Deep learning for photoacoustic imaging: a survey

Changchun Yang†, Hengrong Lan‡, Feng Gao, and Fei Gao*  

Abstract— Machine learning has been developed dramatically and witnessed a lot of applications in various fields over the past few years. This boom originated in 2009, when a new model emerged, that is, the deep artificial neural network, which began to surpass other established mature models on some important benchmarks. Later, it was widely used in academia and industry. Ranging from image analysis to natural language processing, it fully exerted its magic and now become the state-of-the-art machine learning models. Deep neural networks have great potential in medical imaging technology, medical data analysis, medical diagnosis and other healthcare issues, and is promoted in both pre-clinical and even clinical stages. In this review, we performed an overview of some new developments and challenges in the application of machine learning to medical image analysis, with a special focus on deep learning in photoacoustic imaging.

The aim of this review is threefold: (i) introducing deep learning with some important basics, (ii) reviewing recent works that apply deep learning in the entire ecological chain of photoacoustic imaging, from image reconstruction to disease diagnosis, (iii) providing some open source materials and other resources for researchers interested in applying deep learning to photoacoustic imaging.

I. INTRODUCTION

The advent of the era of big data, the increase in computing power following Moore's Law, and open-source, user-friendly software frameworks have made remarkable progress in machine learning, which has aroused great interest in both industry and academia. Data-driven artificial neural networks, also known as deep learning (DL), are good at discovering complex patterns from massive data to determine the optimal solution in the parameter space. The growing computing power of GPUs allows neural networks to be flexibly expanded from depth [7, 8], width [9], and cardinality [10], forming numerous cornerstone-like models, which have also become the state-of-the-art (SOTA) approaches for a series of problems, exerting its magical power in computer vision, natural language processing and robotics.

Among the many image analysis benchmarks, deep learning rapidly surpasses the traditional machine learning technology, making it reach the most important position in computer vision. At the 2012 ImageNet Large-Scale Visual Recognition Challenge (ILSVRC, [11]), a convolutional neural network, AlexNet, easily exceeded the support vector machine (SVM) to win the championship with a disparity of more than 10%. That is, since this session, deep learning has replaced SVM, and various innovative neural network models have sprung up. The Top-5 error rate used to measure performance has also reached new lows under fierce competition, gradually surpassing human eye recognition (5~10%), and even close to the Bayes error rate. In addition to shining in computer vision, deep learning also shows its advantages in areas such as natural language processing [12, 13], speech recognition [14, 15]†, and even the field of physics‡.

By hybrid combination of high contrast of optical imaging and high penetration depth of ultrasound imaging, photoacoustic (PA) imaging, one of the non-invasive medical imaging methods [16-19], has undoubtedly become an emerging application of deep learning. Therefore, we will review recent research of PA imaging with deep learning.

II. DEEP LEARNING

It is the core of machine learning to develop some algorithms to make computers capable of learning experience from data to solve tasks. That is to create a mathematical model f, which can produce the desired output through training when fed with input data. Data representation of the training set can provide experiences for machine learning models, which are then tuned to produce accurate predictions according to the difference between current predicted output and the expected by an optimization algorithm. The feedback obtained by training the model on another separate data set, the validation set, is used for further fine-tuning to measure the generalization ability of the model. After iterating through these two steps, the model is finally evaluated on the test set to evaluate the performance when it encounters new, unseen

† Those authors contributed equally to this work.
‡ Modern deep learning methods have also achieved some applications in the field of physics, and these methods are often highly transferable across domains, such as cleaning of interference images in astrophysics [20] and denoising in gravity wave analysis [21], which use a combination of mathematical models and machine learning models to learn the relationship of physics from raw data. These methods can also be used in medical image analysis [16, 22].
Fig. 1. We take the most common artificial neural network, multilayer perceptron or feedforward neural network as an example. The output of the j-th unit at layer i-th is \( y_j^i = \sigma_j^i(x) \), where \( x \) is the value of the output of the previous layer after the nonlinear function transformation called activation function. Typical activation functions are sigmoid function \( \sigma(z) = \frac{1}{1 + e^{-z}} \) and rectified linear unit \( \text{ReLU}(z) = \max(0, z) \).

There are also many types of activation functions, which one to use depends on the task and architecture. Similarly, the output of each layer is passed to the next layer after the activation function, and then the same calculation is performed until the network produces the final output. The training data is fed to the input layer and then the streaming calculation is performed, where the output and derivative of each node are recorded. Finally, the difference between the prediction and the label at the output layer is measured by the loss function. The choice of loss function has a critical impact on the performance of the entire task, so its choice is crucial, common ones include mean absolute error, mean squared error. You can also manually design the loss function, which is more flexible. The derivative of the loss function will be used as a feedback signal, and then propagate backward through the network, and the weight of the network will be updated to reduce errors. This is a technique called backward propagation [4, 5], which follows the chain rule to calculate the gradient of the objective function with respect to the weights in each node. Then gradient descent [6] is used to update all the weights.

Regarding machine learning, especially deep learning, many excellent reviews and surveys have also emerged, you can check the relevant papers [26] for a short introduction, or read free available book [27] for an in-depth understanding. See also [28] for a broad understanding of its application to medical image analysis. Only some basic essentials of deep learning will be introduced in this section, serving as useful foundations for its application in PA imaging.

A. Artificial neural networks

Firstly we will introduce one of the most famous machine learning models that appeared in the 1950s, the artificial neural network, which is used to briefly explain the deep learning principle. A neural network is composed of many layers, which are connected by many computing units, also known as neurons. Data enters the neural network through the input layer, then flows through one or more hidden layers, and finally generates the prediction of the neural network in the output layer, which will be compared with the actual labels (ground truth) by an objective function (loss/cost function). This difference guides to change the weights of the network to optimize the objective function until the neural network produces a good prediction. This means that the neural network learns experiences from the data and can generalize to new data sets for prediction.

