Minireview

Sounding out the hidden data: A concise review of deep learning in photoacoustic imaging

Anthony DiSpirito III[1], Tri Vu[1], Manojit Pramanik[2] and Junjie Yao[1]

[1]Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA; [2]School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore 637459, Singapore
Corresponding author: Junjie Yao. Email: junjie.yao@duke.edu

Impact statement

With the rapidly developing integration of deep learning in photoacoustic tomography (PAT) over recent years, comes the pressing need to succinctly summarize previous work and present advances. This concise review seeks to properly orient current researchers who are new to either deep learning or PAT, and serves as a condensed exhibition meant to share the exciting innovations emerging from the intersection of PAT and deep learning with the broader research community. This review seeks to shed light on the applications of artificial intelligence in PAT, aiming to capture the attention of interested researchers and spawn the next wave of future innovation within the field.

Abstract

The rapidly evolving field of photoacoustic tomography utilizes endogenous chromophores to extract both functional and structural information from deep within tissues. It is this power to perform precise quantitative measurements *in vivo*—with endogenous or exogenous contrast—that makes photoacoustic tomography highly promising for clinical translation in functional brain imaging, early cancer detection, real-time surgical guidance, and the visualization of dynamic drug responses. Considering photoacoustic tomography has benefited from numerous engineering innovations, it is of no surprise that many of photoacoustic tomography's current cutting-edge developments incorporate advances from the equally novel field of artificial intelligence. More specifically, alongside the growth and prevalence of graphical processing unit capabilities within recent years has emerged an offshoot of artificial intelligence known as deep learning. Rooted in the solid foundation of signal processing, deep learning typically utilizes a method of optimization known as gradient descent to minimize a loss function and update model parameters. There are already a number of innovative efforts in photoacoustic tomography utilizing deep learning techniques for a variety of purposes, including resolution enhancement, reconstruction artifact removal, undersampling correction, and improved quantification. Most of these efforts have proven to be highly promising in addressing long-standing technical obstacles where traditional solutions either completely fail or make only incremental progress. This concise review focuses on the history of applied artificial intelligence in photoacoustic tomography, presents recent advances at this multifaceted intersection of fields, and outlines the most exciting advances that will likely propagate into promising future innovations.

Keywords: Photoacoustic tomography, deep learning, convolutional neural networks, artificial intelligence, photoacoustic computed tomography, photoacoustic microscopy

Experimental Biology and Medicine **2021; 246: 1355–1367. DOI: 10.1177/15353702211000310**

Introduction

The hybrid imaging modality of photoacoustic tomography (PAT) combines optical excitation and ultrasound detection to achieve an unparalleled balance of spatial resolution, penetration depth, and imaging speed.[1–3] PAT relies on the photoacoustic effect, by which the absorption of excitation light by endogenous or exogenous chromophores causes a transient temperature rise that generates a pressure rise proportional to the optical absorption.[1,4] This rapid pressure rise propagates through the tissue as ultrasound waves that are detected by an external ultrasound transducer or transducer array. PAT has two major implementations: photoacoustic computed tomography (PACT) using wide-field light illumination and parallel acoustic detection, and photoacoustic microscopy (PAM) using focused light illumination and point-by-point acoustic detection. Both PAT implementations introduce their own unique set of challenges, which were often restricted to hardware solutions in the past, like more expensive and complex transducer arrays in PACT or novel and equally costly scanning mechanisms in PAM. However, with the advent of traditional iterative reconstruction methods and

ISSN 1535-3702

dictionary learning at first, and later deep learning techniques, there now exists promising new pure-software solutions to address these persistent technical challenges. Many of these software solutions rest on the same fundamental premise—there exists a lack of certain key pieces of information due to imperfect measurement methods. This incomplete data can be approximated via optimization methods that take advantage of either established mathematical models of the imaging progress, some overarching property of the targets like smoothness (i.e., total variation minimization), or features extracted from simulation data, phantom studies, or *in vivo* data during the process of model training.

This review begins with a brief summary of PAT—the fundamentals and current challenges, and then proceeds onto a survey of deep learning principles, with an emphasis on how deep learning is uniquely suited to address the obstacles in PAT. This review extensively explores the history of deep learning optimization methods in PAT, putting current advances in context as part of a continuous line of progress that emanates from the past and propagates forward onto a future of further innovative research. The review then concludes with an in-depth discussion of the significance of the most promising current advances and a look toward the future.

Fundamentals of photoacoustic imaging

PAT relies on a physical phenomenon known as the photoacoustic effect. First reported by Alexander Graham Bell in 1880, the photoacoustic effect refers the physical phenomenon by which light is absorbed by a material and converted into acoustic energy (see Figure 1(A)).[6,7] This conversion occurs when the optical absorption causes a rise in temperature, which causes a rise in pressure through thermo-elastic expansion, which then propagates through the tissue as ultrasound waves—called the photoacoustic wave.[1,4] The most important advantage of PAT is thus its ability to combine optical excitation, and therefore optical absorption contrast, with the spatial resolution of ultrasound for imaging deep within optically scattering tissues. Two key timescales must be met in PAT optical excitation in order to maximize the initial pressure wave: the thermal relaxation time (τ_{th}) and the stress relaxation time (τ_s).[8] In short, the thermal relaxation time refers to the time it takes the thermal energy of an excitation pulse to propagate out of the heated region. When the excitation pulsewidth is much shorter than the thermal relaxation time, thermal conduction during the laser pulse excitation is considered negligible and the excitation is in thermal confinement. Similarly, the stress relaxation time refers to the time required for stress (i.e., pressure) to propagate out of the heated volume. When the laser pulse duration is less than the stress relaxation time (i.e., stress confinement) and thermal confinement has been met, the fractional volume expansion $\frac{\Delta V}{V}$ can be considered negligible[9] (as shown in equation (1))

$$\frac{\Delta V}{V} = -\kappa \Delta p + \beta \Delta T = 0 \qquad (1)$$

where Δp and ΔT represent changes in pressure and temperature respectively; κ represents isothermal compressibility, and β denotes the thermal coefficient of volume expansion. This state of thermal and stress confinement allows the fractional volume expansion to be considered negligible and thus the initial local pressure rise (p_0) immediately after the laser excitation pulse can be computed as (equation (2))

$$p_0 = \frac{\beta}{\kappa \rho C_v} \eta_{th} \mu_a F = \Gamma \eta_{th} \mu_a F = \Gamma E_a \qquad (2)$$

where ρ denotes the mass density, C_v is the specific heat capacity at a constant volume, η_{th} is the percentage of specific optical absorption converted to heat, μ_a is the optical absorption coefficient, F is the optical fluence, and E_a is the absorbed optical energy. Therefore, the initial local pressure rise is directly proportional to the non-radiative optical energy absorption via the proportionality factor Γ, known as the Grüneisen coefficient. With 100% sensitivity to the optical absorption contrast, PAT is fundamentally an optical imaging modality.

Major photoacoustic imaging implementations

So far, PAT has developed two primary implementations based on the image formation methodologies, known as photoacoustic computed tomography (PACT) and photoacoustic microscopy (PAM). In this section, we briefly discuss the fundamentals of PAM and PACT, and highlight how the nature of these different techniques introduces unique challenges to be overcome by hardware or software interventions. PACT usually employs diffused light illumination and parallel acoustic detection by an ultrasound transducer array (Figure 1(B)).[8] The ultrasound transducer array captures the photoacoustic waves at different projection angles, typically through tomographic scanning (i.e., a linear detection geometry) or volumetric imaging (i.e., a spherical detection geometry). The ultrasound signals from different projection angles can then be assembled and backprojected using various reconstruction techniques to estimate the initial pressure distribution.[4] This initial pressure distribution is approximately proportional to the optical energy deposition within the tissue. PACT can achieve deep imaging depths of several centimeters, far beyond the optical diffusion limit in soft tissue (~1 mm), benefiting from the diffusive optical illumination and relatively low-frequency ultrasound detection. The spatial resolution of PACT is determined primarily by the ultrasound detection and not the optical excitation.[1,9]

PAM differs from PACT in its image formation methodology. PAM typically utilizes a focused single-element ultrasound transducer to form images through point-by-point scanning (see Figure 1(B)). The scanning method can vary significantly among PAM implementations.[8] For example, some traditional PAM systems utilize slow mechanical raster scanning, while modern PAM systems utilize high-speed optical scanning mirrors.[1,5,10–14] Although all PAM systems utilize focused ultrasound detection, some systems only use weakly focused optical

