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, 62 Nanyang Drive, 637459, Singapore
Corresponding author: Junjie Yao.
Email: junjie.yao@duke.edu

Abstract

The rapidly evolving field of photoacoustic (PA) tomography (PAT) 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, makes PAT highly promising for clinical translation in functional brain imaging, early cancer detection, real-time surgical guidance, and the visualization of dynamic drug responses. Considering PAT has benefited from numerous engineering innovations, it is of no surprise that many of PAT'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 PAT 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 difficulties where traditional solutions either completely fail or make only incremental progress. This concise review focuses on the history of applied artificial intelligence in PAT, presents recent advances at this multifaceted intersection of fields, and outlines the most exciting advances that will likely propagate into promising future innovations.

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.

Keywords

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

Introduction

The hybrid imaging modality of photoacoustic tomography (PAT) combines optical excitation and ultrasonic detection to achieve an unparalleled balance of spatial resolution, penetration depth, and imaging speed[1-3]. PAT relies on the photothermal 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 dictionary learning at first, and later deep learning techniques, there now exists promising new pure-software solutions. 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 Fig. 1A)[5, 6]. 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)[7]. 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 typically determined by the effective spatial resolution. 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 defined by the effective spatial resolution. 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 $^{\Delta V}/_{V}$ can be considered negligible[8] (as shown in Eqn. 1):

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

Here, Δp and ΔT represent changes in pressure and temperature respectively; κ represents isothermal compressibility, and β denotes the thermal coefficient of volume expansion. This particular 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 (Eqn. 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.

