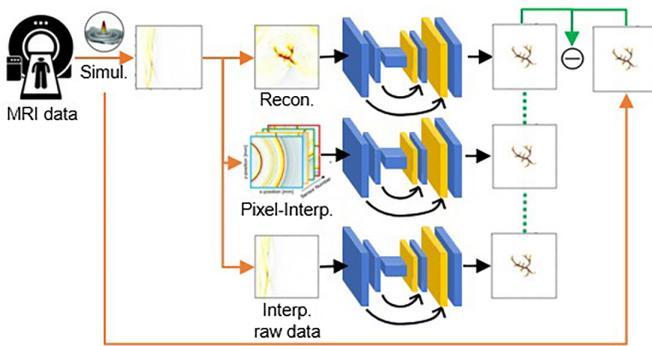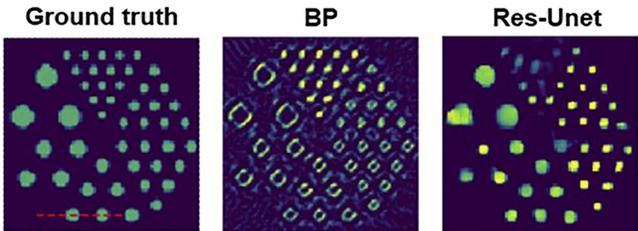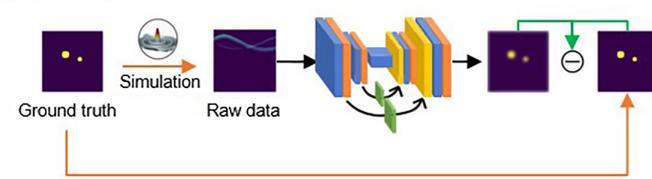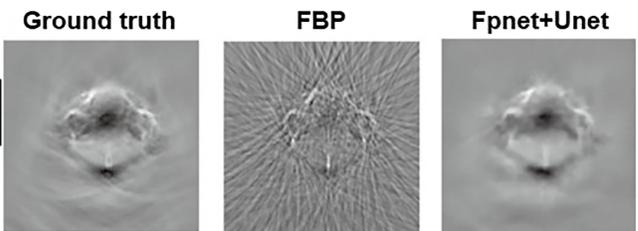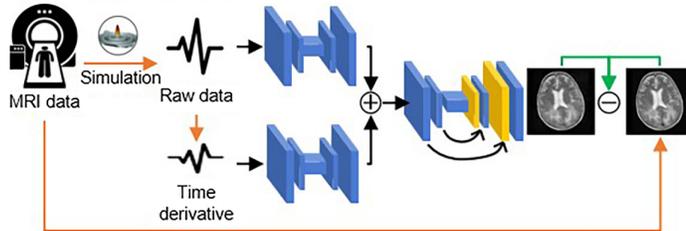
Fig. 7 Representative studies to optimize conventional reconstruction algorithms or replace them with DL. (a) Pixel-wise interpolation approach followed by an FD-UNet for limited-view and sparse PAT image reconstruction. (b) End-to-end U-Net with residual blocks to reconstruct PA images. (c) Two-step PA image reconstruction process with FPnet and U-Net. The images are adapted with permission from Ref. 137, © 2020 Nature Publishing Group; Ref. 138, © 2020 Optica; and Ref. 55, © 2020 Elsevier GmbH.

methods. They also introduced fast-forward PAT (FF-PAT), a modified version of DGD, which addressed artifacts using a small multiscale network.[145]

All the research reviewed in this section is summarized in Table 6.

3.5 Addressing Tissue Heterogeneity

Biological tissues are acoustically nonuniform, making it crucial to use a locally appropriate speed of sound (SoS) value for accurate PA reconstruction. SoS mismatch or a discontinuity in hard textured tissue can create acoustic reflection and imaging artifacts[74] that make it hard to detect the source of the PA signal, which is especially troublesome for interventional applications. DL methods have been used to detect point sources, remove reflections, and mitigate the difficulties presented by acoustic heterogeneity.

Highly echogenic structures can cause a reflection of a PA wave to appear to be a true signal,[146] which makes it hard to find point targets or real sources in PAI. Reiter et al.[147] trained a CNN to identify and remove reflection noise, locate point targets,

and calculate absorber sizes in PAI. Later, Allman et al.[148] found Fast-RCNN[149] to be more effective than VGG16[84] for source detection and artifact elimination. Shan et al.[150] incorporated a DNN into an iterative algorithm to correct reflection artifacts, achieving superior results compared with other methods.[151,152]

Jeon et al.[56] proposed a generalized solution to mitigate SoS aberration in heterogeneous tissue by DL. They proposed a hybrid DNN model, named SegU-net, based on U-Net and SegNet[153] [Fig. 8(a)]. The architecture is similar to SegNet, but has an additional connection between the encoder and decoder through concatenation layers, like U-Net. The training data were generated using the k-Wave toolbox with different SoS values. They tested the model with phantoms with homogeneous media and in heterogeneous media. The proposed method showed better results than the multistencil fast marching[155] method and automatic SoS selection.[156] It not only resolved the SoS aberration but also removed streak artifacts in images of healthy human limbs and melanoma.

All the research reviewed in this section is summarized in Table 7.

Table 6 Summary of methods to optimize or replace conventional reconstruction algorithms.

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded) Source	Data amount	Test data set	Specified task	Representative evaluation results
Guan et al.[137]	Pixel-DL	U-Net	k-Wave simulation: circles, Shepp-Logan, and vasculature, vasculature phantom from micro-CT images of mouse brain	1000 (circles),/1000 (realistic vasculature)	k-Wave simulation: circles, Shepp-Logan, vasculature, vasculature phantom from micro-CT images of mouse brain	Reconstruct PA images from PA signal	PSNR (versus TR) 17.49 → 24.57 SSIM 0.52 → 079
Waibel et al.[139]	U-Net	U-Net	Monte Carlo and k-Wave simulation	2304	Monte Carlo and k-Wave simulation	Reconstruct PA images from PA signal	IQR (versus DAS) 98% → 10%
Antholzer et al.[140]	U-Net	U-Net	Numerical simulation of ring-shaped phantoms	1000	Numerical simulation of ring-shaped phantoms	Reconstruct PA images from PA signal	MSE (versus general CNN) 0.33 → 0.026
Lan et al.[141]	DU-Net	U-Net	k-Wave simulation: disc phantom and segmented fundus oculi/vessels CT	4000	k-Wave simulation: disc phantom and segmented fundus oculi/ vessels CT	Reconstruct PA images from PA signal	PSNR (versus DAS) 26.843 → 44.47 SSIM 0.394 → 0.994
Feng et al.[138]	Res-UNet	U-Net	k-Wave simulation: disc bread, spider (from "quick draw"), simple wires, logos, natural phantom	58,126 (80% of 27,000, 13,000, 10,800, 6000, 240, 15,000)	k-Wave simulation: disc, PAT, vessel/in vitro phantom	Reconstruct PA images from PA signal	Vessel phantom PC (versus MRR) 0.41 → 0.80 PSNR 6.57 → 13.29
Tong et al.[55]	FPnet+ U-Net	U-Net	Numerical simulation: brain from MRI, abdomen from MRI, vessel from DRIVE data set/in vivo mouse brain and abdomen	15,757: 2211 (brain), 8273 (abdomen), 4000 (vessel)/698 (mouse brain), 575 (mouse abdomen)	Numerical simulation: brain, abdomen and liver cancer from MRI, vessel/in vivo mouse brain and abdomen	Reconstruct PA images from PA signal	MSOT-Abdomen PSNR (versus FBP) 16.0532 → 30.3972 SSIM 0.2647 → 0.9073 RMSE 0.4771 → 0.0910
Yang et al.[142]	RIM	—	k-Wave simulation of segmented blood vessels from DRIVE data set	2400	k-Wave simulation of segmented blood vessels from DRIVE	Reconstruct PA images from PA signal	PSNR (versus DGD) 42.37 → 44.26
Kim et al.[143]	upgUNET	U-Net	Monte Carlo simulation	128,000 (after augmentation)	Monte Carlo simulation/in vitro metal-wire phantom/in vivo human finger	Reconstruct PA images from PA signal	PSNR (versus DAS) 20.97 → 27.73 SSIM 0.208 → 0.754
Hauptmann et al.[144]	Updated DGD (deep gradient descent)	DGD (deep gradient descent)	k-Wave simulation pf human lung from 50 whole-lung CT scans	1024 (from 50 CT scans)	k-Wave simulation pf human lung from 50 whole-lung CT scans/ in vivo human palm	Reconstruct PA images from PA signal	PSNR (versus U-Net) 40.81 → 41.40 SSIM 0.933 → 0.945
Hauptmann et al.[145]	FF-PAT	U-Net	k-Wave simulation of human lung from lung CT scans	1024 (from 50 CT scans)	k-Wave simulation of human lung/in vivo data	Reconstruct PA images from PA signal	PSNR (versus BP) 33.5672 → 42.1749