Next, we will briefly explain how the artificial neural networks are constructed. Since the neural network can be very complex and deep, we only introduce its basic form shown in Fig. 1. Map input \( x \) to output \( y \) through parameterized function \( y = f(x; \theta) \), which consists of a lot of nonlinear changes \( f(x) = (f_1 \circ \cdots \circ f_k) (x) \). Each component/layer \( f_k \) can be represented as \( f_k = \sigma_k (\theta_k f_{k-1}) \), followed by a simple linear transformation and a nonlinear function. \( \theta_k \) is the model’s weights, and the nonlinear functions are differentiable, typically are sigmoid function and ReLUs. The basic idea of training a neural network is simple: the forward order of the training data flows into the network, and the gradient of the loss function with respect to every weight is computed using the chain rule, finally gradient descent (GD) is used to change these weights to make the loss smaller. The reason why such a simple but effective idea did not start explosive growth until recent decade is because that when faced with the exponential growth of parameters and nodes, the calculation will be a huge challenge. Therefore, many techniques such as the design and training of neural networks are open and interesting problems.

3 You can find the free online version: https://www.deeplearningbook.org/.
4 These are basic but essential concepts and components, compared to recurrent neural networks (RNNs), generate adversarial networks (GANs), graph convolutional networks (GCNs), etc.
5 As we shall see, the modern architecture is significantly more complicated with multiple outputs, the input connected with the output, multiple branches, multiple loss functions, and much more. How to design a novel architecture that can solve specific task is the core.
**B. Deep learning**

Traditional machine learning models will extract features from the data manually or with the help of other simple machine learning models before training. However, deep learning will automatically learn representations and features from the data during the training process, without the need for manual design. The main common feature of various deep learning models and their variants is that they all focus on feature learning: automatic learning of data representation, which is also the main difference between modern deep learning and classic machine learning methods. Learning features and performing tasks are improved simultaneously in the model. You can get a comprehensive understanding through the reference [26, 27], or have a detailed study of deep learning with an interactive book [29] with code using the most popular deep learning framework [30] in academia today.

Deep learning that has made rapid progress in medical imaging is currently focused on convolutional neural networks [5] (CNNs). This benefits from the fact that CNNs can easily learn specific features from images or other structured data. Next, we will briefly introduce the essential components of CNNs, understand why it is powerful and have an insight to design your own architecture.

**C. Building CNNs**

Although the introduced artificial neural network can be directly applied to the image, the efficiency of connecting each node of all layers to all nodes of the next layer is very low. Fortunately, structured data such as images can be trimmed and connected based on domain knowledge. CNNs are neural networks that can preserve the spatial relationship between data with few connections. The training process of a CNN is exactly the same as ANNs, except that the structure is often convolutional layers accompanied by activation functions, and pooling layers (as shown in Fig. 2).

(i) **Convolutional layers.** In the convolutional layer, the activations from the previous layer performs a convolution operation with a parameterized filter. Each filter shares weights over the entire scope, which not only can greatly reduce the number of network parameters, but also has the translational equivariance. The reason why weight sharing can work is because features that appear in one part of the image may also appear in other parts. For example, this parameterized filter can detect the horizontal line above the image after training to determine the weight, then it can still detect the horizontal line below. After the convolution operation of the convolution layers, a tensor of feature maps will be generated.

(ii) **Activation layers.** Similar to the introduction in ANNs, the activation layer is often composed of nonlinear activation functions. Because of the existence of these nonlinear activation functions, combined with linear operations such as convolution, the neural network can almost approximate any nonlinear function [31, 32]. See [33, 34] to learn more about the role of activation function. The activation layers also generate new tensors of feature maps.

(iii) **Pooling.** In neural networks, feature maps are often pooled in the pooling layer. The pooling operation generates a number for each small grid of the input feature maps, which is often achieved by max-pooling or average pooling. The significance of the pooling layer is to give the neural network translational invariance, since a small shift in input will result in changes in the activation maps. For example, to determine

---

6 Free available link: https://d2l.ai/.
7 Deep learning frameworks undoubtedly facilitate the process of machine learning practitioners turning ideas into code. The well-known ones are: tensorflow, pytorch, MXNet, and caffe2. The first two are currently the most commonly used frameworks in industry and academia.
8 CNN has been used for medical image analysis in the early 1990s, but it was limited by the data and computing power at that time.
9 It can be compared to the filter in digital image processing, the “convolution” operation here is strictly the correlation operation, but called the “convolution” neural network is also appropriate.
10 No matter how deep the linear operation is, it can still only approximate the linear function without these nonlinear activation functions.
whether there is red in the image, translation/rotation will not affect the judgment result. Recent research [35] also proves another alternative to pooling: using convolutional layers with larger stride lengths instead, which can simplify the network structure without reducing performance. Fully convolutional networks (FCNs) [36] are also favored by more and more image analysis tasks.

(iv) **Dropout.** A very simple idea has greatly improved the performance of CNNs. Dropout [37] is an averaging technique based on stochastic sampling of neurons to prevent CNNs from overfitting. By randomly removing neurons during the training process, each batch of data uses a slightly changing network, and the weights of the network will be tuned based on the optimization of multiple variations of the network.

(v) **Batch normalization.** Batch normalization (BN) [38] is an effective technique to accelerate deep networks training by reducing internal covariate shift. Due to changes in network parameters during training, the distribution of network layer output results is different, making network training difficult. BN can produce normalized feature maps by subtracting the mean and dividing by the standard deviation for each training batch. The data will be periodically changed to zero mean and unit standard deviation with BN, which will greatly speed up the training.

In the actual design and improvement of the CNN architecture, above basic components will be combined in a very complex way, with some new and effective operations. When building a CNN for a task, you often have to consider a lot of details to get your CNN perform well on this task. You need to fully understand the task to be solved and find insights to decide how to process the data set before fed to the network.

In the early days of deep learning, the construction of modules was often simple. But then more and more complicated structures emerged, and people designed new effective architectures based on previous ideas and insights resulting in updates to the SOTA. See [39] for a survey about recent architectures of CNNs.