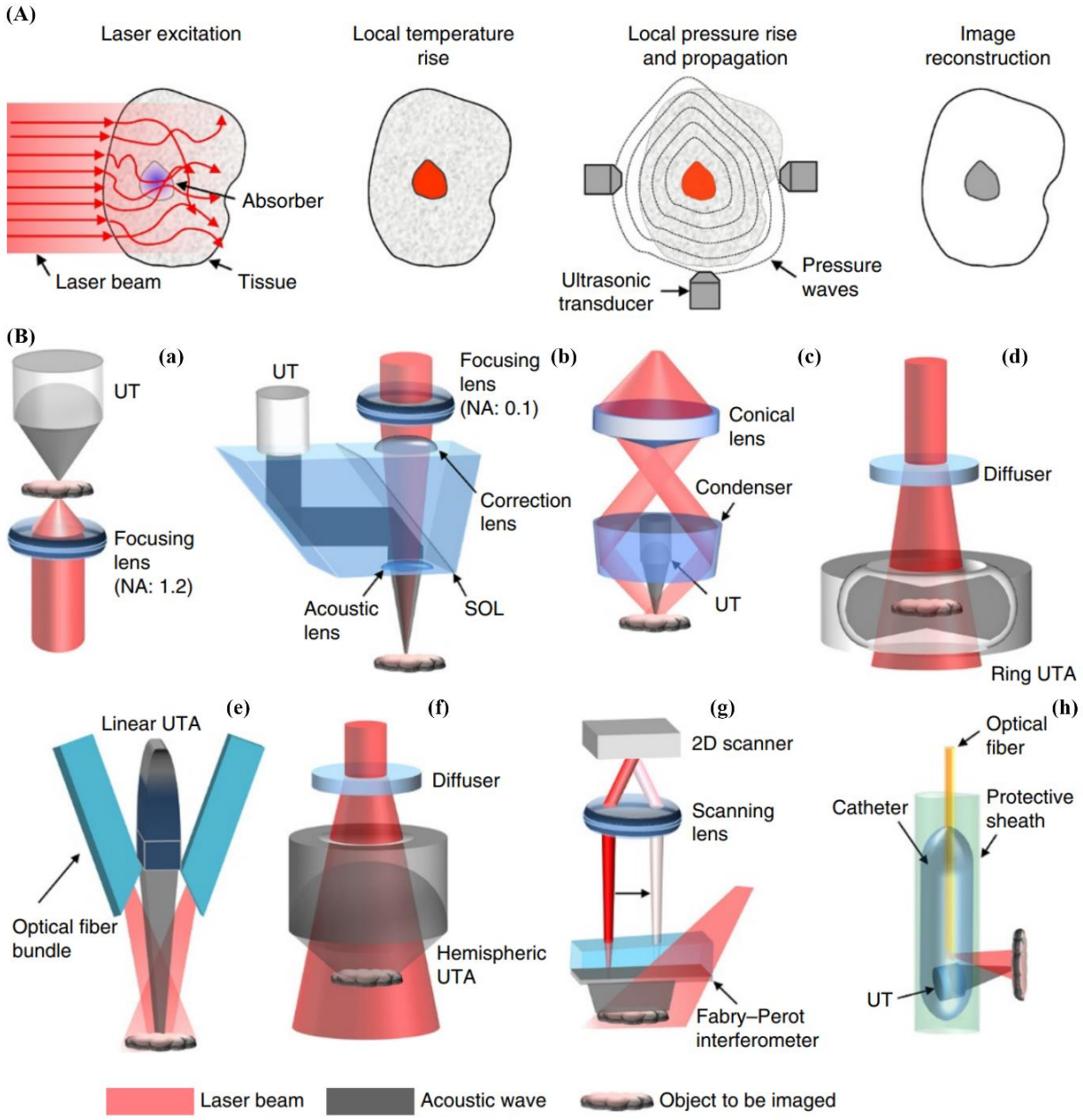

Figure 1. Principle of photoacoustic tomography (PAT). (A) The imaging process of PAT. (B) The major implementations of PAT, including (a) transmission-mode OR-PAM system, (b) reflection-mode OR-PAM system, (c) AR-PAM system with a dark-field illumination, (d) PACT system with a ring-shaped ultrasound 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-Perot interferometer as the acoustic sensor, and (h) side-viewing intravascular PA catheter. Adapted with permission from Ref. xxx. [update the reference]

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 different techniques introduces unique challenges to be overcome by hardware or software interventions. PACT forms tomographic images through diffused light illumination of tissue and parallel acoustic detection by an ultrasonic transducer array (see Fig. 1B)[7]. Similar to X-ray computed tomography, the ultrasonic transducer array captures the photoacoustic wave at different projection angles, which can then be combined 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 ultrasonic detection and not the optical excitation[1, 8].

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 Fig. 1B). The scanning method can vary significantly among PAM implementations[7]. For example, some traditional PAM systems utilize slow mechanical raster scanning, while modern PAM systems utilize high-speed optical scanning mirrors[1, 9-13]. Although all PAM systems utilize focused ultrasonic detection, some systems only use weakly-focused optical excitation – known as acoustic resolution PAM (AR-PAM) – while other systems use tightly-focused optical excitation – known as optical resolution PAM (OR-PAM)[7]. OR-PAM can provide optical-diffraction-limited resolutions within the (quasi) ballistic regime (), while AR-PAM can provide acoustic-diffraction-limited resolutions within the quasi-diffusive regime ().

Technical Challenges in PAT:

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 large fields of view with sufficient spatial resolution[14]. This has led to spatial undersampling in traditional PAM systems in order to reduce the image acquisition time, and has triggered 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[14]. In addition, the slow scanning in PAM also results in motion artifacts for dynamic imaging[15]. 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 focus [16].

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[17]. 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[18, 19]. 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[20]. Additional reconstruction artifacts can also come from inadequate frequency sampling due to limited-bandwidth ultrasound transducers[21]. In PACT, the above challenges are often present simultaneously and it is difficult to separately address their individual impacts on the final imaging performance.

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 ultrasonic arrays, and high-speed scanning mirrors, as well as time-consuming imaging reconstructions and processing, such as iterative-based methods. Moreover, these engineering solutions often have to 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[22]. Inspired by the biological structure of neurons, these networks used nodes connected by edge weights and a nonlinear activation function (Fig. 2(A))[22, 23]. 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 (Fig. 2(B))[24-26]. A subset of neural networks known as convolutional neural networks (CNN) were developed for imaging applications and take advantage of spatial neighborhood relationships[26]. CNNs shifted the focus from optimizing edge weights among various layers of interconnected nodes to optimizing layered convolutional kernel weights (Fig. 2(C)). The convolutional operation can be represented as a Toeplitz matrix multiplication, and a bias term can be incorporated into the matrix multiplication[27]. 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[25, 26, 28, 29].