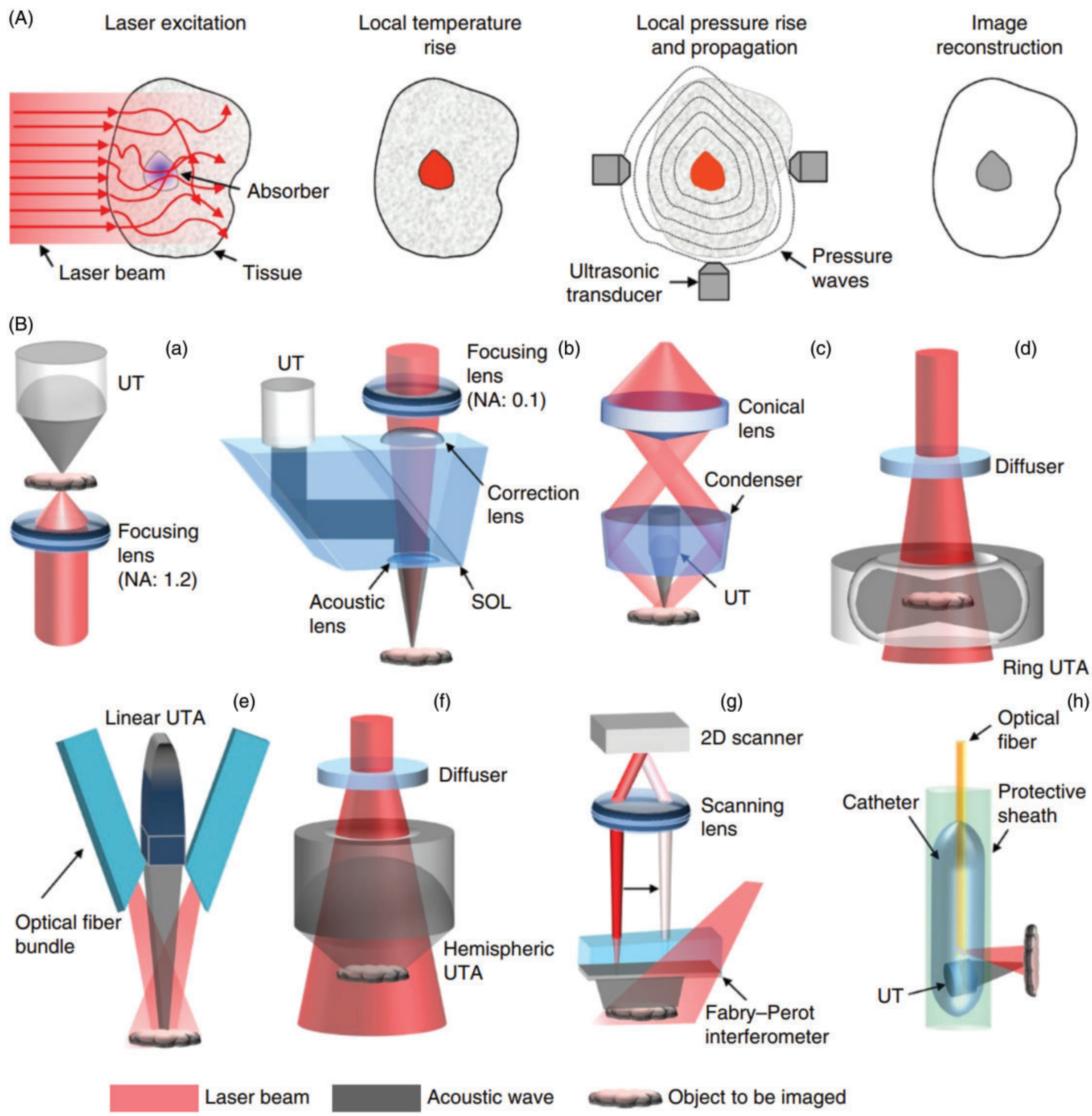

Figure 1. Principle of photoacoustic tomography (PAT). (A) The imaging process of PAT. (B) The major implementations of PAT, including (a) transmission-mode OR-PAM system, (b) reflection-mode OR-PAM system, (c) AR-PAM system with a dark-field illumination, (d) PACT system with a ring-shaped ultrasound transducer array (UTA), (e) PACT system with a linear UTA, (f) PACT system with a hemispherically-shaped UTA, (g) PACT system with a 2D Fabry-Pérot interferometer as the acoustic sensor, and (h) side-viewing intravascular PA catheter. Adapted with permission from Wang and Yao.[5] (A color version of this figure is available in the online journal.)

excitation—known as acoustic resolution PAM (AR-PAM)—while other systems use tightly focused optical excitation—known as optical resolution PAM (OR-PAM).[8] In other words, OR-PAM and AR-PAM are different in terms of which imaging process is more focused, the optical excitation (OR-PAM), or the ultrasound detection (AR-PAM). As such, OR-PAM can provide optical-diffraction-limited resolutions within the (quasi) ballistic regime (<1 mm), while AR-PAM can provide acoustic-diffraction-limited resolutions within the quasi-diffusive regime (<10 mm).[5]

Technical challenges in PAT

In PACT, many challenges arise from solving the PA inverse problem, which is typically ill-posed—mostly due to partial and/or sparse detection geometries.[15] In sparse-sampling PACT, fewer ultrasound transducers are used to reduce the system cost and complexity, at the price of spatial sampling density and projection angles.[16,17] The lack of adequate projection angles results in reconstruction artifacts, reduced image contrast, and diminished quantification accuracy. Similarly, PACT may also suffer from the limited-view problem with low visibility of certain target structures.[18]

Additional reconstruction artifacts can also come from inadequate frequency sampling due to limited-bandwidth of the ultrasound transducers.[19] In PACT, the above challenges are often present simultaneously and it is difficult to separately address their individual impacts on the final imaging performance.

One of the most pressing challenges in PAM is the slow speed resulting from the point-by-point scanning. The point-by-point scanning in PAM often results in long image acquisition times in order to cover a large field of view with a fine spatial resolution.[20] Consequently, this has led to spatial undersampling in traditional PAM systems in order to reduce the image acquisition time, and has inspired the development of fast-scanning systems with or without undersampling.[1] Traditionally, in order to maintain high image quality, interpolation methods were used to upsample the downsampled PAM images. However, as PAM undersampling is not a blurring procedure, but rather a process of skipping effective pixels, the interpolation procedure can result in severe aliasing artifacts and image blurring.[20] In addition, the slow scanning in PAM also results in motion artifacts for dynamic imaging.[21] Fast scanning or undersampling can improve the imaging speed, but often at the cost of inferior image quality and resolution. Besides the imaging speed, PAM systems also suffer from quickly deteriorating spatial resolutions outside of the optical and/or acoustic focal zone.[22]

While numerous engineering efforts have been spent to address the above technical challenges in PAT, many of them rely on complex and expensive hardware such as powerful light sources, 2D ultrasound arrays, and high-speed scanning mirrors, as well as time-consuming image reconstructions and processing, such as iterative-based methods. Moreover, these engineering solutions often must make trade-offs between different imaging parameters such as the imaging speed versus the field of view, and the spatial resolution versus the penetration depth. There exists an acute need in PAT for innovative solutions that can approach these challenges from a completely different perspective, without the need for upgrading the imaging system's hardware or software. As is the case in many other scientific disciplines, deep learning methods have emerged as a viable path to efficiently address many of PAT's long-standing technical challenges.

Deep learning in PAT

Brief introduction to deep learning

Deep learning (DL) developed out of computer science, originally taking the form of simple neural networks, like the perceptron.[23] Inspired by the biological structure of neurons, these networks used nodes connected by edge weights and a nonlinear activation function (Figure 2 (A)).[24] Eventually, researchers devised schemes like the multilayered perceptron that used layered nodes with "hidden layers," which are layers not directly observed by the inputs and outputs (a "black box"). The layer weights were optimized through loss backpropagation and gradient descent (Figure 2(B)).[23,25] A subset of neural networks known as convolutional neural networks (CNN) were developed for imaging applications and take advantage of spatial neighborhood relationships.[23] CNNs shifted the focus from optimizing edge weights among various layers of interconnected nodes to optimizing layered convolutional kernel weights (Figure 2(C)). The convolutional operation can be represented as a Toeplitz matrix multiplication, and a bias term can be incorporated into the matrix multiplication.[23] The reformulation of convolutional kernel weights as matrix operations has enabled the combination of DL methods with graphical processing units (GPUs), which are particularly efficient at matrix operations.[23,26,27]

Deep learning formulation

The typical process of DL involves using input data (either simulation or experimental results) to find a near-optimal set of model parameters that minimize a specified loss function, at which point the DL model has approximated a desired end-to-end functional mapping $f:\ X \rightarrow Y$. The model parameters are typically optimized using a variant of stochastic optimization strategies, such as stochastic gradient descent (SGD) or Adam, by which a mini-batch of input data is processed and a loss function is calculated. The derivative of the model loss with respect to each of the model parameters can then be backpropagated and the model parameters updated accordingly.