![Fig. 3. Number of papers of DL in PA imaging in recent years. For 2020 year’s data, only the quantity before June 30 is calculated.](image)

**Table 1.** Number of papers of DL in PA based on task.

| Task                        | Number |
|-----------------------------|--------|
| Image reconstruction        | 38     |
| Quantitative imaging        | 10     |
| Image detection             | 5      |
| Image classification        | 2      |
| Image segmentation          | 4      |
| PA-assisted intervention     | 2      |

III. **DEEP LEARNING FOR PHOTOACOUSTIC IMAGING**

Deep learning has been widely used in clinics to help doctors get a better diagnosis, and successful examples are growing. Because there are too many applications of deep learning in medical imaging of other modalities, we cannot give a comprehensive overview here, but only focus on deep learning in PA imaging. If you are interested, you can see [1, 40-42] for a more thorough review and overview of deep learning in medical imaging, because successful examples often involve multiple organs, and have different ideas and technical details, which is an extremely rich, intersecting, and interesting subject.

Although PA imaging is a new imaging method compared with other modalities of medical imaging, deep learning still shows great application prospects. More specifically, deep learning has been applied at every step of the entire PA imaging workflow. How to obtain high-quality PA image from the sensor data is relevant to the physics of PA. Disease segmentation, classification, and detection with the reconstructed PA images are relevant to the image domain. Providing functional imaging capability without exogenous contrast: i.e. quantitative imaging of oxygen saturation, is also a unique advantage of PA imaging compared to other imaging modalities. PA-assisted intervention is also involved for pre-clinical or even clinical applications. Based on our statistical analysis of the related literature, the number of papers versus year is shown in Fig. 3. It can be seen that the number on DL used in PA imaging has increased significantly every year. And the number classified by category is given in Table 1. Next, we will review in detail from the perspective of each task.

A. **PA Image Reconstruction**

Image reconstruction is one of the fundamental components for photoacoustic tomography (PAT), which converts raw signals received by ultrasound transducers to the image of initial pressure distribution. It is a challenging task for PA image reconstruction because of the ill-posed nature and the absence of exact inverse model in practical cases (limited-view and sparse sampling). In PAT, image reconstruction aims to retrieve the initial PA pressure distribution, which indicates the optical absorption of

---

11 There is no unified conclusion about whether the position of BN is placed after convolution before activation, or after activation. The original paper [38] put it before activation.

12 When counting according to the task category, two papers are about joint task, hence the calculation is repeated twice.
biological tissues. The PA wave propagation at the position \( \mathbf{r} \) and time \( t \) has the pressure \( p(\mathbf{r},t) \) given by [43]:

\[
\nabla^2 \frac{1}{v_s^2} \frac{\partial^2}{\partial t^2} p(\mathbf{r},t) = \beta \frac{\partial^2 T(\mathbf{r},t)}{\partial t^2} \kappa \frac{\partial^2}{\partial t^2} \nu \tag{1}
\]

where \( v_s \) denotes the acoustic velocity (m/s), \( \kappa \) denotes the isothermal compressibility (Pa\(^{-1}\)), \( \beta \) denotes the thermal coefficient of volume expansion (K\(^{-1}\)), and \( T \) represents the temperature rise. The transducer array receives PA signals excited by a short-pulsed laser light at different locations, which are used to form the PA pressure when \( t=0 \) by several methods, such as universal back-projection (UBP). We can formulate the UBP reconstruction based on [44]:

\[
0 \int_{\Omega_0} \frac{1}{\Omega_0} \int_{\Omega_0} \times 2 \left[ p(\mathbf{r},t) - \frac{\partial\rho(\mathbf{r},t)}{\partial t} \right]_{\mathbf{r}} \tag{2}
\]

where \( \Omega_0 \) is the solid angle of the entire detection surface \( S \) with respect to a source point at \( \mathbf{r} \). The term of square brackets is the back-projection term. These direct reconstructed algorithms can retrieve satisfactory image if the transducers cover enough angle with sufficient element number to detect the PA signals. However, an ill-posed inverse problem is caused by limited-view or sparse view. Meanwhile, the secondary generated artifacts confuse final PA image.

Model-based reconstruction algorithms show flexibility to be applicable in many unfeasible configurations for direct methods. The inverse problem can be formulated by solving the non-negative least-squares problem:

\[
p_0 = \arg \min p_0 \|4p_0 - y\| + \alpha R(p_0) \tag{3}
\]

where \( p_0 \) is initial PA pressure, \( y \) is the received PA data from transducers, \( A \) is the PA forward model, \( R(p_0) \) is regularization term, \( \alpha \) decides the properties between data fit and the regularization term. This problem can be solved by traditional optimization, such as iterative soft-thresholding algorithm (ISTA) [45], GD with repeated iteration. However, these methods are suffering computational complexity and tedious parameters adjustment for different reconstruction tasks. Deep learning methods have found applicability for the reconstruction of PA images, which are driven by a large number of datasets for training. We will review DL-based PA image reconstruction algorithms of two types.

(i) Non-iterative reconstruction. Fig. 4 show three different schemes for non-iterative reconstructions: direct estimation, PA signals and PA image enhancements via deep learning. Direct reconstruction methods only solve the PA wave equation, which take the PA signals as input and capture the mapping from signal to image.

In [46], Dominik Waibel established a direct estimation from raw sensor data for PA imaging. 128-element linear detector’s synthetic data were taken into a modified U-net to reconstruct the final initial PA pressure. Similarly, Emran Mohammad Abu Anas [47] proposed a dense CNN architecture for beamforming PA data, which consists of five dense blocks with different dilated convolutions. In this paper, the authors investigated the effect of varying speed of sound for proposed method, and verified the robustness on the speed of sound variation. In PAI, the visibility of deeper object can be affected with the optical scattering, which is a non-negligible issue for the precise depth localization applications. To address this issue, Kerrick Johnstonbaugh [48]
designed an encoder-decoder network to predict the location of circular target in deep tissue. This work considered both acoustic and optical attenuation into simulation, and it further closed to the reality. Derek Allman [49] used VGG16 to beamform the raw data to detect the point sources, and removed reflection artifact. A real experiment was also demonstrated in this work. Up to now, all of experiments used synthetic simple (point-like) phantoms. In [50], Johannes Schwab used deep learning to learn the weights in back projection of different channel’s data, and the authors used the Shepp-Logan phantom with random enhancement to verify their method. Furthermore, vessel phantom was first used to train neural network in direct reconstruction. He also proposed a data-driven regularization method [51] by applying truncated singular value decomposition (SVD) and then recovering the truncated SVD coefficients, which significantly suppressed noise. Hengrong Lan [52] proposed DU-net, which took three different frequencies’ sensor data (2.25MHz, 5MHz and 7.5MHz) as input. In short, DU-net consisted of two U-net, and an additional loss is used to constrain the first U-net. Obviously, the lack of vessel texture structure for raw data, compared with simple phantom, caused a worse result. In [2], Steven Guan also used vessel phantom to compare different reconstruction schemes, and proposed a new type of input in Fig. 4 (from c to e). Tong Tong [53] used in vivo data to train the FPNNet with a U-net as post-processing. Two types of PA signals (time derivative and normalized original data) are fed into a number of Resblock, and implemented signal-to-image reconstruction by learning below transformation:

\[ H_F = F \left( c_1 \left( c_2 p(d, t) + c_1 \frac{\partial p(d, t)}{\partial t} \right) \right) \] (4)