MRR, model-resolution-based regularization algorithm.
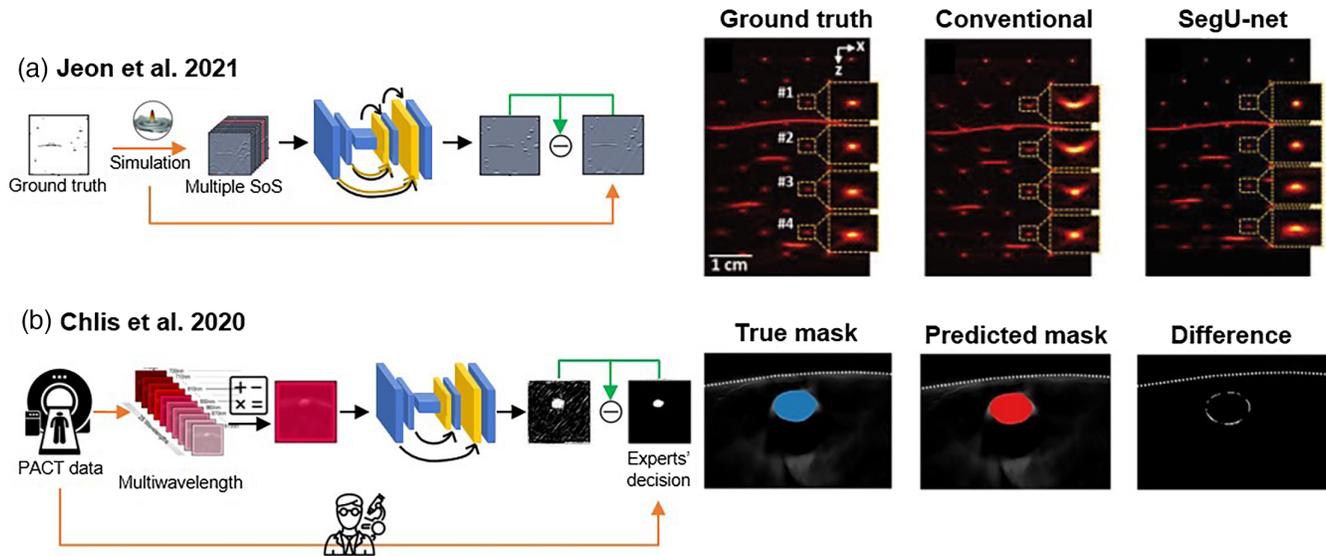
Fig. 8 Representative DL studies to correct the SoS and improve the accuracy of image classification and segmentation. (a) Hybrid DNN model including U-Net and Segnet to mitigate SOS aberration in heterogeneous tissue. (b) Sparse-UNet (S-UNet) for automatic vascular segmentation in MSOT images. The images are adapted with permission from Ref. 153, CC-BY; Ref. 154, © Elsevier GmbH.

Table 7 Summary of methods for addressing tissue heterogeneity.

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded) Source	Data amount	Test data set	Specified task	Representative evaluation results
Reiter et al.[147]	CNN	CNN	k-Wave simulation	19,296	k-Wave simulation/ _in vitro_ vessel-mimicking target phantom	Identify point source	—
Allman et al.[148]	CNN consisting of VGG16/Fast R-CNN	CNN	k-Wave simulation	15,993	k-Wave simulation/ _in vivo_ data	Identify and remove reflection artifacts	Precision, recall, and AUC > 0.96
Allman et al.[157]	CNN consisting of VGG16/fast R-CNN	CNN	k-Wave simulation	15,993	_In vitro_ phantom	Correct reflection artifact	Accuracy (phantom) 74.36%
Shan et al.[150]	U-Net	U-Net	Numerical simulation from 3 cadaver CT	64,000	Numerical simulation from 1 cadaver CT	Correct reflection artifacts	PSNR (versus TR) 9 → 29 SSIM (versus TR) 0.2 → 0.9
Jeon et al.[56]	SegU-net	U-Net	k-Wave simulation of _in silico_ phantom	270	k-Wave simulation of _in silico_ phantom/ _in vivo_ human forearm and foot	Reduce speed-of-sound aberrations	_In silico_ phantom SSIM (versus pre-corrected) + 0.24

3.6 Improving the Accuracy of Image Classification and Segmentation

As PAI gains increasing attention in clinical studies, more accurate classification and segmentation methods are necessary to improve the interpretation of PA images. Image segmentation extracts the outline of objects within an image and identifies the distinct parts of the image that correspond to these objects.[158,159] Image classification predicts a label for an image, identifying its content. In this section, we focus on segmentation and classification techniques.[158]

Segmentation and classification are widely used in image postprocessing, and transfer learning has been utilized to take advantage of pretrained DNNs. Zhang et al.[160] used DL models AlexNet[161] and GoogLeNet[162] for PA image classification, outperforming support vector machine (SVM). Jnawali et al.[163] employed transfer learning with Inception-ResNet-V2[164] for thyroid cancer detection and introduced a deep 3D CNN for cancer detection in multispectral photoacoustic data sets.[165] Moustakidis et al.[166] developed SkinSeg for identifying skin layers in raster-scan optoacoustic mesoscopy (RSOM) images, evaluating decision trees,[167] SVM,[168] and DL algorithms. Nitkunanantharajah et al.[169] achieved good classification performance with ResNet18[164] on RSOM nail fold images.

PACT has demonstrated its great potential for human vascular imaging in several clinical studies.[15] However, segmentation of the vascular structures, particularly the vascular lumen, is still accomplished through manual delineation by expert physicians, which is not only time-consuming but also subjective. To address this issue, Chlis et al.[154] proposed a sparse UNet (S-UNet) for automatic vascular segmentation on MSOT images. GT is obtained from binary images extracted from MSOT images, based on consensus between two clinical experts [Fig. 8(b)]. The MSOT raw data from six healthy humans' vasculature were acquired using a handheld MSOT system, and they were split into training, validation, and test sets. The S-Unet showed performance similar to other U-Net methods, but with smaller parameter sizes and the ability to select wavelengths, indicating its potential for clinical application.[170]

In addition, Lafci et al.[171] proposed a U-Net architecture to accurately segment animal boundaries in hybrid PA and US (PAUS) images. Boink et al.[172] proposed a learned primal-dual (L-PD) algorithm based on a CNN to solve the reconstruction and segmentation problem simultaneously. Ly et al.[57] introduced a modified U-Net DL model for automatic skin and vessel segmentation in *in vivo* PAI. The U-Net architecture showed the best performance.

All the research reviewed in this section is summarized in Table 8.

3.7 Overcoming Other Specified Issues

In addition to the general challenges to PAI mentioned above, researchers encounter several more specific problems with their imaging systems. Collectively, these specific issues constitute a seventh challenge, and among them we have identified six representative categories: motion artifacts, limited spatial resolution, electrical noise and interference, image misalignment, slow accelerating superresolution imaging, and achieving digital histologic staining.

Motion artifacts caused by breathing or heartbeats can significantly reduce image quality in PAM and PA endoscopy (PAE or intravascular PA, IVPA). To address this issue, researchers have presented various breathing artifact removal methods,[173,174] and DL methods have recently been proposed as a potential solution. Chen et al.[175] introduced a CNN approach with three convolutional layers to address motion artifacts and pixel dislocation in *in vivo* rat brain images. Zheng et al.[176] proposed MAC-Net, a network based on VGG16 GAN[134] and spatial transformer networks (STN),[177] to suppress motion artifacts in IVPA. Both methods demonstrated successful improvement in image quality.

OR-PAM can penetrate ~1 mm deep in biological tissue, limited by light scattering. AR-PAM, which does not use focused light, can penetrate up to several centimeters, but it has a lower spatial resolution than OR-PAM. Researchers have applied DL to enhance the spatial resolution of AR-PAM to match that of OR-PAM. Cheng et al.[178] proposed a GAN-based framework called Wasserstein GAN[179] [Fig. 9(a)]. An integrated OR- and AR-PAM system was built for data acquisition and network training. The generator network takes an AR-PAM image as input and generates a high-resolution image, while the discriminator network evaluates the similarity between the generator's output and the GT image obtained from OR-PAM. Using *in vivo* mouse ear vascular images, the proposed method was first compared with the blind deconvolution method, and it improved the spatial resolution and produced superior microvasculature images. Furthermore, the proposed method was shown to be applicable to other types of tissues (e.g., brain vessels) and deep tissues (e.g., a chicken breast tissue slice of 1700 μm thickness) that are not easily accessible by OR-PAM. A similar study was implemented by Zhang et al.,[181] who combined a physical model and a learning-based algorithm, termed MultiResU-Net.