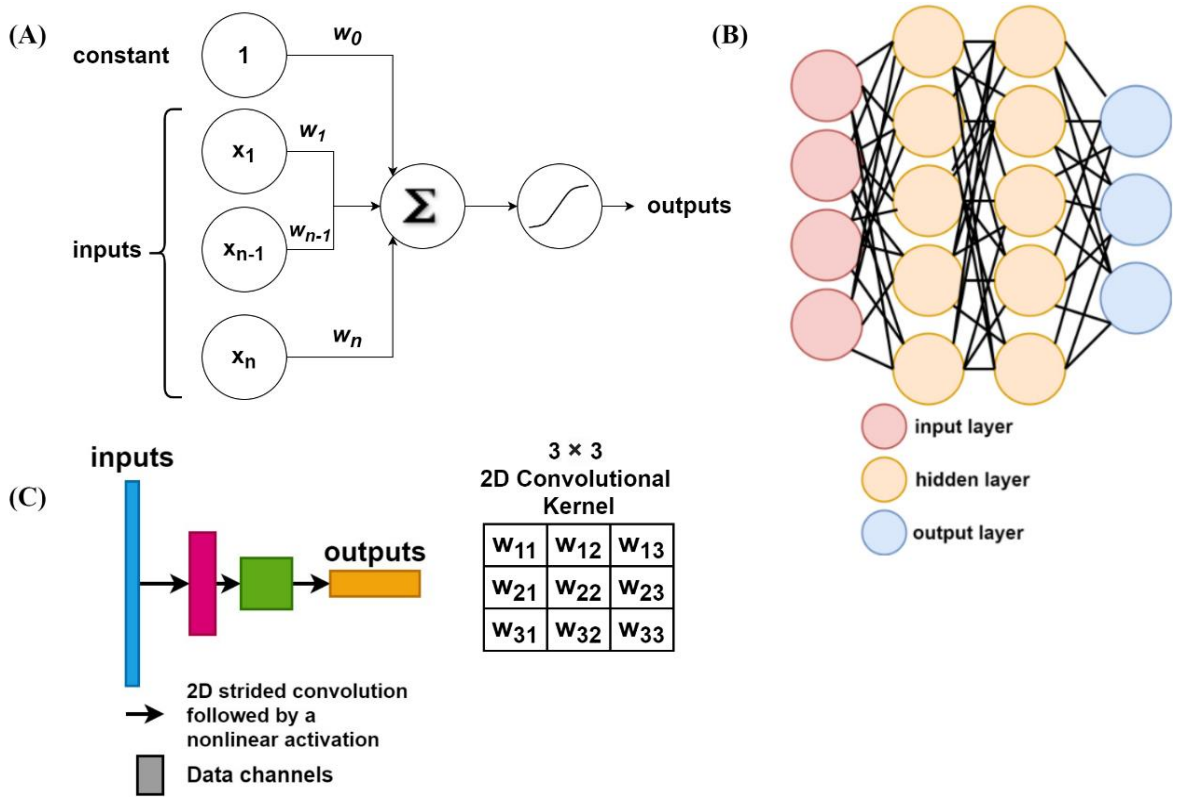

Figure 2. Depiction of representative DL concepts. The (A) perceptron, (B) multilayered neural network, and (C) a simple CNN with a 2D convolutional kernel.

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[30], 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, either replacing the traditional reconstruction altogether or being 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[28, 29, 31-36].

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[18, 37] 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 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 (Fig. 3).

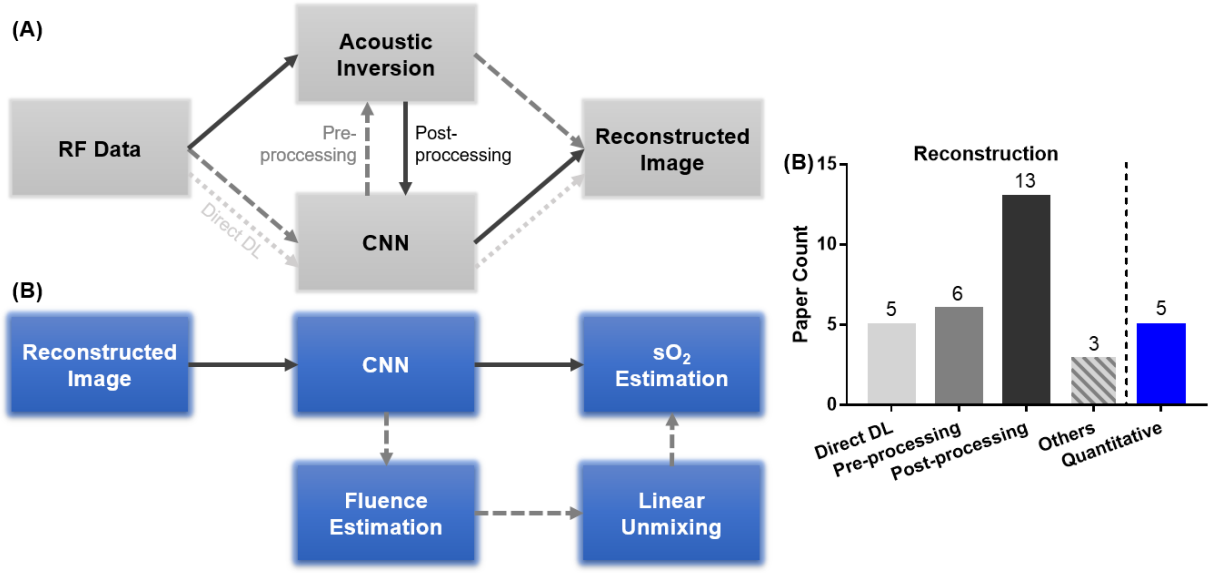

Figure 3. Deep learning strategies in PACT. (A) General pathways to apply DL to PACT reconstruction. Pre-processing CNNs correct raw data before reconstruction, while post-processing DL is applied to post-reconstruction images. CNNs can also be used to replace acoustic inversion altogether. (B) General paths to use DL for quantitative PACT, w.r.t blood oxygenation estimation. (C) Paper count on selected DL-based PACT topics.