In addition, the authors released their code and data in this study. In summary, these direct estimation models capture priori knowledge that being closely related to conventional procedures or not, as shown in Fig. 4 from a to e. On the other hands, PA signal and image enhancement can be performed in the data domain as preprocessing (from a to d in Fig. 4) or in the image domain as postprocessing (from b to e in Fig. 4) instead of solving the PA wave equation. For this scheme, a signal processing or image processing problem usually exists, which is also important for the imaging result of photoacoustic microscopy (PAM).

The blurring image may be caused by bandwidth-limited/noise-polluted/spatial-sparse PA signals, even though using a refined reconstruction algorithm. Therefore, PA signal enhancement could obviously improve the PA image quality. Sreedevi Gutta [54] proposed the first neural network to enhance the bandwidth of PA data. A five-layers neural network took one channel bandwidth-limited PA signal as input and predicted full bandwidth signal. Navchetan Awasthi used CNN to de-noise and super resolve (from 50 to 100 detectors) the PA sinogram data in [55], which solve the sparse and low-SNR problem in signal domain. The experiments contain both simulation and in-vivo experimental cases. Furthermore, they improve the single channel enhanced scheme in [54] and used U-net to de-noise and enhance the bandwidth limited sonogram data [56]. It resolved two issues, and improved the performance compared with previous works. PA signal preprocessing guarantees the quality of reconstruction by compensating the signal pollution in data capture procedure. Spatial-sampling caused issue (limited-view) perhaps should be further explored in image domain.

Stephan Antholzer [57] performed sparsely reconstructed image with residual connected U-net. In [46], the authors also used a U-net to estimate the initial pressure from DAS result images. Likewise, Johannes Schwab [58] used a CNN to reconstruct better PA image by feeding 64 detectors and half-view PA image. Light emitting diode (LED) based PA imaging system suffers from low SNR signal due to limited output power. Therefore, the result of LED-based imaging usually has a worse quality. In [59, 60], Emran Mohammad Abu Anas proposed two different architectures, recurrent neural network based and CNN (Dense block) based approaches, to enhanced the SNR of input low quality PA image. The blood mimicking phantoms’ data were collected in these works, and generated ground truth by averaging data. After that, Mithun Kuniyil Ajith Singh [61] used higher laser energy to obtain the ground truth, which are used to train a U-net to enhance the lower SNR image from LED-based imaging system. Ref. [62] also used modified U-net to reconstruct PA image, which divided the model into three parts: feature extraction, artifacts reduction (U-net with residual skipped connection), and reconstruction. Davoudi used U-net to recover the quality of sparse PA image [3], both of simulation and realistic data were trained and tested in this work. Moreover, these data and codes are open access. Seungwon Jeon tried to use U-net based CNN to correct the speed of sound aberration in PA image in [63]. In [64] Steven Guan proposed FD-U-Net to remove artifacts caused by sparse data. Meanwhile, Tri Vu [65] designed WGAN-GP to decrease the artifacts in PACT, with both phantom and in vivo results validated. Parastoo Farnia [66] post-processed a TR result image as input of CNN.

In PAM, deep learning also plays an important role for image quality correction. In [67], a simple CNN was used to correct motion artifacts in OR-PAM. Israr Ul Haq [68] used a convolutional autoencoder to enhance the quality of image, and the ground truth came from Gobor filter. In [69], Anthony DiSpirito III released a set of mouse brain data, which were used to train a FD U-net to improve the under-sampled PAM images. Simultaneously, Jiasheng Zhou also did a similar work using Squeeze-and-Excitation (SE) block to extract information and verified its effectiveness [70].

In addition, some recent studies open a new scheme neither direct estimation nor pre/postprocessing. Hengrong Lan [71,
and on the left is the reconstructed 2D PA,
is the 2D coordinate of the tissue. The basic equation
is wavelength dependent in real scenarios, so it are the wavelength-dependent molar absorbance. Some traditional methods (e.g. linear unmixing) assumes a constant when solving sO2. In [73], the authors used intermediate results (pixel-interpolated data) as input of model (from d to e in Fig. 4), where pixel-DL method is proposed. MinWoo Kim pre-delayed every channel’s data and got 3D transformed data as input [73]. The improvement of these results is significant using vessel data. The methods that combine the information of different domains and the physical process could be a new frontier of image reconstruction.

(ii) Iterative reconstruction. Iterative reconstruction resolves image reconstruction as an inverse problem. In [74], Stephan Antholzer trained a regularization to optimize the compressed sensing PAT reconstruction procedure, and also compared with L1-minimization in [75]. Andreas Hauptmann [25, 76] took a well-known optimization (GD) and corrected the internal optimization by deep learning. The realistic experimental results showed the robustness and superiority of this scheme. Furthermore, Yoeri E. Boink [77] proposed learned primal-dual (L-PD) realizing multi-task (reconstruction and segmentation) in single architecture, which took primal-dual hybrid gradient (PDHG) as a basic optimization. In [16], Changchun Yang used the Recurrent Inference Machines to learn and accelerate the optimization procedure. In other work [78], Hongming Shan simultaneously reconstructed the initial pressure and sound speed distribution, where SR-net was proposed by fusing the gradients of every iteration with previous pressure and sound speed distribution. Navchetan Awasthi proposed PA-Fuse model, which combined two images (BP reconstructed image and model-based reconstructed image) as a fused image [79].