DL methods have been applied to address noise and interference issues in PA imaging. Dehner et al.[182] developed a discriminative DNN using a U-Net architecture to separate electrical noise from PA signals, improving PA image contrast and spectral unmixing performance. He et al.[183] proposed an attention-enhanced GAN with a modified U-Net generator to remove noise from PAM images, prioritizing fine-feature restoration. Gulenko et al.[184] evaluated different CNN architectures and found that U-Net demonstrated higher efficiency and accuracy in removing electromagnetic interference noise from PAE systems.

To address image misalignment in PAM, Kim et al.[185] utilized a U-Net framework. Their method effectively addressed nonlinear mismatched cross-sectional B-scan PA images during bidirectional raster scanning, resulting in a significant improvement in imaging speed, doubling the speed compared to conventional approaches.

To improve the temporal resolution of superresolution localization imaging,[186,187] hundreds of thousands of overlapping images are traditionally required. However, this process can be time-consuming. To address this problem, Kim et al.[180] proposed a GAN with U-Net based on pix2pix[188] to reconstruct superresolution images from raw image frames [Fig. 9(b)]. The proposed network can be applied to both 3D label-free localization OR-PAM and 2D labeled localization PACT. The authors trained and validated the network with *in vivo* data from 3D OR-PAM and 2D PACT images. The proposed method reduced the required number of raw frames by 10-fold for OR-PAM and 12-fold for PACT, resulting in a significant improvement in temporal resolution.

Ultraviolet PAM (UV-PAM) takes advantage of the optical absorption contrast of UV light to highlight cell nuclei, generating PA contrast images similar to hematoxylin and eosin (H&E) labeling.[189] DL techniques can be used to digitally generate histological stains using trained NNs based on UV-PAM images, providing label-free alternatives to standard chemical staining methods.[190]

Boktor et al.[191] utilized a DL approach based on GANs to digitally stain total-absorption PA remote sensing (TA-PARS) images, achieving high agreement with the gold standard of histological staining. Cao et al.[192] employed a cycle-consistent adversarial network (CycleGAN)[193] model to virtually stain UV-PAM images, producing pseudo-color PAM images that matched the details of corresponding H&E histology images.

Table 8 Summary of methods for improving the accuracy of image classification and segmentation.

Author	Neural network architecture	Basic network	Training data set (if specified, validation is excluded) — Source	Data amount	Test data set	Specified task	Representative evaluation results
Zhang et al.[160]	AlexNet/GoogLeNet	CNN	k-Wave simulation from in vivo human breast (normal, cancer)	98 (normal)/75 (patient)	k-Wave simulation from in vivo human breast	Classify and segment images	BI-RADS rating accuracy: 83% to 96%
Jnawali et al.[163]	Inception-Resnet-V2	—	Ex vivo human thyroid (normal, benign, cancer)	73	Ex vivo human thyroid	Detect cancer tissue	AUCs for cancer, benign, and normal: 0.73, 0.81, and 0.88
Jnawali et al.[165]	3D CNN	CNN	Thyroid cancer tissue	74 (thyroid)/74 (prostate)	Thyroid and prostate cancer tissue	Detect cancer tissue	AUC, 0.72
Moustakidis et al.[166]	SkinSeg	—	In vivo human	About 26,190 (unclear description)	In vivo human	Identify skin morphology	Per-class accuracy: 84.95%
Nitkunanantharajah et al.[169]	ResNet18	—	In vivo human nailfold (SSc, normal)	990 (from 33 subjects)	In vivo human nailfold	Classify images	Intra-class correlation: 0.902. AUC: 0.897. Sensitivity and specificity: 0.783 and 0.895
Chlis et al.[154]	S-Unet	U-Net	In vivo human vasculature	98 pairs (one pair has 28-wavelength PA data)	In vivo human vasculature	Segment human vasculature	Dice coeff. (versus U-Net): 0.75 → 0.86
Schellenberg et al.[170]	nnU-Net	U-Net/CNN	PA/US images of forearm, calf, and neck from 10 volunteers	144 PA and US image pairs	36 images for validation and 108 images from six volunteers	Segment images	Dice coeff. (versus FCNN): 0.66 → 0.85. Normalized surface distance: 0.61 → 0.89
Lafci et al.[171]	U-Net	U-Net	In vivo mice brain, kidney, and liver	174 images from 12 mice (brain)/97 images from 13 mice (kidney)/108 images from 14 mice (liver)	In vivo mice brain, kidney, and liver	Segment hybrid PA/US image	Dice coeff. 0.95
Boink et al.[172]	L-PD	—	Retinal blood vessels from DRIVE data set	768	Retinal blood vessels from DRIVE dataset/in vitro phantom	Reconstruct and segment images	PSNR (versus FBP): 34 → 42.5
Ly et al.[57]	Modified U-Net	U-Net	In vivo human palm	800	In vivo human palm	Segment blood vessels profile	Global accuracy: 0.9938 (SegNet-5), 0.9920 (FCN-8) → 0.9953. Sensitivity: 0.6406 (SegNet-5), 0.6220 (FCN-8) → 0.8084

AUC, area under the receiver operating characteristic (ROC) curve.
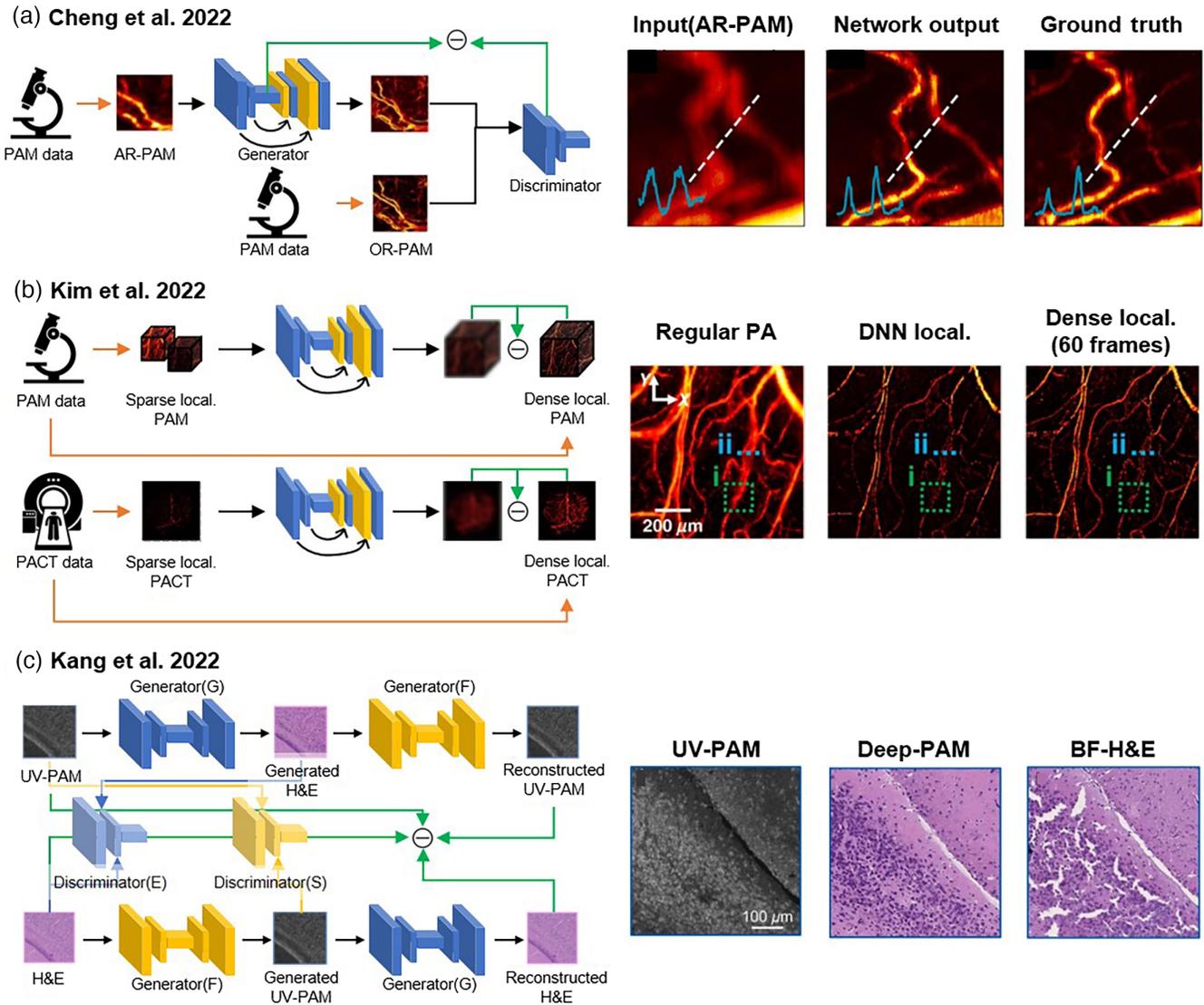
Fig. 9 Representative studies using DL to solve specific issues. (a) GAN-based framework (Wasserstein GAN) to enhance the spatial resolution of AR-PAM. (b) GAN with U-Net to reconstruct superresolution images from raw image frames. (c) Deep-PAM generates virtually stained histological images for both thin sections and thick fresh tissue specimens. The images are adapted with permission from Ref. 178, © 2021 Elsevier GmbH; Ref. 180, © 2022 Springer Nature; and Ref. 58, © 2021 Elsevier GmbH. BF-H&E, brightfield hematoxylin and eosin staining; DNN, deep neural network.