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[38]. A CNN model was trained with simulated data and achieved similar performance with post-processing methods[38]. Nevertheless, this implementation was 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 to the 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 Fig. 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[42], which showed a slight improvement over post-processing models[42]. With similar training times, all 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[44]. 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 bandwidth of the received channel data, and thus increase the SNR of beamformed images by ~6 dB[44]. A follow-up work applied U-Net on channel data for resolution improvement and bandwidth broadening, as shown in Fig. 4(B)[45]. A similar study used a simplified super-resolution CNN for resolution enhancement and signal denoising[46]. Both approaches were validated on experimental data[45, 46]. 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[47-49]. 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 (Fig. 3(C)). Despite typically suffering from artifacts, reconstructed PACT images are 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 ultrasonic arrays[18, 50]. FD U-Net outperformed U-Net on simulated mouse brain vasculature dataset[50]. 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[51]. 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[52]. The performance of these early methods has yet to be tested on *in vivo* data.

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.* has directly utilized experimental data to train a U-Net for removing both sparse-sampling and limited-view artifacts [19]. Full-view reconstructed images were used as the ground truth to train a CNN model to improve sub-aperture reconstructed images[19]. 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 (Fig. 4(C))[19].

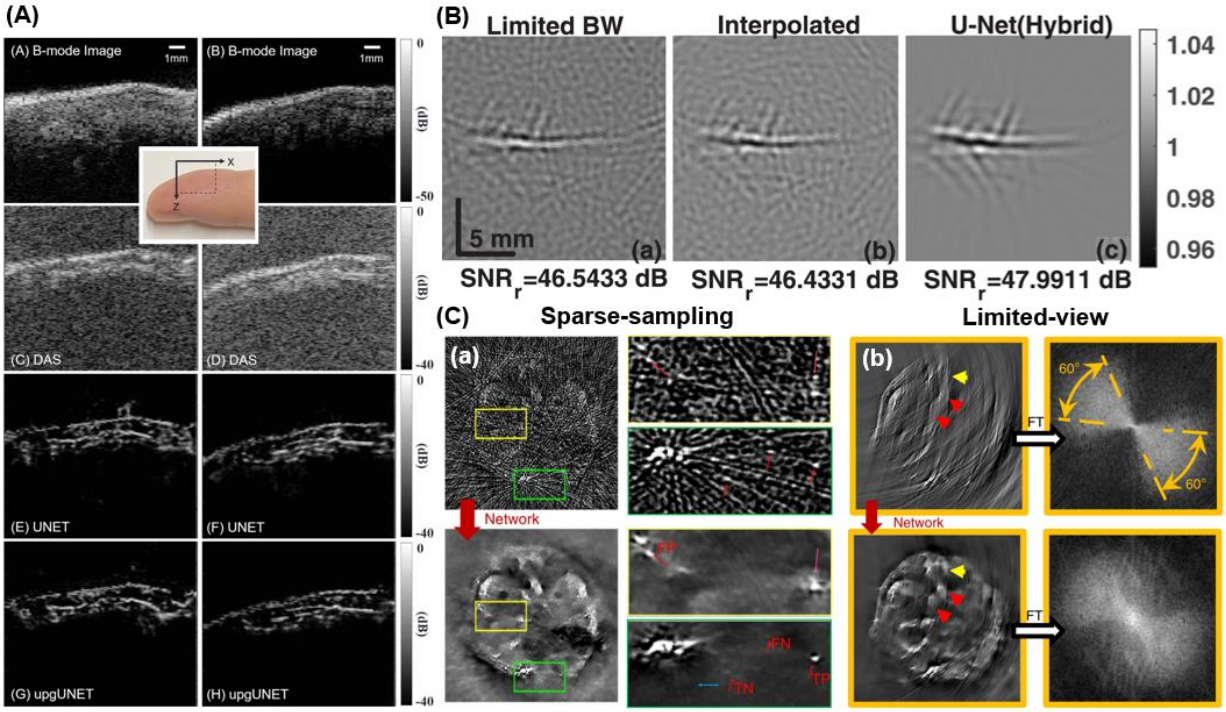

Figure 4. Representative *in vivo* DL-based PACT reconstruction. (A) Direct reconstruction of a human finger using upgUNET. (B) Reconstructed images for in vivo rat brain data with enhanced bandwidth using U-Net. (C) Sparse-sampling and limited-view artifacts of whole-body mouse images are greatly suppressed using U-Net trained by *in vivo* data. [need to cite the references from which these figures are taken in the caption as well]

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]. 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.

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 centre (or close to the transducer surface)[59]. TARES outperformed FD U-Net on both phantom and *in vivo* rat brain data, and has shown great potential for enhancing the resolution of other detection geometries – especially 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[60]. Both the original channel data and its time derivative were used as the model

input[60]. This approach outperformed standalone DL methods[60]. 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[20]. Instead, using deep gradient descent (DGD) to correct limited-view artifacts, Hauptmann *et al.* replaced TV regularization with a trained CNN[20], and demonstrated better reconstruction quality on human palm vasculature images[20]. Boink *et al.* replaced both the primal and dual domain with CNNs in their learned primal-dual method (L-PD)[61], which improved the joint reconstruction and vessel segmentations with limited-view detection[61].

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[62]. 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 (Fig. 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.