B. Quantitative imaging

PAT plays an increasingly important role in both preclinical research and clinical practice. In view of the fact that hemoglobin is one of the major absorbers in human tissue at wavelengths below 1000 nm, PAT can image vascular structure and oxygen saturation of hemoglobin (sO2) by quantifying oxygenated hemoglobin (HbO2) and deoxygenated hemoglobin (HbR). Oxygen saturation is a very important physiological parameter of the human body and can be used to predict the presence of tumors, since the concentration of sO2 of normal tissues is often higher than that of malignant tissues. See [80, 81] for the detailed background of blood oxygenation.

Blood oxygenation, or sO2, is defined as the fraction of HbO2 relative to total hemoglobin concentration in blood:

\[ sO2(x, y) = \frac{C_{HbO2}(x, y)}{C_{HbO2}(x, y) + C_{HbR}(x, y)} \times 100\% \quad (5) \]

where \( x \) and \( y \) is the 2D coordinate of the tissue. The basic reason why we can use PAT for quantitative blood oxygenation imaging is the distinct absorption of HbO2 and HbR at different wavelengths:

\[ P(\lambda, x, y) = \Phi(\lambda_i)(\varepsilon_{\text{HbR}}(\lambda_i)C_{\text{HbR}}(x, y) + \varepsilon_{\text{HbO2}}(\lambda_i)C_{\text{HbO2}}(x, y)) \quad (6) \]

where \( P(\lambda_i, x, y) \) is the reconstructed 2D PA image at a specific wavelength \( \lambda_i \), \( \Phi(\lambda_i) \) is wavelength-dependent optical fluence, which is affected by heterogeneous tissue’s optical properties and light propagation. \( \varepsilon_{\text{HbR}}(\lambda_i) \) and \( \varepsilon_{\text{HbO2}}(\lambda_i) \) are the wavelength-dependent molar extinction coefficients (cm\(^{-1}\)M\(^{-1}\)) of HbR and HbO2, which can be obtained from [82] or recorded information\(^{14}\). And the concentrations of HbR and HbO2 are denoted by \( C_{\text{HbR}}(x, y) \) and \( C_{\text{HbO2}}(x, y) \). Some traditional methods (e.g. linear unmixing) assumes \( \Phi(\lambda_i) \) as a constant when solving sO2. In this case, at least two wavelengths are needed to solve the equations, achieving the quantification of sO2. However, \( \Phi(\lambda_i) \) is wavelength dependent in real scenarios, so it actually confounds the spectral unmixing of HbO2 and HbR. This leads to the fact that linear unmixing cannot accurately give quantitative results (e.g. shown in Fig. 5 (d)). Solving sO2 corresponding to spatial coordinates from reconstructed PA images of multiple wavelengths is a typical ill-posed inverse problem, for which deep learning is undoubtedly a good choice\(^{15}\).

Chuangjian Cai [83] proposed the first deep learning framework, ResU-net, for quantitative PA imaging. Their model takes the initial pressure images at different wavelengths as inputs, and the outputs are quantitative results.

\(^{14}\) See https://omlc.org/software/mc/mcxyz/index.html for a detailed tissue properties in the wavelength range of 300–1000 nm.

\(^{15}\) In fact, deep learning can be used to solve a wide variety of inverse problems arising in computational imaging [95, 96].
of sO$_2$. They use the CNN architecture implemented by U-net to achieve the following functions: (1) determining the contour of the imaged object, (2) performing optical inversion, (3) denoising, (4) determining the type of main absorbing chromophores and estimating their absorption spectrum, (5) unmixing the HbO$_2$ and HbR. Therefore, the author added a residual learning mechanism [84] to solve the gradient explosion problem on the basis of U-net composed of a contraction path that captures comprehensive context and an extension path that realizes precise localization, so that the network can enrich features through deep layers stacking and simultaneously, greatly increase the quantitative accuracy. The author conducted a simulation experiment to simulate a simple numerical phantom based on 21 wavelengths (700-800 nm, step size 5nm) with random gaussian noise. And the quantitative comparison results are given in Fig. 5. It can be seen that for linear unmixing, the central region of Fig. 5(d) produced a large estimation error due to spectral coloring caused by light absorption and scattering. In contrast, ResU-net's relative reconstruction error is much smaller (Fig. 5(e)), which implied that deep learning has learned this inverse mapping and can compensate for spectral coloring well. Later, there are some recent works [85-89] that are also based on U-net to quantify sO$_2$. In order to investigate whether it is theoretically possible to give accurate quantitative results using only two wavelengths, Ref. [85] conducted some simulation experiment verifications. They think that the feature maps obtained by deep layer of ResU-net can recursively convolve each row of vectors to obtain sequence information, which is beneficial for blood oxygen saturation of each coordinate in space with contextual information. Clinically-obtained numerical breast phantom is used for optical and acoustic simulations in Ref. [87] avoiding the use of simple layers of tissues. And in order to take full advantage of multi-wavelength as inputs, aggregator was added with encoder and decoder. They believe that shallow features are propagated through different stages of aggregation and will be refined [90]. The authors also simply verified the impact on the quantitative results corresponding to the number of wavelengths. When the number of input wavelengths increases, the neural network can produce more accurate results, but the improvement is finally limited. Because the concentration of sO$_2$ in blood vessels is more concerned than other tissues (such as fat), blood vessel segmentation and quantification were performed simultaneously in Ref. [88]. Two symmetric U-nets were used for segmentation and quantification, then focused the analysis result of the entire domain to the segmented blood vessels. Ref. [89] expanded quantitative blood oxygenation from 2D to 3D PA images since the spatial information provided by the 3D images improves the neural network’s ability to learn an optical fluence correction and the voxel-wise approach can use the full information. Different from the above work that considers the global spatial information of each wavelength separately, Janek Grohler [91] computed sO$_2$ from pixel-wise initial pressure spectra, which are vectors comprised of the initial pressure at the same spatial location over all recorded wavelengths. Ref. [92] learned the inverse mapping by directly regressing from a set of input spectra to the desired fluence based on eigenspectra multispectral optoacoustic tomography [93]. Both spectral and spatial features were used for inference, which was verified in simulations and experimental dataset obtained from blood phantoms and mice in vivo. This is also the first in vivo experiment to prove PA sO$_2$ imaging with deep learning.