In a recent study, Kang et al.[58] combined UV-PAM with DL to generate rapid and label-free histological images [Fig. 9(c)]. Their proposed method, termed deep-PAM, can generate virtually stained histological images for both thin sections and thick fresh tissue specimens. By utilizing an unpaired image-to-image translation network, a CycleGAN, they were able to process GM-UV-PAM images and instantly produce H&E-equivalent images of unprocessed tissues. This groundbreaking approach has significant implications for the field of histology and may offer an alternative to traditional staining methods.

All the research reviewed in this section is summarized in Table 9.

4 Discussion and Conclusion

PAI is a rapidly growing biomedical imaging modality that utilizes endogenous chromophores to noninvasively provide biometric information, such as vascular structure and sO_2. However, as shown in Fig. 1, PAI still faces seven significant challenges: (1) overcoming limited detection capabilities, (2) compensating for low-dosage light delivery, (3) improving the accuracy of quantitative PA imaging, (4) optimizing or replacing conventional reconstruction methods, (5) addressing tissue heterogeneity, (6) improving the accuracy of image classification and segmentation, and (7) overcoming other specified issues. In this review paper, we have summarized DL studies

Table 9 Summary of methods for addressing other specified issues.

Author	Neural network architecture	Basic network	Training data set (if specified; validation is excluded)			Test data set	Specified task	Representative evaluation results
			Source	Amount				
Chen et al.[175]	CNN	CNN	—	—		Simulation/*in vivo* rat brain	Correct motion artifact	—
Zheng et al.[176]	MAC-Net	GAN	Simulation	7680		Simulation/*in vivo* IVUS and IVOCT	Correct motion artifact	AIFD (versus pre-corrected) 0.1007 → 0.0075
Cheng et al.[178]	WGAN-GP	GAN	*In vivo* mouse ear	528		*In vivo* mouse ear	Transform AR-PAM images into OR-PAM	PSNR (versus blind deconv.) 18.05 → 20.02 SSIM 0.27 → 0.61 PC 0.76 → 0.78
Zhang et al.[181]	AOTDL-GAN/ MultiResU-Net	GAN/U-Net	Simulation/*in vitro* phantom/*in vivo* mouse	3500		Simulation/*in vitro* phantom/*in vivo* mouse	Transform AR-PAM images into OR-PAM	SNR (versus Deconv) 4.853 → 5.70 CNR 6.93 → 12.50 Lateral resolution 45 μm → 15 μm
Dehner et al.[182]	U-Net	U-Net	Simulated PA image/Pure electrical noise/Simulated white Gaussian noise	3000/2110/—		Simulated PA image/ Pure electrical noise/ phantom/*in vivo* human breast/simulated white Gaussian noise	Remove noise	SNR of sinograms (versus pre-corrected) + 10.9 dB
He et al.[183]	GAN	GAN	Leaf phantom/*in vivo* mouse ear	236/149		Leaf phantom/*in vivo* mouse ear	Remove noise	SNR (versus input) 29.08 → 90.73 CNR 4.80 → 7.63
Kim et al.[185]	MS-FD-U-Net	GAN	*In vivo* mouse ear	830		*In vivo* mouse ear	Align bidirectional raster scanning	SSIM (versus input) 0.993 → 0.994 PSNR 50.22 → 50.62 MSE 1.09 → 0.99
Gulenko et al.[184]	Modified U-Net	U-Net	*In vivo* rat colorectum/ *in vivo* rabbit transurethral	700		*In vivo* rat colorectum/*in vivo* rabbit transurethral	Remove noise	Log(RMSE) (versus Segnet) 2.9 → 2.5 Log(SSIM) −2.2 → -0.25 Log(MAE) 2.6 → 2.0
Kim et al.[180]	U-Net	U-Net	*In vivo* OR-PAM mouse ear/*in vivo* PACT mouse brain	3000/500		*In vivo* OR-PAM mouse ear/*in vivo* PACT mouse brain	Accelerate localization process	PSNR (versus input) 38.47 → 40.70 MS-SSIM 0.89 → 0.97
Boktor et al.[191]	Pix2Pix GAN	GAN	Experiments	15,000		Experiments	Perform virtual staining	SSIM between H&E and UV-PAM: 0.91

Table 9 (*Continued*).

Author	Neural network architecture	Basic network	Training data set (if specified; validation is excluded)		Test data set	Specified task	Representative evaluation results
			Source	Amount			
Cao et al.[192]	CycleGAN	GAN	Experiments	17,940 (UV-PAM)/ 26,565 (H&E)	Experiments	Perform virtual staining	H&E versus UV-PAM Cell count: 5549 and 5423. Nuclear area (μm^2): 24.2 and 22.0. Internuclear dist.: 10.14 and 10.18
Kang et al.[58]	CycleGAN	GAN	Experiments	400 (thin section)/ 800 (thick and fresh tissue)	Experiments	Perform virtual staining	H&E versus UV-PAM Cell count: 289 and 283. Nuclear area (μm^2): 70.66 and 72.75

over the past five years that have addressed these general challenges in PAI. Further, we have discussed how DL can be used to solve several more specific problems in PAI.

CNN, U-Net, and GAN have been the most representative networks used in PAI-related research. While some studies use basic architectures to achieve their goals, others modify or develop new architectures to solve particular problems in PAI. These networks can be used in various ways, such as postprocessing reconstructed images with different types of noise or directly reconstructing PA images from the time domain in the image domain.

Furthermore, recent research has aimed to extract more accurate quantitative information by using multiple networks, rather than solely focusing on enhancing image quality with one network. This approach can provide more comprehensive and detailed information, improving the overall performance of PAI. While SSIM is commonly used as the loss function, other metrics, such as PSNR and the Pearson correlation, may be added to improve information extraction and convergence speed. The continued exploration and refinement of these network architectures and loss functions will likely contribute to continued advancements in PAI.

Several obstacles remain. The success of DL approaches in PAI is highly dependent on the availability of high-quality data sets, and there is a scarcity of experimental training data. DL approaches in PAI also lack a standardized PA-image format and publicly available data that are accessible to all groups. Consequently, researchers rely on data generated from experiments or simulations, and even publishing PA data is difficult because there is no standard format. The k-Wave[79] toolbox is the most generally used to generate the PA initial pressure, along with other light transport simulators, such as mcxlab,[194] to generate the light distribution. However, creating reliable simulation data requires GT data from the real world. Commonly, x-ray CT or MRI images of blood vessels and organs are used for PA simulation. Fortunately, there are public data sets of x-ray CT and MRI images, and many groups have used these open data sets to generate PA GTs. However, the varying information obtained by different imaging modalities may not align with PAI. Recently, the International Photoacoustic Standardization Consortium (IPASC)[195] has been working to overcome this challenge by bringing together researchers, device developers, and government regulators to achieve standardization of PAI through community-led consensus building. With the efforts and involvement of IPASC, the generalization ability of DL, which is the fundamental problem in medical imaging field, will increase.

While DL has shown improved image qualities in PAI, there are still concerns regarding its applications in biomedical images. Therefore, the efforts of researchers who aim to advance PAI without applying DL are still valuable. For examples, new restoration algorithms[196] are being developed to enhance the image quality affected by limited-detection capabilities. The development of ultrawide detection bandwidth transducers[197] aims to mitigate the limited bandwidth of traditional US transducers, thereby improving the overall sensitivity and resolution of PAI. Furthermore, specially designed PACT systems with fast-sweep laser scanning techniques offer automatic fluence compensation and motion correction.[129] Combining PACT with transmission-mode US-computed tomography enables the mapping of the distribution of SoS, further enhancing PACT image quality.[198] Moreover, a wide range of exogenous contrast agents has been developed to improve the SNR of PAI or to overcome the resolution limitations.[4]

Despite the challenges faced in applying DL to PAI, there is no doubt that DL will have a great impact on the biomedical imaging field, well beyond PAI.[199–207] PAI's fundamental problems, caused by hardware limitations and the lack of tissue information, are ripe for solution by the information extraction, convergence, and high-speed processing enabled by DL. The result will be new opportunities for PAI to take off as a major imaging modality, opening an exciting era of DL-based PAI.