Early attempts by Cai *et al.* and Yang *et al.* explored ResU-Net and DR2U-Net with multi-wavelength reconstructed PAT images[63, 64]. The DL-reconstructed sO_2 map suggested better performance than linear unmixing[63]. Olefir *et al.* used dimensionality-reduced spectra as the input in their bi-directional RNN model, named DL-eMSOT[65]. DL-eMSOT predicted maps of eigenfluence in deep tissue[65], which were subsequently used for linear unmixing of the oxy- and

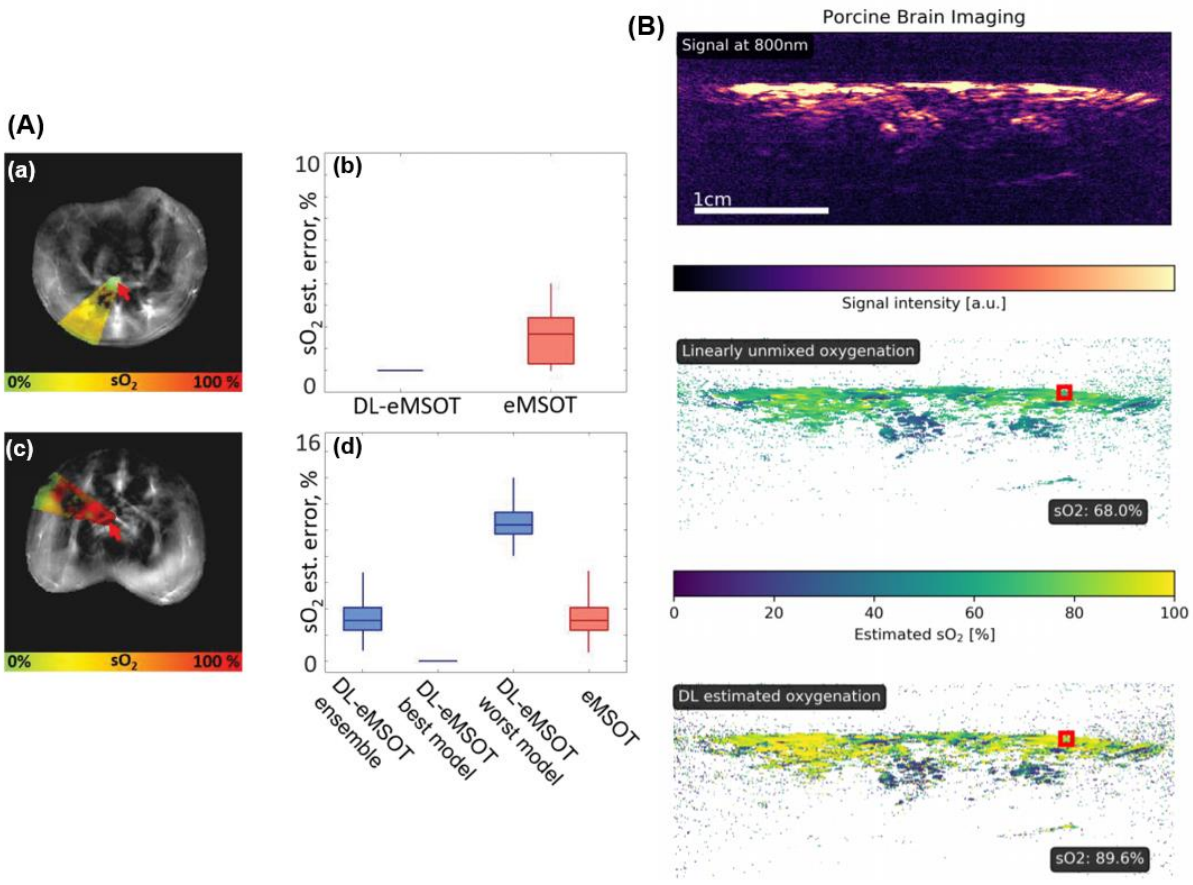

Figure 5. Representative *in vivo* DL-based quantitative PACT. (A) DL-eMSOT estimation of blood oxygenation shows significantly reduced error when compared to conventional eMSOT, on abdominal cross-sections of two mice. (B) sO_2 estimation derived from a multispectral PA image of a pig brain using LSD-qPAI, showing improved accuracy compared to linear unmixing. [need to cite the references from which these figures are taken in the caption as well]

deoxy-hemoglobin concentrations[65]. DL-eMSOT takes advantage of a sequential-learning RNN and achieved less error than the conventional eMSOT approach (Fig. 5(A))[65]. In the LSD-qPAI approach, Gröhl *et al.* applied a fully-connected neural network on multi-spectral (26 wavelengths) pressure maps[66]. This method yielded accurate sO_2 estimations on phantom experiments and *in vivo* porcine brain data (Fig. 5(B))[66]. 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[67]. This method, however, has not yet been adapted to *in vivo* data due to the complexity of tissue properties[67].

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[68]. 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 Fig. 6(C))[15].