The above reviewed methods are all supervised learning for quantitative blood oxygen imaging. However, since the ground truth of sO$_2$ is difficult to obtain for in vivo imaging experiments, it seems that unsupervised learning may be a better choice to solve this problem. Almost all the studies are currently in the stage of feasibility study, Ref. [94] firstly tried to use unsupervised learning to identify regions containing HbO$_2$ and HbR. Although these existing works are still at very preliminary stage, we can still see the dawn of quantitative PA imaging using deep learning. With more investigations in clinical use, more realistic data will be obtained and methods will be verified.

C. PA Image detection

Image detection algorithm is to process the image and detect objects in it. As a promising imaging method, PA imaging, even in the early development stage, also has some published work with deep learning for cancer detection [97-101]. Arjun et. al explored deep neural networks to diagnose prostate cancer based on multispectral PA image database [99]. In this work, a combination of an adaptive greedy forward with backward removal features’ selector was used to select features as a key strategy, and then an optimal detection result was given along with a CNN detection/classification model. This is the first time deep learning has been used to detect prostate cancer on ex-vivo cancer bearing human prostate tissue with PA imaging. Based on this work, the detection of prostate cancer has expanded to 3D [97], and the experiment was verified based on transfer learning to the prostate using the thyroid database [98]. In order to solve the problem of body hair signal that hampers the visibility of blood vessels in PA imaging, a novel semi-supervised learning (SSL) method was proposed in [100], since the actual training data and test data are very limited. Due to the directional similarity between adjacent body hairs, the author introduced this fact as a priori knowledge into SSL, which enabled the effective learning of the discrimination model for detecting body hair from a small training data set. Finally, the effectiveness of their method was quantitatively verified based on experimental data and successfully applied to virtual hair removal to improve the visibility of blood vessels in PA imaging. Xiang Ma [101] firstly presented the principle of using deep learning to evaluate the average adipocyte size. A deep neural network with fully connected layers was used to fit the relationship between PA spectrum and average adipocyte size. Since the size of adipocyte in obese people is directly related to the risk
of metabolic diseases, deep learning has the potential for noninvasive assessment of adipose dysfunction.

D. PA Image classification

Image classification is to label objects in the image, firstly detecting the different objects and then classifying them, which can be regarded as an advanced version of image detection. There are only a few preliminary works [102, 103] about classification in PA imaging. Yongping Lin [102] achieved automatic classification of early endometrial cancer based on co-registered PA and ultrasonic (CRPU) signals. The CRPU images were obtained based on Monte Carlo simulations in case standardized CRPU images were hardly got with most PAI systems still in prototype phase. Jiayao Zhang [103] explored the deep learning algorithms for breast cancer diagnostics, where transfer learning was used to achieve better classification performance. Facing the problem of limited data sets, the author used a pre-processing algorithm to enhance the quality and uniformity of input breast cancer images for the experiments.

E. PA Image segmentation

Image segmentation is the holy grail of quantitative image analysis, the goal of which is outputting a pixel-wise mask of images. The image is divided into several regions or parts based on the characteristics of the pixels. Some recent works have been reported about segmentation of the PA image domain using deep learning [77, 103-105], which is often performed jointly \(^{16}\) with other tasks. A partially-learned algorithm is proposed in Ref. [77] for joint PA reconstruction and segmentation. Different from other reconstruction works, the authors assumed that segmentation of the vascular geometry is of high importance since the visualization of blood vessels is the main task in PAI. The unique advantages of PA imaging make it possible to unmix HbO\(_2\) and Hb in multispectral optoacoustic tomography (MSOT), which is beneficial to the precise segmentation. A deep learning approach with a sparse UNET (S-UNET) for automatic vascular segmentation in MSOT images was used in Ref. [104] to avoid the rigorous and time-consuming manual segmentation. The wavelength selection module helped to select the optimal wavelengths for the current segmentation task. It also helped to reduce the longitudinal scanning time and data volume in multi-wavelength experiments. Facing the errors in the co-registration of compound images of optoacoustic and ultrasound (OPUS) imaging, proper segmentation of different regions turns essential. An automatic segmentation method based on deep learning was proposed in Ref. [105] for segmenting the mouse boundary in a pre-clinal OPUS system. The experimental results were shown to be superior than another SOTA method in a series of experimental OPUS images of the mouse brain, liver and kidney regions.

F. PA-assisted intervention

PA imaging has been developed to guide surgery in some research works [109, 110] benefiting from its real-time imaging capability. Ref. [109] used SOTA deep learning methods to improve image quality by learning from the physics of sound propagation. Those deep learning methods were used in PA-assisted intervention to hold promise for visualization and visual servo of surgical tool tips, and evaluated the distance between critical human structures near the tools used for surgery (for example, serious complications, paralysis or death of the patient will occur when major blood vessels and nerves are injured). Neurosurgery, spinal fusion surgery, hysterectomies, and biopsies, these surgeries and procedures will be beneficially affected clinically. Catheter guidance is often used in cardiac interventions, and PA signal can be generated at catheter tips. It is important to accurately identify the position of the catheter tip because the projection will lack depth information during fluoroscopy. Novel deep learning PA point source detection technique was used [110] to identify catheter tips in the presence of artifacts in raw data, prior to implementing the beamforming steps required to form PA images.

IV. OPEN SOURCE FOR DEEP LEARNING IN PA IMAGING

Most researchers tend to upload their preliminary manuscript to the arXiv\(^{17}\) preprint server as soon as possible and share their corresponding code on github\(^{18}\). You can also find most public data sets through various repositories, which is also the starting point for many competitions or challenges. Many competitions every year often attract many new participants, who put forward their own methods on the same issue to fight for the best results, which always push the SOTA to a new level. With such a rich variety of open access resources, if finding a problem of interest, everyone can easily start their own research based on the public data set, the method described in the preprint, and the project implemented on github.