References

1. L. V. Wang and S. Hu, "Photoacoustic tomography: *in vivo* imaging from organelles to organs," *Science* **335**(6075), 1458–1462 (2012).
2. L. V. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," *Nat. Methods* **13**(8), 627–638 (2016).
3. J. Yang, S. Choi, and C. Kim, "Practical review on photoacoustic computed tomography using curved ultrasound array transducer," *Biomed. Eng. Lett.* **12**(1), 19–35 (2022).
4. W. Choi et al., "Recent advances in contrast-enhanced photoacoustic imaging: overcoming the physical and practical challenges," *Chem. Rev.* **123**, 7379–7419 (2023).
5. W. Choi et al., "Three-dimensional multistructural quantitative photoacoustic and US imaging of human feet *in vivo*," *Radiology* **303**(2), 467–473 (2022).
6. S. Lei et al., "*In vivo* three-dimensional multispectral photoacoustic imaging of dual enzyme-driven cyclic cascade reaction for tumor catalytic therapy," *Nat. Commun.* **13**(1), 1298 (2022).
7. E.-Y. Park et al., "Simultaneous dual-modal multispectral photoacoustic and ultrasound macroscopy for three-dimensional whole-body imaging of small animals," *Photonics* **8**(1), 13 (2021).
8. N. Kwon et al., "Hexa-BODIPY-cyclotriphosphazene based nanoparticle for NIR fluorescence/photoacoustic dual-modal imaging and photothermal cancer therapy," *Biosens. Bioelectron.* **216**, 114612 (2022).
9. C. Kim, C. Favazza, and L. V. Wang, "*In vivo* photoacoustic tomography of chemicals: high-resolution functional and molecular optical imaging at new depths," *Chem. Rev.* **110**(5), 2756–2782 (2010).
10. J. Yang et al., "Assessment of nonalcoholic fatty liver function by photoacoustic imaging," *J. Biomed. Opt.* **28**(1), 016003 (2023).
11. S. Cho et al., "3D PHOVIS: 3D photoacoustic visualization studio," *Photoacoustics* **18**, 100168 (2020).

12. J. Kim et al., "Real-time photoacoustic thermometry combined with clinical ultrasound imaging and high-intensity focused ultrasound," *IEEE Trans. Biomed. Eng.* **66**(12), 3330–3338 (2019).

13. B. Park et al., "Functional photoacoustic imaging: from nano- and micro- to macro-scale," *Nano Converg.* **10**(1), 29 (2023).

14. J. Li et al., "Spatial heterogeneity of oxygenation and haemodynamics in breast cancer resolved *in vivo* by conical multispectral optoacoustic mesoscopy," *Light Sci. Appl.* **9**(1), 57 (2020).

15. J. Yang et al., "Photoacoustic assessment of hemodynamic changes in foot vessels," *J. Biophotonics* **12**(6), e201900004 (2019).

16. J. Yang et al., "Detecting hemodynamic changes in the foot vessels of diabetic patients by photoacoustic tomography," *J. Biophotonics* **13**, e202000011 (2020).

17. J. Yang et al., "Photoacoustic imaging of hemodynamic changes in forearm skeletal muscle during cuff occlusion," *Biomed. Opt. Express* **11**(8), 4560–4570 (2020).

18. M. R. Tomaszewski et al., "Oxygen-enhanced and dynamic contrast-enhanced optoacoustic tomography provide surrogate biomarkers of tumor vascular function, hypoxia, and necrosis," *Cancer Res.* **78**(20), 5980–5991 (2018).

19. V. M. Sciortino et al., "Longitudinal cortex-wide monitoring of cerebral hemodynamics and oxygen metabolism in awake mice using multi-parametric photoacoustic microscopy," *J Cereb. Blood Flow Metab.* **41**(12), 3187–3199 (2021).

20. J. Yang et al., "Intracerebral haemorrhage-induced injury progression assessed by cross-sectional photoacoustic tomography," *Biomed. Opt. Express* **8**(12), 5814–5824 (2017).

21. J. Kim et al., "Multiparametric photoacoustic analysis of human thyroid cancers *in vivo*," *Cancer Res.* **81**(18), 4849–4860 (2021).

22. B. Park et al., "3D wide-field multispectral photoacoustic imaging of human melanomas *in vivo*: a pilot study," *J. Eur. Acad. Dermatol. Venereol.* **35**(3), 669–676 (2021).

23. N. Nikhila and X. Jun, "Photoacoustic imaging of breast cancer: a mini review of system design and image features," *J. Biomed. Opt.* **24**(12), 121911 (2019).

24. B. Park, C. Kim, and J. Kim, "Recent advances in ultrasound and photoacoustic analysis for thyroid cancer diagnosis," *Adv. Phys. Res.* **2**(4), 2200070 (2023).

25. B. Park et al., "Listening to drug delivery and responses via photoacoustic imaging," *Adv. Drug Delivery Rev.* **184**, 114235 (2022).

26. T. Qiu et al., "Assessment of liver function reserve by photoacoustic tomography: a feasibility study," *Biomed. Opt. Express* **11**(7), 3985–3995 (2020).

27. H. Jung et al., "A peptide probe enables photoacoustic-guided imaging and drug delivery to lung tumors in K-rasLA2 mutant mice," *Cancer Res.* **79**(16), 4271–4282 (2019).

28. S. K. Kalva et al., "Rapid volumetric optoacoustic tracking of nanoparticle kinetics across murine organs," *ACS Appl. Mater. Interfaces* **14**(1), 172–178 (2022).

29. H. H. Han et al., "Bimetallic hyaluronate-modified Au@Pt nanoparticles for noninvasive photoacoustic imaging and photothermal therapy of skin cancer," *ACS Appl. Mater. Interfaces* **15**(9), 11609–11620 (2023).

30. T. G. Nguyen Cao et al., "Engineered extracellular vesicle-based sonotheranostics for dual stimuli-sensitive drug release and photoacoustic imaging-guided chemo-sonodynamic cancer therapy," *Theranostics* **12**(3), 1247–1266 (2022).

31. J. Yao and L. V. Wang, "Photoacoustic microscopy," *Laser Photonics Rev.* **7**(5), 758–778 (2013).

32. J. Ahn et al., "Fully integrated photoacoustic microscopy and photoplethysmography of human *in vivo*," *Photoacoustics* **27**, 100374 (2022).

33. J. Park et al., "Quadruple ultrasound, photoacoustic, optical coherence, and fluorescence fusion imaging with a transparent ultrasound transducer," *Proc. Natl. Acad. Sci. U. S. A.* **118**(11), e1920879118 (2021).

34. S.-W. Cho et al., "High-speed photoacoustic microscopy: a review dedicated on light sources," *Photoacoustics* **24**, 100291 (2021).

35. J. W. Baik et al., "Intraoperative label-free photoacoustic histopathology of clinical specimens," *Laser Photonics Rev.* **15**(10), 2100124 (2021).

36. J. Ahn et al., "High-resolution functional photoacoustic monitoring of vascular dynamics in human fingers," *Photoacoustics* **23**, 100282 (2021).

37. J. Ahn et al., "*In vivo* photoacoustic monitoring of vasoconstriction induced by acute hyperglycemia," *Photoacoustics* **30**, 100485 (2023).

38. B. Park et al., "Shear-force photoacoustic microscopy: toward super-resolution near-field imaging," *Laser Photonics Rev.* **16**(12), 2200296 (2022).

39. J. W. Baik et al., "Super wide-field photoacoustic microscopy of animals and humans *in vivo*," *IEEE Trans. Med. Imaging* **39**(4), 975–984 (2020).

40. C. Lee et al., "Three-dimensional clinical handheld photoacoustic/ultrasound scanner," *Photoacoustics* **18**, 100173 (2020).

41. C. Lee et al., "Panoramic volumetric clinical handheld photoacoustic and ultrasound imaging," *Photoacoustics* **31**, 100512 (2023).

42. W. Kim et al., "Wide-field three-dimensional photoacoustic/ultrasound scanner using a two-dimensional matrix transducer array," *Opt. Lett.* **48**(2), 343–346 (2023).

43. W. Choi, D. Oh, and C. Kim, "Practical photoacoustic tomography: realistic limitations and technical solutions," *J. Appl. Phys.* **127**(23), 230903 (2020).