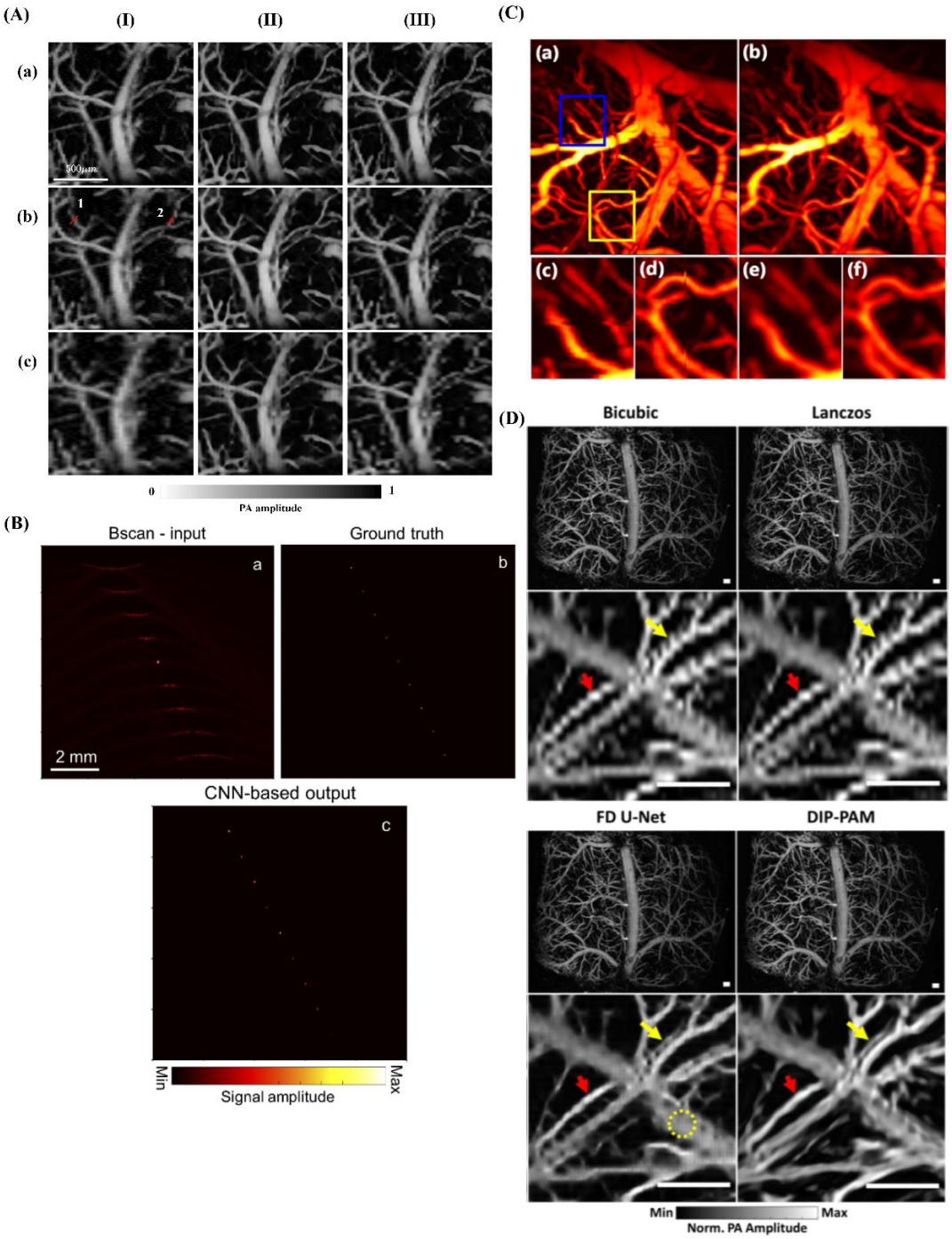

Figure 6. Representative DL methods in PAM. (A) FD U-net upsampling performance on mouse brain OR-PAM images at various downsampling ratios of 20%, 5% and 2% effective pixels, respectively. (B) DL resolution enhancement of out of focus plane signal. (C) Motion artifact removal using a CNN. (D) Comparison of Deep Prior with other upsampling methods on a mouse brain image. [need to cite the references from which these figures are taken in the caption as well]

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[69]. We trained a modified fully dense U-net architecture (FD U-net)[14]. 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 ratios (See Fig. 6(A)). 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 utilized a CNN architecture with Squeeze-and-Excitation (SE) blocks to perform a similar upsampling procedure[70].

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" (see Fig. 6(D))[71]. 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 Song *et al.* to improve the low SNR of PAM images[72]. 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 (Fig. 6(B))[16].

Discussion:

Both the fields of photoacoustic imaging and deep learning have progressed at an exceedingly accelerated rate over recent years. 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 lacks 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. The future incorporation of deep learning into photoacoustic imaging technology, especially if successful clinical adoption is to be achieved, 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.

Acknowledgments:

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

Declaration of Conflicting Interests:

The authors declared no conflicting interests.

Funding:

National Institutes of Health (R01 EB028143, R01 NS111039, RF1 NS115581, R01GM134036, 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.