Deep learning needed in PACT and PAM

There is a clear difference between the DL formulation in PACT and PAM. In PACT, the final image needs to be reconstructed from the signals received by different transducer elements. DL models in PACT can be used as a pre-processing or post-processing step in the image reconstruction, replace the traditional reconstruction altogether, or be incorporated into an iterative reconstruction. As PAM does not require inverse reconstruction, deep learning models can directly map input signals to output images and improve the image quality accordingly. DL is especially suited for many of PAT's current challenges, like improving ill-posed reconstruction, removing artifacts, denoising channel data, improving spatial resolution, and upsampling sparse scanning input data, as DL is an efficient, GPU-accelerated method for the robust approximation of non-linear spatial mappings in reasonable optimization time scales.[26–35]

Deep learning in PACT

PACT reconstruction is often ill posed and prone to artifacts, mostly due to heterogeneous target properties (e.g., speed of sound) and system parameters such as limited-view, limited-bandwidth detection, and sparse sampling. Traditional reconstruction methods often incorporate implicit or explicit prior knowledge such as l_1, l_2, and total variation (TV) regularization[16,36] to optimize the ill-posed inverse process, which are typically very time consuming and highly sensitive to noise. By contrast, DL-based approaches, such as model-based learning, have replaced

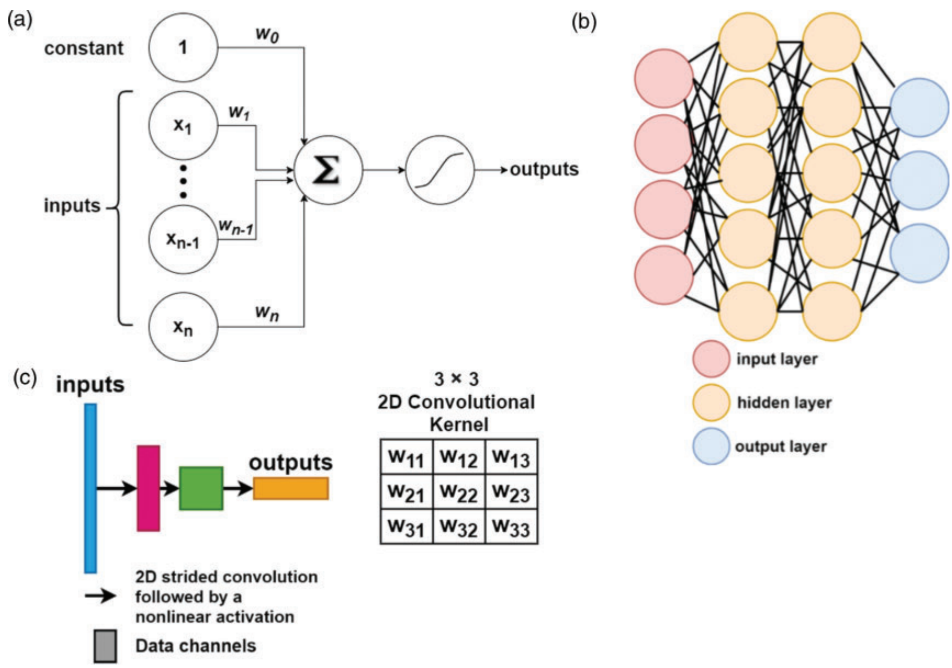

Figure 2. Depiction of representative DL concepts. The (A) perceptron-style neuron, (B) multilayered neural network, and (C) a simple CNN with a 2D convolutional kernel. (A color version of this figure is available in the online journal.)

the traditional regularization terms with a learned regularization term, and thus can be less time consuming.

So far, there has been a variety of deep learning formulations proposed to address the ill-posed reconstruction problem. Several deep learning approaches train pre-processing models to improve the channel data, while other post-processing models have been used to remove reconstruction artifacts. Some deep learning models have even been used to replace the inverse operator entirely, and others are used to improve quantitative imaging, like functional or molecular imaging (Figure 3).

DL for direct image reconstruction

One of the earliest works on deep learning-based direct image reconstruction was reported by Waibel *et al.*, in which the authors used U-Net to estimate the initial pressure distribution directly from the detected channel data.[37] A CNN model was trained with simulated data and achieved similar performance with post-processing methods.[37] Around the same time, Anas *et al.* trained a deep CNN with large 9×9 convolutional kernels and dense blocks with dilated convolutions on simulation data to directly perform beamforming on channel data—notably outperforming traditional delay-and-sum (DAS) beamformed results.[38] Nevertheless, both of these early implementations were not tested with experimental data. Another similar study employed Res-UNet and achieved success on phantom experiments.[39] Subsequent recent works have proven that using modified channel data yields improved performance.[40–42] For example, Kim *et al.* applied a delay on the channel data for each spatial point before feeding the data into a U-Net.[41] This so called upgUNET simplified the learning process by exposing the model to the back-propagation of channel data[41] and improved the structural similarity index (SSIM)[43] modestly on both simulated and experimental data (see Figure 4 (A)).[41] Guan *et al.* also applied a similar approach using FD U-Net—an advanced U-Net architecture with a four-layered dense block at each level,[40] which outperformed the same model trained for post-processing reconstructed images and was able to correct limited-view and sparse-sampling artifacts.[40] However, Guan *et al.* did not test their model on experimental data. Conversely, Lan *et al.* utilized both delay-and-sum images and raw channel data to train a dual-encoder Y-Net as the inverse model, which showed a slight improvement over post-processing models.[42] More recently, Lan *et al.* proposed their BSR-Net utilizing a novel residual separation block which combines the positional information obtained by applying a channel delay with the reconstruction result (unreviewed preprint, URL: https://arxiv.org/abs/2012.02472). In addition, BSR-Net relies on a space-based calibration and removal module (SCRM) and two novel losses (response loss and overlay loss) to produce results that have even fewer artifacts than the ground truth images when tested on simulation data. With similar training times, all of these approaches have demonstrated improved image quality over traditional backprojection-based image reconstruction.

DL for pre-processing channel data

Besides direct image reconstruction, DL has also been used to process raw channel data before performing traditional beamforming. For example, Gutta *et al.* used a fully connected deep neural network (FC-DNN) to correct the sonograms acquired by each transducer channel.[19] The ultrasound transducer is effectively a bandpass filter that blurs sharp edges and suppresses low-frequency signal components. This approach was able to broaden the

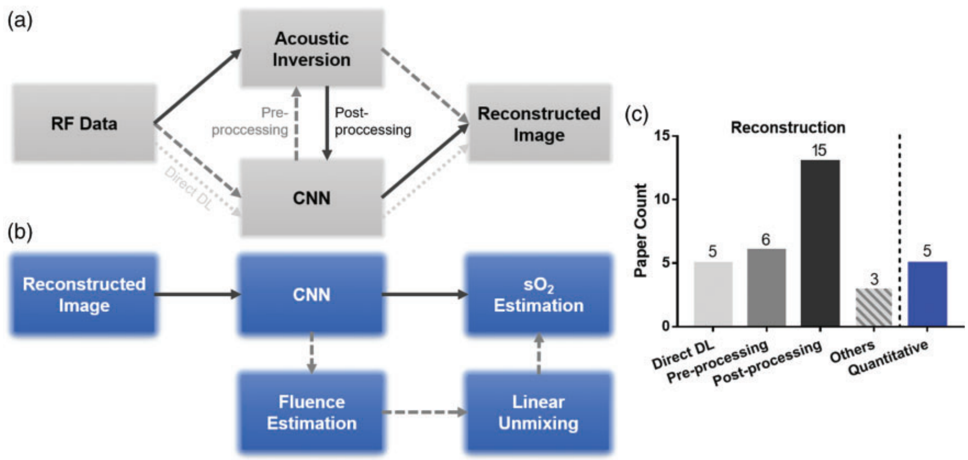

Figure 3. Deep learning strategies in PACT. (A) General pathways to apply DL to PACT reconstruction. Pre-processing CNNs correct raw data before reconstruction, while post-processing DL is applied to post-reconstruction images. CNNs can also be used to replace acoustic inversion altogether. (B) General paths to use DL for quantitative PACT, w.r.t blood oxygenation estimation. (C) Paper count on selected DL-based PACT topics. Only referenced papers are included in the counts. (A color version of this figure is available in the online journal.)