44. S. K. Kalva and M. Pramanik, "Experimental validation of tangential resolution improvement in photoacoustic tomography using modified delay-and-sum reconstruction algorithm," *J. Biomed. Opt.* **21**(8), 086011 (2016).

45. S. Cho et al., "Nonlinear pth root spectral magnitude scaling beamforming for clinical photoacoustic and ultrasound imaging," *Opt. Lett.* **45**(16), 4575–4578 (2020).

46. S. Jeon et al., "Real-time delay-multiply-and-sum beamforming with coherence factor for *in vivo* clinical photoacoustic imaging of humans," *Photoacoustics* **15**, 100136 (2019).

47. M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Phys. Rev. E* **71**(1), 016706 (2005).

48. K. P. Köstli and P. C. Beard, "Two-dimensional photoacoustic imaging by use of Fourier-transform image reconstruction and a detector with an anisotropic response," *Appl. Opt.* **42**(10), 1899–1908 (2003).

49. B. E. Treeby, E. Z. Zhang, and B. T. Cox, "Photoacoustic tomography in absorbing acoustic media using time reversal," *Inverse Prob.* **26**(11), 115003 (2010).

50. I. Steinberg et al., "Superiorized photo-acoustic non-negative reconstruction (spanner) for clinical photoacoustic imaging," *IEEE Trans. Med. Imaging* **40**(7), 1888–1897 (2021).

51. S. Bu et al., "Model-based reconstruction integrated with fluence compensation for photoacoustic tomography," *IEEE Trans. Biomed. Eng.* **59**(5), 1354–1363 (2012).

52. S. Choi et al., "Deep learning enhances multiparametric dynamic volumetric photoacoustic computed tomography *in vivo* (DL-PACT)," *Adv. Sci. (Weinh.)* **10**(1), 2202089 (2023).

53. A. Hariri et al., "Deep learning improves contrast in low-fluence photoacoustic imaging," *Biomed. Opt. Express* **11**(6), 3360–3373 (2020).

54. J. Li et al., "Deep learning-based quantitative optoacoustic tomography of deep tissues in the absence of labeled experimental data," *Optica* **9**(1), 32–41 (2022).

55. T. Tong et al., "Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data," *Photoacoustics* **19**, 100190 (2020).

56. S. Jeon et al., "A deep learning-based model that reduces speed of sound aberrations for improved *in vivo* photoacoustic imaging," *IEEE Trans. Image Process.* **30**, 8773–8784 (2021).

57. C. D. Ly et al., "Full-view *in vivo* skin and blood vessels profile segmentation in photoacoustic imaging based on deep learning," *Photoacoustics* **25**, 100310 (2022).

58. L. Kang et al., "Deep learning enables ultraviolet photoacoustic microscopy based histological imaging with near real-time virtual staining," *Photoacoustics* **25**, 100308 (2022).

59. J. Gröhl et al., "Deep learning for biomedical photoacoustic imaging: a review," *Photoacoustics* **22**, 100241 (2021).

60. X. Zhu et al., "Real-time whole-brain imaging of hemodynamics and oxygenation at micro-vessel resolution with ultrafast wide-field photoacoustic microscopy," *Light Sci. Appl.* **11**(1), 138 (2022).

61. T. C. Benjamin et al., "Quantitative spectroscopic photoacoustic imaging: a review," *J. Biomed. Opt.* **17**(6), 061202 (2012).

62. C. Huang et al., "Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media," *IEEE Trans. Med. Imaging* **32**(6), 1097–1110 (2013).

63. F. Y. Wang et al., "Where does AlphaGo go: from Church-Turing thesis to AlphaGo thesis and beyond," *IEEE/CAA J. Autom. Sin.* **3**(2), 113–120 (2016).

64. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Int. Conf. Mach. Learn.*, pp. 448–456 (2015).

65. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**(6088), 533–536 (1986).

66. Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.* **1**(4), 541–551 (1989).

67. S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747 (2016).

68. C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," *Nat. Methods* **16**(12), 1215–1225 (2019).

69. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).

70. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 3431–3440 (2015).

71. I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM* **63**(11), 139–144 (2020).

72. D. Berthelot, T. Schumm, and L. Metz, "BEGAN: boundary equilibrium generative adversarial networks," arXiv:1703.10717 (2017).

73. A. Hauptmann and B. Cox, "Deep learning in photoacoustic tomography: current approaches and future directions," *J. Biomed. Opt.* **25**(11), 112903 (2020).

74. H. Deng et al., "Deep learning in photoacoustic imaging: a review," *J. Biomed. Opt.* **26**(4), 040901 (2021).

75. G. Wissmeyer et al., "Looking at sound: optoacoustics with all-optical ultrasound detection," *Light Sci. Appl.* **7**(1), 53 (2018).

76. G. Sreedevi et al., "Deep neural network-based bandwidth enhancement of photoacoustic data," *J. Biomed. Opt.* **22**(11), 116001 (2017).

77. T. Lu et al., "LV-GAN: a deep learning approach for limited-view optoacoustic imaging based on hybrid datasets," *J. Biophotonics* **14**(2), e202000325 (2021).

78. H. Lan et al., "Y-Net: hybrid deep learning image reconstruction for photoacoustic tomography *in vivo*," *Photoacoustics* **20**, 100197 (2020).

79. B. E. Treeby, J. Jaros, and B. T. Cox, "Advanced photoacoustic image reconstruction using the k-Wave toolbox," *Proc. SPIE* **9708**, 97082P (2016).

80. S. Ma, S. Yang, and H. Guo, "Limited-view photoacoustic imaging based on linear-array detection and filtered mean-backprojection-iterative reconstruction," *J. Appl. Phys.* **106**(12), 123104 (2009).

81. Y. Tang et al., "High-fidelity deep functional photoacoustic tomography enhanced by virtual point sources," *Photoacoustics* **29**, 100450 (2023).

82. S. Vilov et al., "Photoacoustic fluctuation imaging: theory and application to blood flow imaging," *Optica* **7**(11), 1495–1505 (2020).

83. H. Deng et al., "Machine-learning enhanced photoacoustic computed tomography in a limited view configuration," *Proc. SPIE* **11186**, 111860J (2019).

84. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).

85. J. Zhang et al., "Limited-view photoacoustic imaging reconstruction with dual domain inputs based on mutual information," in *IEEE 18th Int. Symp. Biomed. Imaging (ISBI)*, pp. 1522–1526 (2021).

86. J. Staal et al., "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging* **23**(4), 501–509 (2004).

87. Y. Xu, D. Feng, and L. V. Wang, "Exact frequency-domain reconstruction for thermoacoustic tomography. I. Planar geometry," *IEEE Trans. Med. Imaging* **21**(7), 823–828 (2002).

88. L. Li and L. V. Wang, "Recent advances in photoacoustic tomography," *BME Front.* **2021**, 9823268 (2021).

89. S. Guan et al., "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE J. Biomed. Health. Inf.* **24**(2), 568–576 (2019).

90. P. Farnia et al., "High-quality photoacoustic image reconstruction based on deep convolutional neural network: towards intraoperative photoacoustic imaging," *Biomed. Phys. Eng. Express* **6**(4), 045019 (2020).

91. M. Guo et al., "AS-Net: fast photoacoustic reconstruction with multi-feature fusion from sparse data," *IEEE Trans. Comput. Imaging* **8**, 215–223 (2022).

92. H. Lan et al., "Ki-GAN: knowledge infusion generative adversarial network for photoacoustic image reconstruction *in vivo*," *Lect. Notes Comput. Sci.* **11764**, 273–281 (2019).

93. A. DiSpirito et al., "Reconstructing undersampled photoacoustic microscopy images using deep learning," *IEEE Trans. Med. Imaging* **40**(2), 562–570 (2020).

94. M. Chen et al., "Simultaneous photoacoustic imaging of intravascular and tissue oxygenation," *Opt. Lett.* **44**(15), 3773–3776 (2019).

95. T. Vu et al., "Deep image prior for undersampling high-speed photoacoustic microscopy," *Photoacoustics* **22**, 100266 (2021).

96. G. Godefroy, B. Arnal, and E. Bossy, "Compensating for visibility artefacts in photoacoustic imaging with a deep learning approach providing prediction uncertainties," *Photoacoustics* **21**, 100218 (2021).

97. Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: representing model uncertainty in deep learning," in *Int. Conf. Mach. Learn.*, pp. 1050–1059 (2016).

98. T. Vu et al., "A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer," *Exp. Biol. Med. (Maywood)* **245**(7), 597–605 (2020).

99. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4681–4690 (2017).

100. H. Zhang et al., "A new deep learning network for mitigating limited-view and under-sampling artifacts in ring-shaped photoacoustic tomography," *Comput. Med. Imaging Graph.* **84**, 101720 (2020).