There have been many excellent articles summarizing the implementation, data sets and challenges. For example, you can find relevant summaries about a short list of publicly available codes for DL in medical imaging, medical imaging data sets and repositories in [1]. Since we can design or improve our own methods and models inspired by excellent, open-accessed deep learning implementation, we briefly summarize the available data sets in PA imaging for deep learning research. Although limited by the status quo that PA imaging equipment has not been clinically available, and the number of open data sets is small, we believe that these can

\(^{16}\) Joint task refers to the execution of multiple categories of tasks together, producing information complementarity, or one task may have a supervisory effect on another task. There are many examples [106-108] in medical image analysis.

\(^{17}\) Deep learning is moving at a breakneck speed, too fast for the standard peer-review process to keep up. In the past few years, many of the most famous and influential papers in machine learning can only be available as preprints before they were published in conference proceedings. Hence you can see the latest research through the preprint website arXiv: https://arxiv.org/.

\(^{18}\) Most researchers are often willing to share code and data on github (https://github.com/), which to a certain extent also promotes the rapid development of this community.
still help peers to conduct high-quality research based on these data sets. Existing data sets can be divided into two categories: one is physiologically digital phantom, and the other is experimental PA imaging data. Yang Lou in Ref. [11] proposed a computational methodology using clinical contrast-enhanced magnetic resonance imaging data to generate anatomically 3D realistic numerical breast phantoms, and finally provide three different BI-RADS breast density levels’ digital breast models named Optical and Acoustic Breast Phantom Database (OA-Breast)19. In [2], the authors released two real experimental datasets20 (MSOT-Brain and MSOT-Abdomen) and some numerical simulation data. In [3], the realistic mouse cross-sectional data includes sampling data of different probe numbers can be downloadable21. In [68], a set of mouse brain data acquired from OR-PAM are shared22.

V. CHALLENGES AND FUTURE PERSPECTIVES

Deep learning shows its potential for medical imaging in various modalities including PA imaging, and achieve SOTA results when tasked with producing precise decisions based on complicated data sets. But there are still many challenges and limitations that need to be overcome. In addition to its data-driven black box-nature, the methods often used for comparison are based on traditional manual design when introducing deep learning into PA imaging. Nevertheless, neural network-based methods can easily outperform these baselines, which makes this comparison meaningless. Largely due to the inability to obtain open data sets, researchers always verify on their own closed data sets with most of the PA imaging systems still in prototype phase. Therefore, although the research on using deep learning to solve some problems in PAI has improved greatly, there is still a lack of fair comparison on training and testing results using large-scale standardizad real data sets. We believe when scientists and clinicians gradually develop the current clinical application standardization for PAI, a large amount of real patient data will be available. These problems will be conquered and the entire PA imaging community will develop more harmoniously. As machine learning researchers and clinicians gain more experience, it will be easier to solve current clinical problems using reasonable solutions. Once there are enough mature solution systems based on mathematics, computer science, physics and engineering entering the daily workflow in the clinic, computational medicine will become mainstream. Machine learning and other computational medicine-based technology ecosystem will eventually be established in most biomedical, and PA imaging will finally find its unique position.

REFERENCES

This research was funded by Start-up grant of ShanghaiTech University (F-0203-17-004), National Science Foundation of Shanghai (18ZR1425000), and National Natural Science Foundation of China (61805139).

[1] A.S. Lunderervold, A.J.Z.F.M.P. Lunderervold, An overview of deep learning in medical imaging focusing on MRI, Zeitschrift für Medizinische Physik, 29(2) (2019) 102-127.
[2] S. Guan, A.A. Khan, S. Sikdar, P.V.J.S.R. Chitnis, Limited-View and Sparse photoacoustic tomography for neuroimaging with Deep Learning, Scientific Reports, 10(1) (2020) 1-12.
[3] N. Davoudi, X.L. Deán-Ben, D.J.N.M.I. Razansky, Deep learning optoacoustic tomography with sparse data, Nature Machine Intelligence, 1(10) (2019) 453-460.
[4] D.E. Rumelhart, G.E. Hinton, R.J.J.n. Williams, Learning representations by back-propagating errors, Nature, 323(6088) (1986) 533-536.
[5] Y. LeCun, L. Bottou, Y. Bengio, P.J.P.o.t.I. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86(11) (1998) 2278-2324.
[6] S.J.a.p.a. Ruder, An overview of gradient descent optimization algorithms, p. arXiv:1609.04747, (2016).
[7] K. He, X. Zhang, S. Ren, J.J.a.e.-p. Sun, Deep Residual Learning for Image Recognition, 2015, p. arXiv:1512.03385.
[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A.J.a.e.-p. Rabinovich, Going Deep with Convolutions, 2014, p. arXiv:1409.4842.
[9] S. Zagoruyko, N.J.a.e.-p. Komodakis, Wide Residual Networks, 2016, p. arXiv:1605.07146.
[10] S. Xie, R. Girshick, P. Dollár, Z. Tu, J.J.a.e.-p. He, Aggregated Residual Transformations for Deep Neural Networks, 2016, p. arXiv:1603.08388.
[11] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012, pp. 1097-1105.
[12] J. Howard, S.J.a.e.-p. Ruder, Universal Language Model Fine-tuning for Text Classification, 2018, p. arXiv:1801.06146.
[13] T. Young, D. Hazarika, S. Poria, E.J.a.e.-p. Cambria, Recent Trends in Deep Learning Based Natural Language Processing, 2017, p. arXiv:1708.02709.
[14] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K.J.a.e.-p. Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, 2016, p. arXiv:1609.03499.
[15] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, J.J.a.e.-p. Stolcke, The Microsoft 2017 Conversational Speech Recognition System, 2017, p. arXiv:1708.06073.
[16] C. Yang, H. Lan, F. Gao, Accelerated Photoacoustic Tomography Reconstruction via Recurrent Inference Machines, 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 6371-6374.
[17] P.J.I.f. Beard, Biomedical photoacoustic imaging, 1(4) (2011) 602-631.
[18] W. Choi, E.-Y. Park, S. Jeon, C.J.B.e.l. Kim, Clinical photoacoustic imaging platforms, 8(2) (2018) 139-155.
[19] F. Gao, X. Feng, Y.J.J.o.O. Zheng, Advanced photoacoustic and thermal-acoustic imaging beyond pulsed absorption contrast, Journal of Optics, 18(7) (2016) 074006.
[20] W.R. Morningstar, L. Perreault Levasseur, Y.D. Hezaveh, R. Blandford, P. Marshall, P. Putzky, T.D. Rueter, R. Wechsler, M. Welling, Data-driven Reconstruction of Gravitationally Lensed Galaxies Using Recurrent Inference Machines, The Astrophysical Journal, 883 (2019) 14.
[21] H. Shen, D. George, E.A. Huerta, Z.J.a.e.-p. Zhao, Denoising Gravitational Waves using Deep Learning with Recurrent Denoising Autoencoders, 2017, p. arXiv:1711.09919.
[22] K. Lenning, P. Putzky, M.W. Caan, M. Welling, Recurrent inference machines for accelerated MRI reconstruction, (2018).
[23] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A.J.n. Bolton, Mastering the game of go without human knowledge, Nature, 550(7676) (2017) 354-359.
[24] G. Wang, J.C. Ye, K. Mueller, J.A.J.I.t.o.m.i. Fessler, Image reconstruction is a new frontier of machine learning, IEEE transactions on medical imaging, 37(6) (2018) 1289-1296.