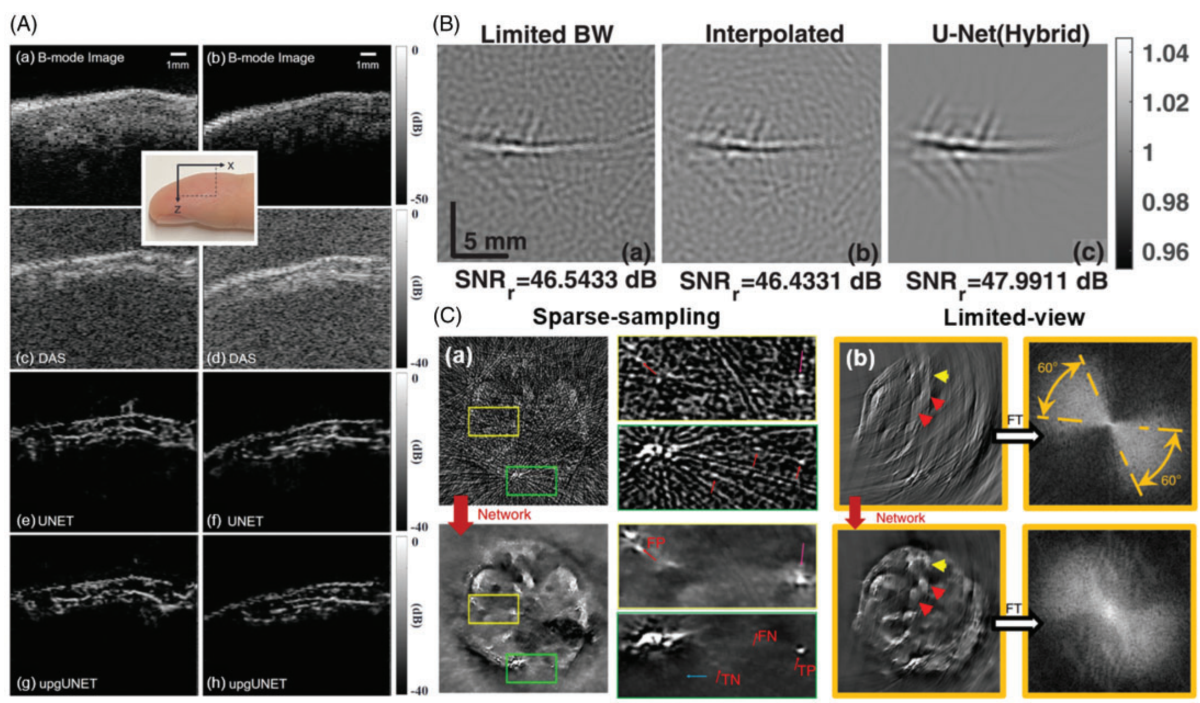

Figure 4. Representative *in vivo* DL-based PACT reconstruction. (A) Direct reconstruction of a human finger using upgUNET.[41] (B) Reconstructed images for *in vivo* rat brain data with enhanced bandwidth using U-Net.[44] (C) Sparse-sampling and limited-view artifacts of whole-body mouse images are greatly suppressed using U-Net trained by *in vivo* data.[17] (A color version of this figure is available in the online journal.)

bandwidth of the received channel data, and thus increase the SNR of beamformed images by ~6 dB[19]. A follow-up work applied U-Net on channel data for resolution improvement and bandwidth broadening, as shown in Figure 4(B), and was validated on experimental data.[44] Allman *et al.* employed VGGNet to identify point sources from the channel data and reduce reflection artifacts in the reconstructed images, which was useful for detecting the catheter tip in PACT-guided surgical intervention.[45–47] In summary, pre-processing DL applications in PACT are able to identify and remove noise, reflection artifacts, and bandlimited artifacts directly from the channel data, which would otherwise very difficult to separate from reconstructed images.

DL for post-processing reconstructed images

DL has also been applied in PACT as a post-processing method for artifact removal (Figure 3(C)). Despite typically suffering from artifacts, reconstructed PACT images are

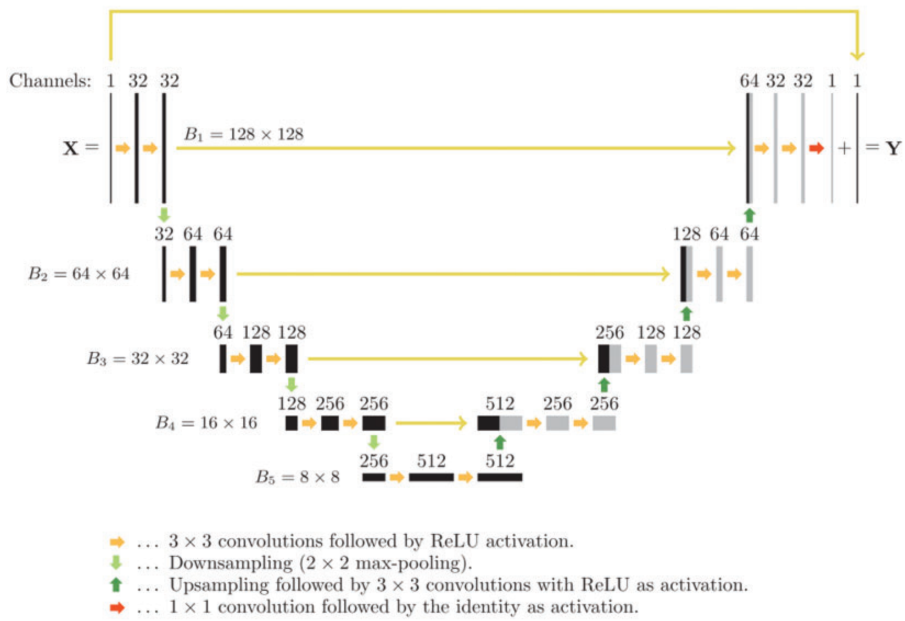

Figure 5. Depiction of a modified U-Net implemented by Antholzer *et al.*[16] (A color version of this figure is available in the online journal.)

able to provide a good approximation of the initial pressure distribution. Thus, there are fewer features and filters for the neural network to learn, which simplifies and stabilizes the training process. Specifically, post-processing DL methods have been applied for identifying and reducing artifacts that result from sparse-sampling, limited-view, and limited-bandwidth detection. For instance, Antholzer *et al.* and Guan *et al.* have used U-Net and FD U-Net respectively to remove undersampling artifacts in reconstructed PACT images acquired by ring-shaped ultrasound arrays (see Figure 5)[16,48] FD U-Net outperformed U-Net on a simulated mouse brain vasculature dataset.[48] The distinctive curved-stripe artifacts resulting from sparse sampling were significantly suppressed in simulated data, improving structural visibility in the reconstructed images. Knowledge-infusion generative adversarial network (GAN), an advanced model architecture with two sub-networks competing against each other, has also been proposed for addressing the sparse sampling issue.[49] Similarly, a deep CNN has been applied to the truncated singular-value-decomposition (SVD) of reconstructed images, in order to resolve the limited-view issue.[50] The performance of these early methods has yet to be tested on *in vivo* data. A novel approach by Zhang *et al.*, known as Dual Domain U-net (DuDoUnet), utilizes input information from both the time domain and frequency domain in the form of DAS image and k-space image inputs, respectively (unreviewed preprint, URL: https://arxiv.org/abs/2011.06147). This DuDoUnet is able to effectively utilize information from both domains in its limited-view artifact removal procedure through its use of information sharing blocks (ISBs) and a mutual information (MI) constraint. This dual-domain technique outperformed the SSIM of U-net, Y-Net, and DEU-net when trained on simulation data. The PA-Fuse method by Awasthi *et al.* first performs both linear backprojection and regularized inversion (TV or Lanczos Tikhonov), and then fuses these reconstruction results together using a Siamese network.[51] This method was shown to outperform both traditional regularized inversion methods and contemporary fusion methods, such as modified guided filtering, under a variety of noise levels. Similarly, Antholzer *et al.* showed how compressed sensing PAT inversion, such as data acquired using a sparse sampling or Bernoulli measurements, can benefit from joint l_1-minimization and Tikhonov regularization using a learned regularization term (i.e. NETT).[52]