101. N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," *Nat. Mach. Intell.* **1**(10), 453–460 (2019).

102. N. Davoudi et al., "Deep learning of image-and time-domain data enhances the visibility of structures in optoacoustic tomography," *Opt. Lett.* **46**(13), 3029–3032 (2021).

103. N. Awasthi et al., "Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography," *IEEE Trans. Ultrasonics, Ferroelectr. Freq. Control* **67**(12), 2660–2673 (2020).

104. D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUS)," arXiv:1511.07289 (2015).

105. J. Schwab et al., "Real-time photoacoustic projection imaging using deep learning," arXiv:1801.06693 (2018).

106. T. Karras et al., "Progressive growing of GANS for improved quality, stability, and variation," arXiv:1710.10196 (2017).

107. K. Daoudi et al., "Handheld probe integrating laser diode and ultrasound transducer array for ultrasound/photoacoustic dual modality imaging," *Opt. Express* **22**(21), 26365–26374 (2014).

108. A. Hariri et al., "The characterization of an economic and portable LED-based photoacoustic imaging system to facilitate molecular imaging," *Photoacoustics* **9**, 10–20 (2018).

109. P. Rajendran and M. Pramanik, "High frame rate (∼3 Hz) circular photoacoustic tomography using single-element ultrasound transducer aided with deep learning," *J. Biomed. Opt.* **27**(6), 066005 (2022).

110. H. Zhao et al., "Deep learning enables superior photoacoustic imaging at ultralow laser dosages," *Adv. Sci. (Weinh.)* **8**(3), 2003097 (2021).

111. M. K. A. Singh et al., "Deep learning-enhanced LED-based photoacoustic imaging," *Proc. SPIE* **11240**, 1124038 (2020).

112. E. M. A. Anas et al., "Towards a fast and safe LED-based photoacoustic imaging using deep convolutional neural network," *Lect. Notes Comput. Sci.* **11073**, 159–167 (2018).

113. L. R. Medsker and L. Jain, *Recurrent Neural Networks*, CRC Press, Inc. (2001).

114. E. M. A. Anas et al., "Enabling fast and high quality LED photoacoustic imaging: a recurrent neural networks based approach," *Biomed. Opt. Express* **9**(8), 3852–3866 (2018).

115. M. Li, Y. Tang, and J. Yao, "Photoacoustic tomography of blood oxygenation: a mini review," *Photoacoustics* **10**, 65–73 (2018).

116. S. Tzoumas et al., "Eigenspectra optoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues," *Nat. Commun.* **7**(1), 1–10 (2016).

117. A. Rosenthal, D. Razansky, and V. Ntziachristos, "Fast semi-analytical model-based acoustic inversion for quantitative optoacoustic tomography," *IEEE Trans. Med. Imaging* **29**(6), 1275–1285 (2010).

118. X. L. Deán-Ben and D. Razansky, "A practical guide for model-based reconstruction in optoacoustic imaging," *Front. Phys.* **10**, 1057 (2022).

119. C. Cai et al., "End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging," *Opt. Lett.* **43**(12), 2752–2755 (2018).

120. C. Yang et al., "Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network," in *IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019)*, pp. 741–744 (2019).

121. G. P. Luke et al., "O-Net: a convolutional neural network for quantitative photoacoustic image segmentation and oximetry," arXiv:1911.01935 (2019).

122. C. Yang and F. Gao, "EDA-Net: dense aggregation of deep and shallow information achieves quantitative photoacoustic blood oxygenation imaging deep in human breast," *Lect. Notes Comput. Sci.* **11764**, 246–254 (2019).

123. J. Gröhl et al., "Estimation of blood oxygenation with learned spectral decoloring for quantitative photoacoustic imaging (LSD-qPAI)," arXiv:1902.05839 (2019).

124. C. Bench, A. Hauptmann, and B. Cox, "Toward accurate quantitative photoacoustic imaging: learning vascular blood oxygen saturation in three dimensions," *J. Biomed. Opt.* **25**(8), 085003 (2020).

125. Y. Zou et al., "Ultrasound-enhanced Unet model for quantitative photoacoustic tomography of ovarian lesions," *Photoacoustics* **28**, 100420 (2022).

126. T. Chen et al., "A deep learning method based on U-Net for quantitative photoacoustic imaging," *Proc. SPIE* **11240**, 112403V (2020).

127. J. Gröhl et al., "Confidence estimation for machine learning-based quantitative photoacoustics," *J. Imaging* **4**(12), 147 (2018).

128. Y. Wang et al., "Nonlinear iterative perturbation scheme with simplified spherical harmonics (SP3) light propagation model for quantitative photoacoustic tomography," *J. Biophotonics* **14**(6), e202000446 (2021).

129. G.-S. Jeng et al., "Real-time interleaved spectroscopic photoacoustic and ultrasound (PAUS) scanning with simultaneous fluence compensation and motion correction," *Nat. Commun.* **12**(1), 716 (2021).

130. S. Park et al., "Normalization of optical fluence distribution for three-dimensional functional optoacoustic tomography of the breast," *J. Biomed. Opt.* **27**(3), 036001 (2022).

131. J. Zhu et al., "Self-fluence-compensated functional photoacoustic microscopy," *IEEE Trans. Med. Imaging* **40**(12), 3856–3866 (2021).

132. A. Madasamy et al., "Deep learning methods hold promise for light fluence compensation in three-dimensional optoacoustic imaging," *J. Biomed. Opt.* **27**(10), 106004 (2022).

133. Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.* **15**(5), 749–753 (2018).

134. A. Creswell et al., "Generative adversarial networks: an overview," *IEEE Signal Process Mag.* **35**(1), 53–65 (2018).

135. D. A. Durairaj et al., "Unsupervised deep learning approach for photoacoustic spectral unmixing," *Proc. SPIE* **11240**, 112403H (2020).

136. I. Olefir et al., "Deep learning-based spectral unmixing for optoacoustic imaging of tissue oxygen saturation," *IEEE Trans. Med. Imaging* **39**(11), 3643–3654 (2020).

137. S. Guan et al., "Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning," *Sci. Rep.* **10**(1), 8510 (2020).

138. J. Feng et al., "End-to-end Res-Unet based reconstruction algorithm for photoacoustic imaging," *Biomed. Opt. Express* **11**(9), 5321–5340 (2020).

139. W. Dominik et al., "Reconstruction of initial pressure from limited view photoacoustic images using deep learning," *Proc. SPIE* **10494**, 104942S (2018).

140. A. Stephan et al., "Photoacoustic image reconstruction via deep learning," *Proc. SPIE* **10494**, 104944U (2018).

141. H. Lan et al., "Reconstruct the photoacoustic image based on deep learning with multi-frequency ring-shape transducer array," in *41st Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc. (EMBC)*, pp. 7115–7118 (2019).

142. C. Yang, H. Lan, and F. Gao, "Accelerated photoacoustic tomography reconstruction via recurrent inference machines," in *41st Annu. Int. Conf. IEEE Eng. in Med. and Biol. Soc. (EMBC)*, pp. 6371–6374 (2019).

143. M. Kim et al., "Deep-learning image reconstruction for real-time photoacoustic system," *IEEE Trans. Med. Imaging* **39**(11), 3379–3390 (2020).

144. A. Hauptmann et al., "Model-based learning for accelerated, limited-view 3-D photoacoustic tomography," *IEEE Trans. Med. Imaging* **37**(6), 1382–1393 (2018).

145. A. Hauptmann et al., "Approximate k-space models and deep learning for fast photoacoustic reconstruction," *Lect. Notes Comput. Sci.* **11074**, 103–111 (2018).

146. M. K. A. Singh and W. Steenbergen, "Photoacoustic-guided focused ultrasound (PAFUSion) for identifying reflection artifacts in photoacoustic imaging," *Photoacoustics* **3**(4), 123–131 (2015).

147. R. Austin and A. L. B. Muyinatu, "A machine learning approach to identifying point source locations in photoacoustic data," *Proc. SPIE* **10064**, 100643J (2017).

148. D. Allman, A. Reiter, and M. A. L. Bell, "A machine learning method to identify and remove reflection artifacts in photoacoustic channel data," in *IEEE Int. Ultrasonics Symp. (IUS)*, pp. 1–4 (2017).

149. S. Ren et al., "Faster R-CNN: towards real-time object detection with region proposal networks," in *Adv. Neural Inf. Process. Syst.,* **28** (2015).

150. H. Shan, G. Wang, and Y. Yang, "Accelerated correction of reflection artifacts by deep neural networks in photo-acoustic tomography," *Appl. Sci.* **9**(13), 2615 (2019).

151. P. Stefanov and Y. Yang, "Multiwave tomography in a closed domain: averaged sharp time reversal," *Inverse Prob.* **31**(6), 065007 (2015).