References:

1. Xia J, Yao J, Wang LV. Photoacoustic tomography: principles and advances. *Electromagn Waves (Camb)* 2014;**147**:1-22
2. Upputuri PK, Pramanik M. Recent advances toward preclinical and clinical translation of photoacoustic tomography: a review. *J Biomed Opt* 2017;**22**:041006
3. Das D, Sharma A, Rajendran P, Pramanik M. Another decade of photoacoustic imaging. *Phys Med Biol* 2020 (In Press);
4. Beard P. Biomedical photoacoustic imaging. *Interface Focus* 2011;**1**:602-31
5. Bell AG. On the production and reproduction of sound by light. *American Journal of Science* 1880;**s3-20**:305-24
6. Manohar SR, Daniel PB. Photoacoustics: a historical review. *Adv Opt Photon* 2016;**8**:586-617
7. Xu M, Wang LV. Photoacoustic imaging in biomedicine. *Rev Sci Instrum* 2006;**77**:041101
8. Wang LV. Multiscale photoacoustic microscopy and computed tomography. *Nat Photonics* 2009;**3**:503-09
9. Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, Baraniuk RG. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine* 2008;**25**:83-91
10. Random-access optical-resolution photoacoustic microscopy using a digital micromirror device. *Optics Letters* 2013;**38**:2683-86
11. Haltmeier M, Berer T, Moon S, Burgholzer P. Compressed sensing and sparsity in photoacoustic tomography. *Journal of Optics* 2016;**18**:114004
12. Wang LV, Yao J. A practical guide to photoacoustic tomography in the life sciences. *Nature Methods* 2016;**13**:627-38
13. Jeon S, Kim J, Lee D, Baik JW, Kim C. Review on practical photoacoustic microscopy. *Photoacoustics* 2019;**15**
14. III AD, Li D, Vu T, Chen M, Zhang D, Luo J, Horstmeyer R, Yao J. Reconstructing Undersampled Photoacoustic Microscopy Images using Deep Learning. *IEEE Transactions on Medical Imaging* 2020:1-1
15. Chen X, Qi W, Xi L. Deep-learning-based motion-correction algorithm in optical resolution photoacoustic microscopy. *Visual Computing for Industry, Biomedicine, and Art* 2019;**2**:12
16. Sharma A, Pramanik M. Convolutional neural network for resolution enhancement and noise reduction in acoustic resolution photoacoustic microscopy. *Biomedical Optics Express* 2020;**11**:6826-39
17. Agranovsky M, Kuchment P. Uniqueness of reconstruction and an inversion procedure for thermoacoustic and photoacoustic tomography. *Inverse Problems* 2007;**23**:2089-102
18. Antholzer S, Haltmeier M, Schwab J. Deep learning for photoacoustic tomography from sparse data. *Inverse Problems in Science and Engineering* 2019;**27**:987-1005
19. Davoudi N, Deán-Ben XL, Razansky D. Deep learning optoacoustic tomography with sparse data. *Nature Machine Intelligence* 2019;**1**:453-60
20. 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 transactions on medical imaging* 2018;**37**:1382-93
21. Gutta S, Kadimesetty VS, Kalva SK, Pramanik M, Ganapathy S, Yalavarthy PK. Deep neural network-based bandwidth enhancement of photoacoustic data. *Journal on Biomedical Optics* 2017;**22**:116001
22. Wang H, Raj B. On the origin of deep learning. *arXiv preprint arXiv:170207800* 2017;
23. Schmidhuber J. Deep learning in neural networks: An overview. *Neural networks* 2015;**61**:85-117
24. Hinton G, LeCun Y, Bengio Y. Deep learning. *Nature* 2015;**521**:436-44
25. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.
26. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Van Esesn BC, Awwal AAS, Asari VK. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:180301164* 2018;
27. Bengio Y, Goodfellow I, Courville A. Deep learning. MIT press Massachusetts, USA:; 2017.
28. Razzak MI, Naz S, Zaib A. Deep Learning for Medical Image Processing: Overview, Challenges and Future. *ArXiv170406825 Cs* 2017;
29. Litjens G, al. e. A survey on deep learning in medical image analysis. *Medical Image Analysis* 2017;**42**:60-88
30. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:14126980* 2014;

31. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *Journal of Big Data* 2015;**2**:1
32. Lee J-G, al. e. Deep Learning in Medical Imaging: General Overview. *Korean Journal of Radiology* 2017;**18**:570
33. Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. *Optica* 2017;**4**:1437-43
34. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *RadioGraphics* 2017;**37**:505-15
35. Sahiner B, al. e. Deep learning in medical imaging and radiation therapy. *Medical Physics* 2019;**46**:e1-e36
36. Martorell-Marugán J, al. e. Deep Learning in Omics Data Analysis and Precision Medicine. *Computational Biology* 2019:37-53
37. Yalavarthy PK, Kalva SK, Pramanik M, Prakash J. Non-local means improves total-variation constrained photoacoustic image reconstruction. *Journal of biophotonics* 2020 (In Press):e202000191
38. 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. Photons Plus Ultrasound: Imaging and Sensing 2018: International Society for Optics and Photonics; 2018. p. 104942S.
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. *Scientific Reports* 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 Transactions on Medical Imaging* 2020;
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. Gutta S, Kadimesetty VS, Kalva SK, Pramanik M, Ganapathy S, Yalavarthy PK. Deep neural network-based bandwidth enhancement of photoacoustic data. *Journal of biomedical optics* 2017;**22**:116001
45. 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 Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 2020;
46. Awasthi N, Pardasani R, Kalva SK, Pramanik M, Yalavarthy PK. Sinogram super-resolution and denoising convolutional neural network (SRCN) for limited data photoacoustic tomography. *arXiv preprint arXiv:200106434* 2020;
47. Reiter A, Bell MAL. A machine learning approach to identifying point source locations in photoacoustic data. Photons Plus Ultrasound: Imaging and Sensing 2017: International Society for Optics and Photonics; 2017. p. 100643J.
48. Allman D, Reiter A, Bell MAL. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE transactions on medical imaging* 2018;**37**:1464-77
49. 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: Imaging and Sensing 2019: International Society for Optics and Photonics; 2019. p. 108785E.
50. Guan S, Khan AA, Sikdar S, Chitnis PV. Fully Dense UNet for 2-D Sparse Photoacoustic Tomography Artifact Removal. *IEEE journal of biomedical and health informatics* 2019;**24**:568-76
51. 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. International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer; 2019. p. 273-81.
52. Schwab J, Antholzer S, Nuster R, Paltauf G, Haltmeier M. Deep Learning of truncated singular values for limited view photoacoustic tomography. Photons Plus Ultrasound: Imaging and Sensing 2019: International Society for Optics and Photonics; 2019. p. 1087836.
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. *Experimental Biology and Medicine* 2020:1535370220914285
54. Godefroy G, Arnal B, Bossy E. Compensating for visibility artefacts in photoacoustic imaging with a deep learning approach providing prediction uncertainties. *Photoacoustics* 2020:100218

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. *Computerized Medical Imaging and Graphics* 2020;**84**:101720
56. Lu T, Chen T, Gao F, Sun B, Ntziachristos V, Li J. LV-GAN: A deep learning approach for limited-view optoacoustic imaging based on hybrid datasets. *Journal of Biophotonics* 2020:e202000325
57. Singh MKA, Sivasubramanian K, Sato N, Ichihashi F, Sankai Y, Xing L. Deep learning-enhanced LED-based photoacoustic imaging. Photons Plus Ultrasound: Imaging and Sensing 2020: International Society for Optics and Photonics; 2020. p. 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. Rajendran P, Pramanik M. Deep learning approach to improve tangential resolution in photoacoustic tomography. *Biomedical Optics Express* 2020;**11**:7311-23
60. 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* 2020:100190
61. Boink YE, Manohar S, Brune C. A Partially-Learned Algorithm for Joint Photo-acoustic Reconstruction and Segmentation. *IEEE transactions on medical imaging* 2019;**39**:129-39
62. Yang C, Gao F. Eda-net: Dense aggregation of deep and shallow information achieves quantitative photoacoustic blood oxygenation imaging deep in human breast. International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer; 2019. p. 246-54.
63. Cai C, Deng K, Ma C, Luo J. End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging. *Optics letters* 2018;**43**:2752-55
64. Yang C, Lan H, Zhong H, Gao F. Quantitative photoacoustic blood oxygenation imaging using deep residual and recurrent neural network. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019): IEEE; 2019. p. 741-44.
65. 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 transactions on medical imaging* 2020;**39**:3643-54
66. Gröhl J, Kirchner T, Adler T, Maier-Hein L. Estimation of blood oxygenation with learned spectral decoloring for quantitative photoacoustic imaging (LSD-qPAI). *arXiv preprint arXiv:190205839* 2019;
67. Bench C, Hauptmann A, Cox BT. Toward accurate quantitative photoacoustic imaging: learning vascular blood oxygen saturation in three dimensions. *Journal of Biomedical Optics* 2020;**25**:085003
68. 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. *Scientific Reports* 2018;**8**:985
69. Anthony DI. Duke PAM Dataset. Zenodo; 2020.
70. Zhou J, He D, Shang X, Guo Z, Chen S-l, Luo J. Photoacoustic Microscopy with Sparse Data Enabled by Convolutional Neural Networks for Fast Imaging. *arXiv preprint arXiv:200604368* 2020;
71. Vu T, DiSpirito III A, Li D, Zhang Z, Zhu X, Chen M, Zhang D, Luo J, Zhang YS, Horstmeyer R. Deep Image Prior for Sparse-sampling Photoacoustic Microscopy. *arXiv preprint arXiv:201012041* 2020;
72. Song X, Tang K, Wei J, Song L. Denoising convolutional neural networks for photoacoustic microscopy. *arXiv preprint arXiv:200913913* 2020;