DL-based post-processing approaches have also been used to remove different PACT artifacts simultaneously. The DL-based methods generally outperform traditional beamforming methods and can accommodate different detection geometries. For PACT with a linear-array transducer, a stabilized GAN model—Wasserstein GAN (WGAN) with gradient clipping—has been employed to reduce both limited-view and limited-bandwidth artifacts, improving the contrast-to-noise ratio of *in vivo* data.[53] In another study, Godefroy *et al.* employed an external CMOS camera to acquire ground truth for reconstructed PACT images deteriorated by artifacts.[54] A modified U-Net was used with the pairwise camera-and-PACT images for training.[54] Using an optical camera to obtain ground truth is, however, not applicable for deep tissue imaging. For PACT with a ring-array transducer, Zhang *et al.* used a deep CNN with 10 layers to significantly suppress both undersampling and limited-view artifacts in simulated data and *in vivo* mouse brain data.[55] Similarly, Lu *et al.* proposed the use of a GAN model for a ring-array PACT system with a limited view.[56] Rather than using simulated data for training, Davoudi *et al.* have directly utilized experimental data to train a U-Net for removing both sparse-sampling and limited-view artifacts.[17] Full-view reconstructed images were used as the ground truth to train a CNN model to improve sub-aperture reconstructed images.[17] By training the CNN model with experimental data directly, the model avoids biases in simulated training data and improves its quick adaptation onto *in vivo* data (Figure 4(C)).[17]

DL methods have also been used for improving the SNR of PACT systems using laser emission diodes (LED) as the light source. LEDs are cost-effective and compact, but have low output power and generate weak PA signals. Singh *et al.* used a high-power laser to acquire pre- and post-average PACT images in order to train a U-Net model, which was able to improve the SNR of the LED-based images.[57] However, the training and testing data were not from the same imaging platform, which may induce biases like detection geometry and bandwidth. By contrast, Anas *et al.* applied a recurrent neural network (RNN) on multiple consecutive reconstructed images from the same system.[58] The RNN, with its long-short-term-memory, was capable of extracting noise from the signal's temporal information and outperformed both simple averaging and a conventional CNN.[58] Hariri *et al.* utilized a multi-level wavelet CNN (MWCNN) to restore PA image quality and contrast under a variety of optional fluence conditions, such as the low fluence levels provided by LED sources.[59] This unique MWCNN architecture replaced pooling blocks with discrete wavelet transforms and upsampling blocks with inverse wavelet transforms, thereby removing information loss during the downsampling and upsampling procedures. The DL model of Hariri *et al.* showed promising results for a variety of targets under different optical fluence conditions. DL-based denoising approaches can also be applied to improve the image quality of traditional PACT systems when imaging deeper targets, which also suffer from deteriorated SNR due to optical attenuation. Manwar *et al.* successfully used a U-Net to improve the SNR of *in vivo* deep tissue regions when using low laser energy.[60]

DL-based resolution enhancement has been explored in PACT. For a circular detection geometry, Rajendran and Pramanik have applied an advanced FD U-Net, named TARES, to improve tangential resolution of reconstructed PACT images far away from the scanning center (or close to the transducer surface).[61] TARES outperformed FD U-Net on both phantom and *in vivo* rat brain data, and has shown a great potential for enhancing the resolution of other detection geometries—especially those with a linear-array transducer.

Integrated DL-enhanced PACT reconstruction

Instead of targeting only a single step of the PACT image reconstruction, some researchers have applied CNNs at multiple steps of the reconstruction process. For example, to remove limited-view and sparse-sampling artifacts with a circular detection geometry, Tong *et al.* used a CNN model for both direct reconstruction and post-reconstruction processing to optimize final image quality.[62] Both the original channel data and its time derivative were used as the model input.[62] This approach outperformed standalone DL methods.[62] Nonetheless, training two CNNs significantly increases the training time and requires substantially larger datasets.

CNNs can also act as a regularizer in model-based PACT reconstruction (i.e., model-based learning), which incorporates the PA forward operator to account for the imaging system's physical parameters such as the speed of sound and limited detection angles. Model-based methods are superior in accurately estimating the initial pressure distribution, at the cost of time-consuming iterative optimization. Traditional model-based methods employ regularizing terms such as l_1, l_2, and total variation (TV). However, such simple regularizers, which aim for noise reduction or edge preservation, often fail to handle the complex features of experimental data.[18] Instead, using deep gradient descent (DGD) to correct limited-view artifacts, Hauptmann *et al.* replaced TV regularization with a trained CNN,[18] and demonstrated better reconstruction quality on human palm vasculature images.[18] Boink *et al.* replaced both the primal and dual domain with CNNs in their learned primal-dual method (L-PD),[63] which improved the joint reconstruction and vessel segmentations with limited-view detection.[63] Similarly Hauptmann *et al.* devised a learned iterative reconstruction utilizing a fast and approximate forward model that is based on k-space methods for PAT, called fast forward PAT (FF-PAT) reconstruction, for improving subsampled data acquired using a limited detection aperture.[64] This technique was validated *in vivo* and shown to have speeds 32 times faster than traditional TV variation reconstructions, while still maintaining a competitive PSNR. Yang *et al.* also combined deep learning, in the form of recurrent inference machines (RIM), with PAT k-space methods to accelerate iterative PAT reconstruction.[65]

DL-assisted quantitative PACT imaging

Over the past few years, DL-based methods have been investigated for quantitative PACT. Compared with pure ultrasound imaging, PAT is advantageous in functional and molecular imaging. Spectroscopic PA measurements can be performed to quantify the concentrations of different endogenous chromophores (e.g., deoxy- and oxy-hemoglobin) and exogenous probes (e.g., nanoparticles and reporter gene products). However, quantitative PAT has long been challenging for deep-seated targets, because the optical fluence attenuation is highly wavelength dependent in biological tissues, a phenomenon called spectral coloring. The spectral coloring may result in an erroneous quantification of deep tissue components using conventional spectral unmixing methods.[66] Recent DL approaches in PACT have provided a promising solution to deep tissue quantitative imaging, by either completely replacing the spectral unmixing algorithms or by better estimating the optical fluence at different wavelengths (Figure 3(B)). As an example, we will introduce various DL methods for quantifying the oxygen saturation of hemoglobin (sO_2) in blood vessels, which is critical information for studying cancer hypoxia and tissue inflammation.

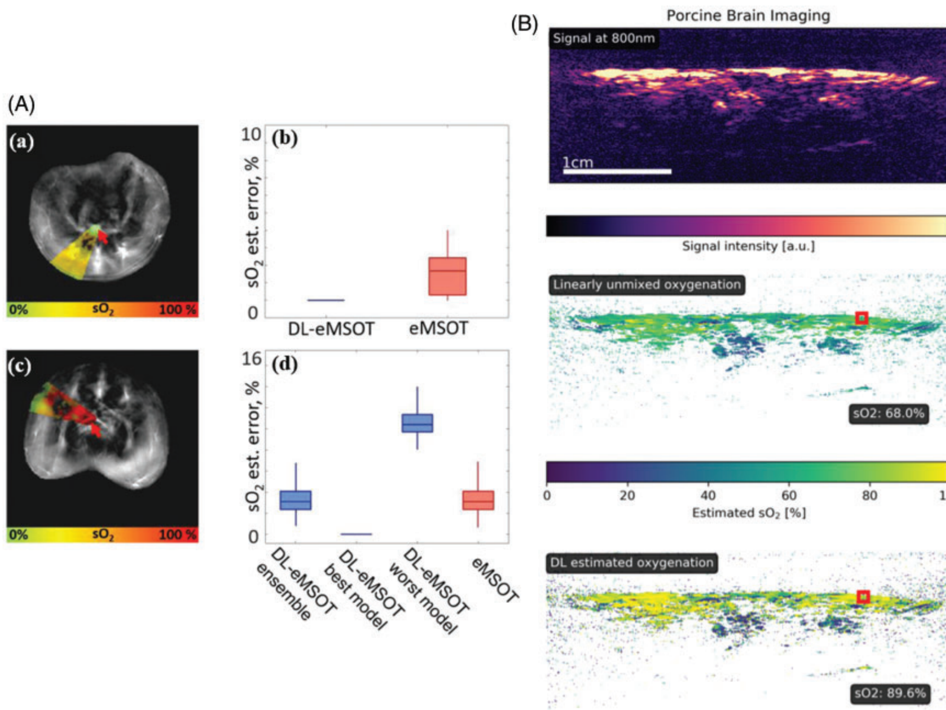

Figure 6. Representative *in vivo* DL-based quantitative PACT. (A) DL-eMSOT estimation of blood oxygenation shows significantly reduced error when compared to conventional eMSOT, on abdominal cross-sections of two mice.[69] (B) sO_2 estimation derived from a multispectral PA image of a pig brain using LSD-qPAI, showing improved accuracy compared to linear unmixing (unreviewed preprint, URL: https://arxiv.org/abs/1902.05839). (A color version of this figure is available in the online journal.)