[25] A. Hauptmann, F. Lucka, M. Betcke, N. Huyhnh, J. Adler, B. Cox, P. Bear, S.J.I.t.o.m.i. Bridge, Model-based learning for accelerated, limited-view 3-d photoacoustic tomography, IEEE transactions on medical imaging, 37(6) (2018) 1382-1393.

[26] Y. LeCun, Y. Bengio, G.J.n. Hinton, Deep learning, Nature, 521(7553) (2015) 436-444.

[27] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press 2016.

[28] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I.J.M.i.a. Sánchez, A survey on deep learning in medical image analysis, Medical image analysis, 42 (2017) 60-88.

[29] A. Zhang, Z.C. Lipton, M. Li, A.J.J.U.D.R. Smola, Dive into deep learning, Unpublished Draft. Retrieved, 19 (2019) 2019.

[30] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, Pytorch: An imperative style, high-performance deep learning library, Advances in neural information processing systems, 2019, pp. 8026-8037.

[31] M. Leshno, V.Y. Lin, A. Pinkus, S.J.N.n. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural networks, 6(6) (1993) 861-867.

[32] S. Sonoda, N.J.A. Murata, C.H. Analysis, Neural network with unbounded activation functions is universal approximator, Applied and Computational Harmonic Analysis, 43(2) (2017) 253-265.

[33] D-A. Clevert, T. Unterthiner, S.J.a.p.a. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), p. arXiv:1511.07289, (2015).

[34] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026-1034.

[35] J.T. Springenberg, A. Dosovitskiy, T. Brox, M.J.a.p.a. Riedmiller, Striving for simplicity: The all convolutional net, p. arXiv:1412.6806, (2014).

[36] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R.J.T.j.o.m.l.r. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15(1) (2014) 1929-1958.

[38] S. Ioffe, C.J.a.p.a. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, p. arXiv:1502.03167, (2015).

[39] A. Khan, A. Sohail, U. Zahoora, A.S.J.A.I.R. Qureshi, A survey of the network training by reducing internal covariate shift, p. arXiv:1502.03167, (2015).

[40] S. Ioffe, C.J.a.p.a. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, p. arXiv:1502.03167, (2015).

[41] N. Srivastava, V.Y. Lin, A. Pinkus, S.J.N.n. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural networks, 6(6) (1993) 861-867.

[42] S. Sonoda, N.J.A. Murata, C.H. Analysis, Neural network with unbounded activation functions is universal approximator, Applied and Computational Harmonic Analysis, 43(2) (2017) 253-265.

[43] D-A. Clevert, T. Unterthiner, S.J.a.p.a. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), p. arXiv:1511.07289, (2015).

[44] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026-1034.

[45] J.T. Springenberg, A. Dosovitskiy, T. Brox, M.J.a.p.a. Riedmiller, Striving for simplicity: The all convolutional net, p. arXiv:1412.6806, (2014).

[46] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[47] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R.J.T.j.o.m.l.r. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15(1) (2014) 1929-1958.

[48] S. Ioffe, C.J.a.p.a. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, p. arXiv:1502.03167, (2015).

[49] A. Khan, A. Sohail, U. Zahoora, A.S.J.A.I.R. Qureshi, A survey of the recent architectures of deep convolutional neural networks,ritificial Intelligence and Theory of Computing, 1-62.

[50] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, Classification in BioApps, Intelligence Review, (2020) 1-62.

[51] M. Leshno, V.Y. Lin, A. Pinkus, S.J.N.n. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural networks, 6(6) (1993) 861-867.

[52] S. Sonoda, N.J.A. Murata, C.H. Analysis, Neural network with unbounded activation functions is universal approximator, Applied and Computational Harmonic Analysis, 43(2) (2017) 253-265.

[53] D-A. Clevert, T. Unterthiner, S.J.a.p.a. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), p. arXiv:1511.07289, (2015).

[54] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026-1034.

[55] J.T. Springenberg, A. Dosovitskiy, T. Brox, M.J.a.p.a. Riedmiller, Striving for simplicity: The all convolutional net, p. arXiv:1412.6806, (2014).

[56] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.

[57] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R.J.T.j.o.m.l.r. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15(1) (2014) 1929-1958.
[108] Z. Xu, M. Niethammer, DeepAtlas: Joint semi-supervised learning of image registration and segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 420-429.

[109] M.A.L. Bell, Deep learning the sound of light to guide surgeries, Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XVII, International Society for Optics and Photonics, 2019, p. 108680G.

[110] D. Allman, F. Assis, J. Chrispin, M.A.L. Bell, Deep learning to detect catheter tips in vivo during photoacoustic-guided catheter interventions: Invited presentation, 2019 53rd Annual Conference on Information Sciences and Systems (CISS), IEEE, 2019, pp. 1-3.

[111] Y. Lou, W. Zhou, T.P. Matthews, C.M. Appleton, M.A.J.J.o.B.O. Anastasio, Generation of anatomically realistic numerical phantoms for photoacoustic and ultrasonic breast imaging, Journal of Biomedical Optics, 22(4) (2017) 041015.