152. Z. Belhachmi, T. Glatz, and O. Scherzer, "A direct method for photoacoustic tomography with inhomogeneous sound speed," *Inverse Prob.* **32**(4), 045005 (2016).

153. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).

154. N.-K. Chlis et al., "A sparse deep learning approach for automatic segmentation of human vasculature in multispectral optoacoustic tomography," *Photoacoustics* **20**, 100203 (2020).

155. X. Lin et al., "Variable speed of sound compensation in the linear-array photoacoustic tomography using a multi-stencils fast marching method," *Biomed. Signal Process. Control* **44**, 67–74 (2018).

156. B. Treeby et al., "Automatic sound speed selection in photoacoustic image reconstruction using an autofocus approach," *J. Biomed. Opt.* **16**(9), 090501 (2011).

157. D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Trans. Med. Imaging* **37**(6), 1464–1477 (2018).

158. E. Moen et al., "Deep learning for cellular image analysis," *Nat. Methods* **16**(12), 1233–1246 (2019).

159. S. Misra et al., "Deep learning-based multimodal fusion network for segmentation and classification of breast cancers using B-mode and elastography ultrasound images," *Bioeng. Transl. Med.* e10480 (2022).

160. J. Zhang et al., "Photoacoustic image classification and segmentation of breast cancer: a feasibility study," *IEEE Access* **7**, 5457–5466 (2019).

161. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM* **60**(6), 84–90 (2017).

162. C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–9 (2015).

163. K. Jnawali et al., "Transfer learning for automatic cancer tissue detection using multispectral photoacoustic imaging," *Proc. SPIE* **10950**, 109503W (2019).

164. C. Szegedy et al., "Inception-v4, inception-ResNet and the impact of residual connections on learning," *Proc. AAAI Conf. Artif. Intell.* **31**(1), 4278–4284 (2017).

165. K. Jnawali et al., "Deep 3D convolutional neural network for automatic cancer tissue detection using multispectral photoacoustic imaging," *Proc. SPIE* **10955**, 109551D (2019).

166. S. Moustakidis et al., "Fully automated identification of skin morphology in raster-scan optoacoustic mesoscopy using artificial intelligence," *Med. Phys.* **46**(9), 4046–4056 (2019).

167. W. A. Belson, "Matching and prediction on the principle of biological classification," *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* **8**(2), 65–75 (1959).

168. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).

169. S. Nitkunanantharajah et al., "Three-dimensional optoacoustic imaging of nailfold capillaries in systemic sclerosis and its potential for disease differentiation using deep learning," *Sci. Rep.* **10**(1), 16444 (2020).

170. M. Schellenberg et al., "Semantic segmentation of multispectral photoacoustic images using deep learning," *Photoacoustics* **26**, 100341 (2022).

171. B. Lafci et al., "Deep learning for automatic segmentation of hybrid optoacoustic ultrasound (OPUS) images," *IEEE Trans. Ultrasonics, Ferroelectr. Freq. Control* **68**(3), 688–696 (2021).

172. Y. E. Boink, S. Manohar, and C. Brune, "A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation," *IEEE Trans. Med. Imaging* **39**(1), 129–139 (2020).

173. M. Schwarz et al., "Motion correction in optoacoustic mesoscopy," *Sci. Rep.* **7**(1), 10386 (2017).

174. X. Tong et al., "Non-invasive 3D photoacoustic tomography of angiographic anatomy and hemodynamics of fatty livers in rats," *Adv. Sci. (Weinh.)* **10**(2), 2205759 (2023).

175. X. Chen, W. Qi, and L. Xi, "Deep-learning-based motion-correction algorithm in optical resolution photoacoustic microscopy," *Vis. Comput. Ind. Biomed. Art* **2**(1), 12 (2019).

176. S. Zheng et al., "A deep learning method for motion artifact correction in intravascular photoacoustic image sequence," *IEEE Trans. Med. Imaging* **42**(1), 66–78 (2023).

177. M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Adv. Neural Inf. Process. Syst.,* **28** (2015).

178. S. Cheng et al., "High-resolution photoacoustic microscopy with deep penetration through learning," *Photoacoustics* **25**, 100314 (2022).

179. I. Gulrajani et al., "Improved training of Wasserstein GANS," in *Adv. Neural Inf. Process. Syst.,* 30 (2017).

180. J. Kim et al., "Deep learning acceleration of multiscale superresolution localization photoacoustic imaging," *Light Sci. Appl.* **11**(1), 131 (2022).

181. Z. Zhang et al., "Deep and domain transfer learning aided photoacoustic microscopy: acoustic resolution to optical resolution," *IEEE Trans. Med. Imaging* **41**(12), 3636–3648 (2022).

182. C. Dehner et al., "Deep-learning-based electrical noise removal enables high spectral optoacoustic contrast in deep tissue," *IEEE Trans. Med. Imaging* **41**(11), 3182–3193 (2022).

183. D. He et al., "De-noising of photoacoustic microscopy images by attentive generative adversarial network," *IEEE Trans. Med. Imaging* **42**, 1349–1362 (2022).

184. O. Gulenko et al., "Deep-learning-based algorithm for the removal of electromagnetic interference noise in photoacoustic endoscopic image processing," *Sensors* **22**(10), 3961 (2022).

185. J. Kim et al., "Deep learning alignment of bidirectional raster scanning in high speed photoacoustic microscopy," *Sci. Rep.* **12**(1), 16238 (2022).

186. J. Kim et al., "Super-resolution localization photoacoustic microscopy using intrinsic red blood cells as contrast absorbers," *Light Sci. Appl.* **8**(1), 103 (2019).

187. W. Choi and C. Kim, "Toward *in vivo* translation of super-resolution localization photoacoustic computed tomography using liquid-state dyed droplets," *Light Sci. Appl.* **8**(1), 57 (2019).

188. P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1125–1134 (2017).

189. T. T. W. Wong et al., "Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy," *Sci. Adv.* **3**(5), e1602168 (2017).

190. B. Bai et al., "Deep learning-enabled virtual histological staining of biological samples," *Light Sci. Appl.* **12**(1), 57 (2023).

191. M. Boktor et al., "Virtual histological staining of label-free total absorption photoacoustic remote sensing (TA-PARS)," *Sci. Rep.* **12**(1), 10296 (2022).

192. R. Cao et al., "Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy," *Nat. Biomed. Eng.* **7**(2), 124–134 (2023).

193. J.-Y. Zhu et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2223–2232 (2017).

194. L. Yu et al., "Scalable and massively parallel Monte Carlo photon transport simulations for heterogeneous computing platforms," *J. Biomed. Opt.* **23**(1), 010504 (2018).

195. S. Bohndiek, "Addressing photoacoustics standards," *Nat. Photonics* **13**(5), 298–298 (2019).

196. L. Qi et al., "Photoacoustic tomography image restoration with measured spatially variant point spread functions," *IEEE Trans. Med. Imaging* **40**(9), 2318–2328 (2021).

197. R. Shnaiderman et al., "A submicrometre silicon-on-insulator resonator for ultrasound detection," *Nature* **585**(7825), 372–378 (2020).

198. E. Merčep et al., "Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals," *Light Sci. Appl.* **8**(1), 18 (2019).

199. G. Kim et al., "Integrated deep learning framework for accelerated optical coherence tomography angiography," *Sci. Rep.* **12**(1), 1289 (2022).

200. S. Misra et al., "Bi-modal transfer learning for classifying breast cancers via combined B-mode and ultrasound strain imaging," *IEEE Trans. Ultrasonics, Ferroelectr. Freq. Control* **69**(1), 222–232 (2022).

201. S. Misra et al., "Multi-channel transfer learning of chest x-ray images for screening of COVID-19," *Electronics* **9**(9), 1388 (2020).

202. C. Yoon et al., "Collaborative multi-modal deep learning and radiomic features for classification of strokes within 6 h," *Expert Syst. Appl.* **228**, 120473 (2023).

203. S. Misra et al., "A voting-based ensemble feature network for semiconductor wafer defect classification," *Sci. Rep.* **12**(1), 16254 (2022).

204. S. Kim et al., "Convolutional neural network–based metal and streak artifacts reduction in dental CT images with sparse-view sampling scheme," *Med. Phys.* **49**(9), 6253–6277 (2022).

205. S. Choi et al., "*In situ* x-ray-induced acoustic computed tomography with a contrast agent: a proof of concept," *Opt. Lett.* **47**(1), 90–93 (2022).

206. S. Choi et al., "Synchrotron x-ray induced acoustic imaging," *Sci. Rep.* **11**(1), 4047 (2021).

207. H. Kim et al., "Deep learning-based histopathological segmentation for whole slide images of colorectal cancer in a compressed domain," *Sci. Rep.* **11**(1), 22520 (2021).

Biographies of the authors are not available.