Early attempts by Cai *et al.* and Yang *et al.* explored ResU-Net and DR2U-Net with multi-wavelength reconstructed PAT images.[67,68] The DL-reconstructed sO_2 map suggested better performance than linear unmixing.[67] Olefir *et al.* used dimensionality-reduced spectra as the input in their bi-directional RNN model, named DL-eMSOT[69]. DL-eMSOT predicted maps of eigenfluence in deep tissue,[69] which were subsequently used for linear unmixing of the oxy- and deoxy-hemoglobin concentrations.[69] DL-eMSOT takes advantage of a sequential-learning RNN and achieved less error than the conventional eMSOT approach (Figure 6(A)).[69] In the LSD-qPAI approach, Gröhl *et al.* applied a fully connected neural network on multi-spectral (26 wavelengths) pressure maps (unreviewed preprint, URL: https://arxiv.org/abs/1902.05839). This method yielded accurate sO_2 estimations on phantom experiments and *in vivo* porcine brain data (Figure 6(B)). A 3D sO_2 estimation is highly desired for volumetric quantification of the tissue's oxygen status. Bench *et al.* applied a 3D encoder-decoder neural network to predict volumetric sO_2 maps.[70] This method, however, has not yet been adapted to *in vivo* data due to the complexity of tissue properties.[70]

Deep learning in photoacoustic microscopy

Without the pressing need for improving inverse image reconstruction, the utilization of DL techniques in PAM has been relatively sparse in comparison to that in PACT. PAM does not suffer from the difficulties that arise from an ill-posed reconstruction, but there are still a number of ways deep learning has been utilized to augment PAM capabilities, including the spatial resolution, imaging speed, and SNR.

Before the widespread utilization of deep learning in PAM, a precursor technique known as dictionary learning was utilized by Govinahallisathyanarayana *et al.* to remove reverberation signals from mouse brain images without compromising the underlying microvasculature structure.[71] One of the first utilizations of deep learning for enhancing PAM was published by Chen *et al.* in which a simple three-layered CNN model, with various kernel sizes tested, was implemented to remove motion artifacts from OR-PAM images(see Figure 7(A)).[21]

One of the major utilizations of deep learning in PAM is to upsample sparsely sampled OR-PAM images, thereby shortening image acquisition time without substantially degrading image quality. We developed the first DL technique for this purpose with a now open source dataset of mouse brain PAM images.[72] We trained a modified fully dense U-net architecture (FD U-net).[20] This pivotal publication utilized fully sampled OR-PAM images as the ground truth and artificially downsampled images, training and testing a CNN with varying degrees of downsampling along either imaging axis (see Figure 7(B)). This work successfully demonstrated the feasibility of using CNNs to upsample PAM images, using only approximately 2% effective pixels, and made open source a large collection of murine brain images for further deep learning research. Soon after this work came a report by Zhou *et al.* that

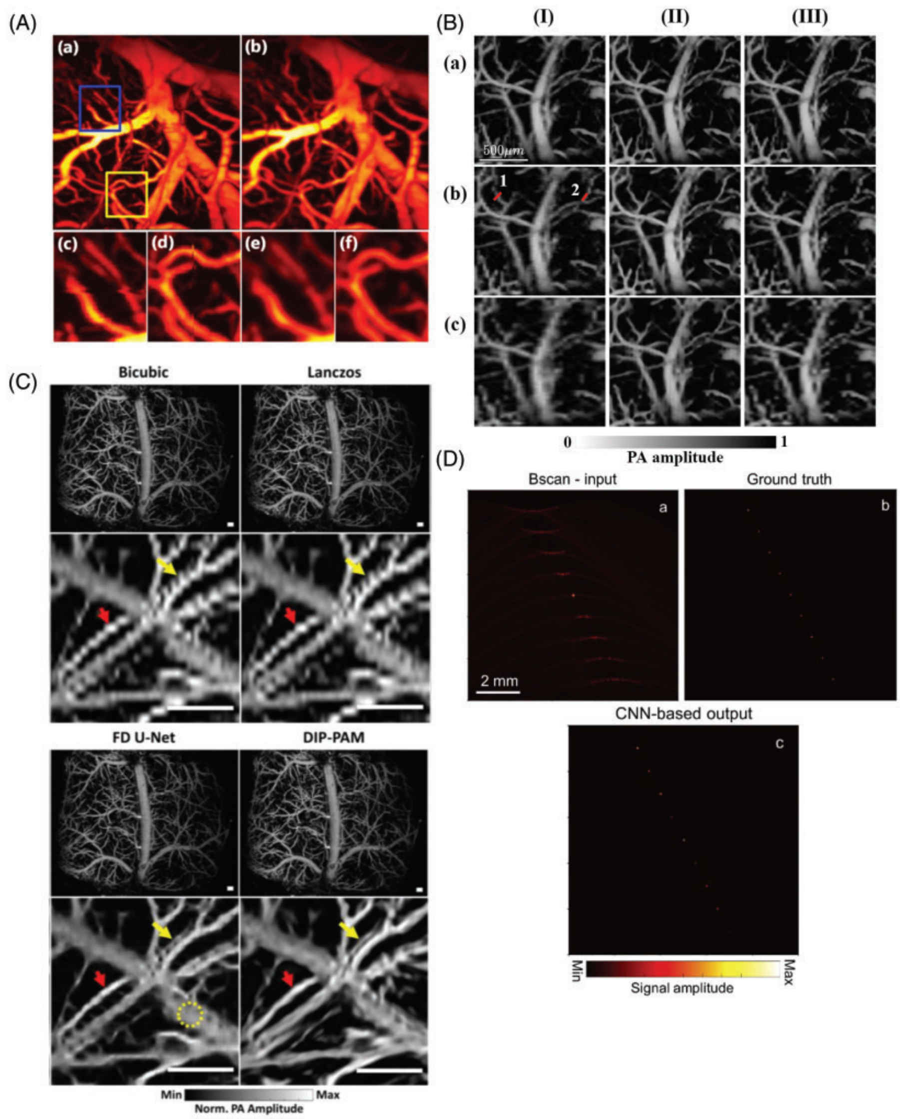

Figure 7. Representative DL methods in PAM. (A) Motion artifact removal using a CNN.[21] (B) FD U-net upsampling performance on OR-PAM *in vivo* mouse brain images at various downsampling ratios of 20%, 5%, and 2% effective pixels, respectively.[20] (C) Comparison of Deep Prior with other upsampling methods on a mouse brain image (unreviewed preprint, URL: https://arxiv.org/abs/2010.12041). (D) DL resolution enhancement of out of focus plane signal.[22] (A color version of this figure is available in the online journal.)

utilized a CNN architecture with squeeze-and-excitation (SE) blocks to perform a similar upsampling procedure.[73]

Following these more traditional CNN implementations, we utilized an innovative deep prior methodology that iteratively refines undersampled PAM images using a deep learning prior—a "Deep Prior" (unreviewed preprint, URL: https://arxiv.org/abs/2010.12041). Results comparing Deep Prior to other DL architecture is shown in Figure 7(C). This work is of particular note because it does not require training on a large PAM dataset with established ground truth, thereby circumventing the data bottleneck that currently exists in many DL-based applications. Deep learning, in the form of a feedforward denoising CNN, has also recently been used by Tang *et al.* to improve the low SNR of PAM images.[74] Most recently, there has been some promising work by Sharma *et al.* that uses an FD U-net to both denoise and enhance the resolution of AR-PAM images, especially outside the focus plane (Figure 7(D)).[22]

Conclusions

Both the fields of photoacoustic imaging and deep learning have progressed at an exceedingly accelerated rate over recent years.[75–77] This innovative intersection of fields has served as the staging ground for a number of important innovations in both PACT and PAM, augmenting the capabilities of both imaging methodologies and overcoming many of the persistent challenges facing the field of photoacoustic imaging. We hope this concise review can succinctly summarize recent exciting technological advances and make them accessible to the broader scientific community.

This concise summary of recent work implementing deep learning in PACT and PAM has highlighted several remaining challenges and avenues for promising future research. One such challenge for implementing deep learning in PACT is the current reliance on simulation data and the lack of large, open source repositories of *in vivo* data. Deep learning models learn various features from training data, but simulation data inherently lack much of the variability that exists in *in vivo* data. This gap between simulation data and *in vivo* data makes model extrapolation to *in vivo* applications difficult. The two apparent solutions that exist to address this concern are for the community to create a large, open source repository of variable *in vivo* training examples, or to improve the quality of simulation data to better mimic *in vivo* cases. *In vivo* data can be readily obtained for training DL models to improve predictably degraded PAT data, such as spatially undersampled OR-PAM data or sparsely sampled/limited-view PACT data, so long as the system's degradation function can be replicated through post-processing. This method of establishing *in vivo* ground truth can be done by acquiring fully sampled or full-view data (i.e. the ground truth), and subsequently applying artificial degradation to synthesize the expected physically degraded input data. However, acquiring ground truth *in vivo* PAT data to train DL models to exceed current state-of-the-art system capabilities remains a looming challenge for PAT and medical DL researchers alike. Despite this challenge, the future incorporation of deep learning into photoacoustic imaging technology and eventual clinical adoption will require robust models that can readily adapt to a variety of *in vivo* conditions—many of which, like sparsely sampled, limited-view, and limited-bandwidth detection, will be in non-ideal environments.

A key area of future research for both PACT and PAM will be the upstream integration of deep learning techniques with system design and engineering. For example, one method that has been used to integrate the PA forward operator into a deep learning formulation has been the model-based learning utilized by Hauptmann *et al.* and Boink *et al.* However, these iterative methods can be time consuming, and have yet to truly integrate the deep learning approach into the system design. Deep learning techniques have typically been applied to pre-existing system configurations, but the next generation of PACT and PAM DL applications will likely be designed with both deep learning and compressed sensing techniques at the forefront. The light source, detector arrangement, scanning mechanism, and data acquisition can be optimized based on the accompanying DL models. This will make it possible to achieve full integration of deep learning and PA imaging, thereby allowing the next generation of "smart" PA technology to far exceed what has come before.

AUTHORS' CONTRIBUTIONS

All authors discussed the topics and wrote the manuscript.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

This work was supported by the National Institutes of Health (R01 EB028143, R01 NS111039, RF1 NS115581, R21 EB027304, R21EB027981, R43 CA243822, R43 CA239830, R44 HL138185); Duke Institute of Brain Science Incubator Award; American Heart Association Collaborative Sciences Award (18CSA34080277); Chan Zuckerberg Initiative Grant on Deep Tissue Imaging 2020–226178 by Silicon Valley Community Foundation.

ORCID iDs

Anthony DiSpirito https://orcid.org/0000-0002-3232-7353
Manojit Pramanik https://orcid.org/0000-0003-2865-5714

ACKNOWLEDGMENTS

The authors thank Dr. Caroline Connor for editing the manuscript.

REFERENCES

1. Xia J, Yao J, Wang LV. Photoacoustic tomography: principles and advances. *Electromagn Waves* 2014;**147**:1–22
2. Upputuri PK, Pramanik M. Recent advances toward preclinical and clinical translation of photoacoustic tomography: a review. *J Biomed Opt* 2016;**22**:041006
3. Das D, Sharma A, Rajendran P, Pramanik M. Another decade of photoacoustic imaging. *Phys Med Biol* 2020;**66**: 05TR01.
4. Beard P. Biomedical photoacoustic imaging. *Interface Focus* 2011;**1**:602–31
5. Wang LV, Yao J. A practical guide to photoacoustic tomography in the life sciences. *Nat Methods* 2016;**13**:627–38
6. Bell AG. On the production and reproduction of sound by light. *Am J Sci* 1880;**s3-20**:305–24
7. Manohar SR, Daniel PB. Photoacoustics: a historical review. *Adv Opt Photon* 2016;**8**:586–617
8. Xu M, Wang LV. Photoacoustic imaging in biomedicine. *Rev Sci Instrum* 2006;**77**:041101
9. Wang LV. Multiscale photoacoustic microscopy and computed tomography. *Nat Photon* 2009;**3**:503–9
10. Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, Baraniuk RG. Single-pixel imaging via compressive sampling. *IEEE Signal Process Mag* 2008;**25**:83–91
11. Liang J, Zhou Y, Winkler AW, Wang L, Maslov KI, Li C, Wang LV. Random-access optical-resolution photoacoustic microscopy using a digital micromirror device. *Opt Lett* 2013;**38**:2683–6
12. Haltmeier M, Berer T, Moon S, Burgholzer P. Compressed sensing and sparsity in photoacoustic tomography. *J Opt* 2016;**18**:114004
13. Jeon S, Kim J, Lee D, Baik JW, Kim C. Review on practical photoacoustic microscopy. *Photoacoustics* 2019;**15**:100141
14. Kim J, Kim JY, Jeon S, Baik JW, Cho SH, Kim C. Super-resolution localization photoacoustic microscopy using intrinsic red blood cells as contrast absorbers. *Light Sci Appl* 2019;**8**:103
15. Agranovsky M, Kuchment P. Uniqueness of reconstruction and an inversion procedure for thermoacoustic and photoacoustic tomography. *Inverse Prob* 2007;**23**:2089–102
16. Antholzer S, Haltmeier M, Schwab J. Deep learning for photoacoustic tomography from sparse data. *Inverse Probl Sci Eng* 2019;**27**:987–1005
17. Davoudi N, Deán-Ben XL, Razansky D. Deep learning optoacoustic tomography with sparse data. *Nat Mach Intell* 2019;**1**:453–60
18. Hauptmann A, Lucka F, Betcke M, Huynh N, Adler J, Cox B, Beard P, Ourselin S, Arridge S. Model-based learning for accelerated, limited-view 3-d photoacoustic tomography. *IEEE Trans Med Imaging* 2018;**37**:1382–93

19. Gutta S, Kadimesetty VS, Kalva SK, Pramanik M, Ganapathy S, Yalavarthy PK. Deep neural network-based bandwidth enhancement of photoacoustic data. *J Biomed Opt* 2017;**22**:116001
20. DiSpirito A, III, Li D, Vu T, Chen M, Zhang D, Luo J, Horstmeyer R, Yao J. Reconstructing undersampled photoacoustic microscopy images using deep learning. *IEEE Trans Med Imaging* 2020;**40**:562–70
21. Chen X, Qi W, Xi L. Deep-learning-based motion-correction algorithm in optical resolution photoacoustic microscopy. *Vis Comput Ind Biomed Art* 2019;**2**:12
22. Sharma A, Pramanik M. Convolutional neural network for resolution enhancement and noise reduction in acoustic resolution photoacoustic microscopy. *Biomed Opt Express* 2020;**11**:6826–39
23. Goodfellow I, Bengio Y, Courville A. *Deep learning*. Cambridge, MA: MIT Press, 2016
24. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;**61**:85–117
25. Hinton G, LeCun Y, Bengio Y. Deep learning. *Nature* 2015;**521**:436–44
26. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: overview, challenges and the future. In: Dey N, Ashour AS, Borra S (eds) *Classification in BioApps: automation of decision making*. Cham: Springer International Publishing, 2018, pp. 323–50
27. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sanchez CI. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;**42**:60–88
28. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data* 2015;**2**:1–21
29. Lee J-G, Jun S, Cho YW, Lee H, Kim GB, Seo JB, Kim N. Deep learning in medical imaging: general overview. *Korean J Radiol* 2017;**18**:570
30. Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. *Optica* 2017;**4**:1437–43
31. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *RadioGraphics* 2017;**37**:505–15
32. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML. Deep learning in medical imaging and radiation therapy. *Med Phys* 2019;**46**:e1–e36
33. Martorell-Marugán J, Siham T, Yassir B, Coral del V, Igor Z, Francisco H, Pedro C-S, Holger H. Deep learning in omics data analysis and precision medicine. *Comput Biol* 2019;**3**:37–53
34. Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: advantages, potential, and challenges. *Phys Med Biol* 2020;**65**:05TR01
35. Pradhan P, Guo S, Ryabchykov O, Popp J, Bocklitz TW. Deep learning a boon for biophotonics? *J Biophotonics* 2020;**13**:e201960186
36. Yalavarthy PK, Kalva SK, Pramanik M, Prakash J. Non-local means improves total-variation constrained photoacoustic image reconstruction. *J Biophotonics* 2021;**14**:e202000191
37. Waibel D, Gröhl J, Isensee F, Kirchner T, Maier-Hein K, Maier-Hein L. Reconstruction of initial pressure from limited view photoacoustic images using deep learning. In: Oraevsky AA, Wang LV (eds) *Photons Plus Ultrasound: Imaging and Sensing 2018*. International Society for Optics and Photonics, SPIE, vol. 10494, 2018, p.104942S
38. Anas EMA, Zhang HK, Audigier C, Boctor EM. *Robust photoacoustic beamforming using dense convolutional neural networks*. Cham: Springer International Publishing, 2018, pp. 3–11
39. Feng J, Deng J, Li Z, Sun Z, Dou H, Jia K. End-to-end Res-Unet based reconstruction algorithm for photoacoustic imaging. *Biomed Opt Express* 2020;**11**:5321–40
40. Guan S, Khan AA, Sikdar S, Chitnis PV. Limited-View and sparse photoacoustic tomography for neuroimaging with deep learning. *Sci Rep* 2020;**10**:1–12
41. Kim MW, Jeng G-S, Pelivanov I, O'Donnell M. Deep-learning image reconstruction for real-time photoacoustic system. *IEEE Trans Med Imag* 2020;**39**:3379–90
42. Lan H, Jiang D, Yang C, Gao F, Gao F. Y-Net: hybrid deep learning image reconstruction for photoacoustic tomography in vivo. *Photoacoustics* 2020;**20**:100197
43. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004;**13**:600–12
44. Awasthi N, Jain G, Kalva SK, Pramanik M, Yalavarthy PK. Deep neural network based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography. *IEEE Trans Ultrason Ferroelectr Freq Control* 2020;**67**:2660–73
45. Reiter A, Bell MAL. A machine learning approach to identifying point source locations in photoacoustic data. *Photons Plus Ultrasound: Imag Sens* 2017; **10064**:100643J
46. Allman D, Reiter A, Bell MAL. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE Trans Med Imaging* 2018;**37**:1464–77
47. Allman D, Assis F, Chrispin J, Bell MAL. A deep learning-based approach to identify in vivo catheter tips during photoacoustic-guided cardiac interventions. *Photons Plus Ultrasound: Imag Sens* 2019; **10878**: 108785E
48. Guan S, Khan AA, Sikdar S, Chitnis PV. Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal. *IEEE J Biomed Health Inform* 2019;**24**:568–76
49. Lan H, Zhou K, Yang C, Cheng J, Liu J, Gao S, Gao F. Ki-GAN: knowledge infusion generative adversarial network for photoacoustic image reconstruction in vivo. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Shenzhen, China, 13–17 October 2019, pp.273–81. Berlin: Springer
50. Schwab J, Antholzer S, Nuster R, Paltauf G, Haltmeier M. Deep Learning of truncated singular values for limited view photoacoustic tomography. *Photons Plus Ultrasound: Imag Sens* 2019;**10878**:1087836
51. Awasthi N, Prabhakar KR, Kalva SK, Pramanik M, Babu RV, Yalavarthy PK. PA-Fuse: deep supervised approach for the fusion of photoacoustic images with distinct reconstruction characteristics. *Biomed Opt Express* 2019;**10**:2227–43
52. Antholzer S, Schwab J, Bauer-Marschallinger J, Burgholzer P, Haltmeier M. NETT regularization for compressed sensing photoacoustic tomography. *SPIE BiOS* 2019;**10878**:108783B
53. Vu T, Li M, Humayun H, Zhou Y, Yao J. A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer. *Exp Biol Med* 2020;**245**:597–605
54. Godefroy G, Arnal B, Bossy E. Compensating for visibility artefacts in photoacoustic imaging with a deep learning approach providing prediction uncertainties. *Photoacoustics* 2020;**21**:100218
55. Zhang H, Hongyu L, Nyayapathi N, Wang D, Le A, Ying L, Xia J. A new deep learning network for mitigating limited-view and under-sampling artifacts in ring-shaped photoacoustic tomography. *Comput Med Imaging Graph* 2020;**84**:101720
56. Lu T, Chen T, Gao F, Sun B, Ntziachristos V, Li JL. LV-GAN. A deep learning approach for limited-view optoacoustic imaging based on hybrid datasets. *J Biophotonics* 2020; **14**:e202000325
57. Singh MKA, Sivasubramanian K, Sato N, Ichihashi F, Sankai Y, Xing L. Deep learning-enhanced LED-based photoacoustic imaging. *Photons Plus Ultrasound: Imag Sens* 2020; **11240**:1124038
58. Anas EMA, Zhang HK, Kang J, Boctor E. Enabling fast and high quality LED photoacoustic imaging: a recurrent neural networks based approach. *Biomed Opt Express* 2018;**9**:3852–66
59. Hariri A, Alipour K, Mantri Y, Schulze JP, Jokerst JV. Deep learning improves contrast in low-fluence photoacoustic imaging. *Biomed Opt Express* 2020;**11**:3360–73
60. Manwar R, Li X, Mahmoodkalayeh S, Asano E, Zhu D, Avanaki K. Deep learning protocol for improved photoacoustic brain imaging. *J Biophotonics* 2020;**13**:e202000212
61. Rajendran P, Pramanik M. Deep learning approach to improve tangential resolution in photoacoustic tomography. *Biomed Opt Express* 2020;**11**:7311–23
62. Tong T, Huang W, Wang K, He Z, Yin L, Yang X, Zhang S, Tian J. Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data. *Photoacoustics* 2020;**19**:100190
63. Boink YE, Manohar S, Brune C. A Partially-Learned algorithm for joint photo-acoustic reconstruction and segmentation. *IEEE Trans Med Imaging* 2019;**39**:129–39

64. Hauptmann A, Cox B, Lucka F, Huynh N, Betcke M, Beard P, Arridge S. *Approximate k-Space models and deep learning for fast photoacoustic reconstruction*. Cham: Springer International Publishing, 2018, pp. 103–11
65. Yang C, Lan H, Gao F. Accelerated photoacoustic tomography reconstruction via recurrent inference machines. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 23–27 July 2019, pp.6371–74. Piscataway, NJ: IEEE.
66. Yang C, Gao F. Eda-net: dense aggregation of deep and shallow information achieves quantitative photoacoustic blood oxygenation imaging deep in human breast. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A (eds) *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin: Springer, 2019, pp. 246–54
67. Cai C, Deng K, Ma C, Luo J. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Opt Lett* 2018;**43**:2752–5
68. Yang C, Lan H, Zhong H, Gao F. Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network. In: *16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 8–11 April 2019, pp.741–44. Piscataway, NJ: IEEE
69. Olefir I, Tzoumas S, Restivo C, Mohajerani P, Xing L, Ntziachristos V. Deep learning-based spectral unmixing for optoacoustic imaging of tissue oxygen saturation. *IEEE Trans Med Imaging* 2020;**39**:3643–54
70. Bench C, Hauptmann A, Cox BT. Toward accurate quantitative photoacoustic imaging: learning vascular blood oxygen saturation in three dimensions. *J Biomed Opt* 2020;**25**:085003
71. Govinahallisathyanarayana S, Ning B, Cao R, Hu S, Hossack JA. Dictionary learning-based reverberation removal enables depth-resolved photoacoustic microscopy of cortical microvasculature in the mouse brain. *Sci Rep* 2018;**8**:985
72. DiSpirito A, III. *Duke PAM dataset*. Switzerland: Zenodo, 2020
73. Zhou J, He D, Shang X, Guo Z, Chen S-L, Luo J. Photoacoustic microscopy with sparse data by convolutional neural networks. *Photoacoustics* 2021; **22**:100242
74. Tang K, Li B, Zhang J, Wei J, Song L, Song X. Denoising method for photoacoustic microscopy using deep learning. In: Kimata M, Shaw JA, Valenta CR (eds) *SPIE future sensing technologies*. Bellingham, Washington: SPIE, 2020, pp. 611–616
75. Hauptmann A, Cox B. Deep learning in photoacoustic tomography: current approaches and future directions. *J Biomed Opt* 2020;**25**:112903
76. Gröhl J, Schellenberg M, Dreher K, Maier-Hein L. Deep learning for biomedical photoacoustic imaging: a review. *Photoacoustics* 2021;**22**:100241
77. Yang C, Lan H, Gao F, Gao F. Review of deep learning for photoacoustic imaging. *Photoacoustics* 2021;**21**:100215