



Deep learning improves contrast in low-fluence photoacoustic imaging

ALI HARIRI,^{1,7} KAMRAN ALIPOUR,^{2,7} YASH MANTRI,³ JURGEN P. SCHULZE,^{2,4} AND JESSE V. JOKERST^{1,5,6,*}

¹Department of NanoEngineering, University of California, San Diego, La Jolla, CA 92093, USA

²Department of Computer Science, University of California, San Diego, La Jolla, CA 92093, USA

³Department of BioEngineering, University of California, San Diego, La Jolla, CA 92093, USA

⁴Qualcomm Institute, University of California, San Diego, La Jolla, CA 92093, USA

⁵Department of Radiology, University of California, San Diego, La Jolla, CA 92093, USA

⁶Material Science Program, University of California, San Diego, La Jolla, CA 92093, USA

⁷These authors contributed equally to this paper

*jjokerst@ucsd.edu

Abstract: Low fluence illumination sources can facilitate clinical transition of photoacoustic imaging because they are rugged, portable, affordable, and safe. However, these sources also decrease image quality due to their low fluence. Here, we propose a denoising method using a multi-level wavelet-convolutional neural network to map low fluence illumination source images to its corresponding high fluence excitation map. Quantitative and qualitative results show a significant potential to remove the background noise and preserve the structures of target. Substantial improvements up to 2.20, 2.25, and 4.3-fold for PSNR, SSIM, and CNR metrics were observed, respectively. We also observed enhanced contrast (up to 1.76-fold) in an in vivo application using our proposed methods. We suggest that this tool can improve the value of such sources in photoacoustic imaging.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Photoacoustic imaging (PAI) combines the high-contrast of optical imaging and the high spatial resolution of ultrasound [1–3]. Short optical pulses serve as an excitation source in PAI systems [4] to generate photoacoustic waves via thermoelastic expansion [5,6]. Wideband ultrasonic transducers detect the propagated waves, and mathematical processing methods (reconstruction algorithms) can transform the detected signals into an image [7–11]. Over the last decade, investigators have demonstrated various applications of PAI in ophthalmology [12–14], oncology [15–17], dermatology [18–20], cardiology [21–23], etc.

PAI traditionally uses solid-state lasers as an excitation source because of their tunability, coherence, and high pulse energy. However, utilization of these lasers is difficult in clinical applications because they are bulky, expensive, unstable (in terms of power intensity fluctuation), and require frequent maintenance [24]. In contrast, pulse laser diodes (PLD) [25–27] and light emitting diodes (LED) [28–30] are a stable, affordable, and compact alternative light source. However, the output pulse energy of PLDs and LEDs is low—on the order of $\mu\text{J}/\text{pulse}$ and nJ/pulse versus mJ/pulse with lasers. Thus, the resulting photoacoustic data needs be averaged hundreds of time to cancel the noise and extract meaningful signal [28]. Unfortunately, performing many averages negatively affects the temporal resolution. Investigators have improved the signal-to-noise ratio (SNR) using classical signal processing methods such as empirical mode decomposition (EMD) [31,32], wavelet-based methods [33,34], Wiener deconvolution [35], principle component analysis (PCA) [36], and adaptive noise canceler (ANC) [37,38]. However, these methods all require some prior information about the signal and noise properties, which is

a significant limitation. Therefore, new tools to increase the SNR in low fluence PLD and LED PAI are needed.

Deep learning is rapidly expanding within various fields and improving performance in pattern recognition and machine learning applications. These relatively new techniques have vastly outperformed other classical methods in recent years. For example, computer vision has extensively utilized deep learning algorithms object detection, image classification, and image reconstruction [39–42]. Convolutional neural networks (CNN) are among the most popular deep learning algorithms [43].

In medical imaging, previous studies focused on denoising CT-specific noise patterns. Kang et al. [44] utilized CNNs for wavelet transform coefficients with low-dose CT images. Chen et al. [45] used CNNs to directly map low-dose CT images to their normal-dose counterparts. Other methods altered the original CNN architecture to either preserve details in the image through *residual* blocks [46,47] or *generator* CNNs to produce the restored image based on encoded features of low-dose images [48–50].

To capture more spatial context, previous approaches used pooling between convolution to reduce feature map resolution. However, pooling extends the receptive field and depth of their CNNs to drastically increase the computational costs in training and deploying such models [51]. Dilation [52] is another alternative to pooling but is limited by sparse sampling in the input layer, which can lead to gridding issues [53].

The concept of denoising in PA images is similar to low-dose CT yet the noise can have very different patterns; hence, the noise requires a different transfer method to be removed. Some of the earlier methods used short-lag spatial coherence [54,55] or singular value decomposition (SVD) [56] to remove reflection artifacts from PA images. Some recent approaches utilize CNNs to identify point sources per image [57,58] or use recurrent neural networks (RNN) to leverage temporal information in PA images to remove artifacts [59]. Antholzer et al. [60] adopted U-net architectures to reconstruct photoacoustic tomography (PAT). Anas et al. [61] utilized skip connections in dense convolutional networks to improve the quality of the PA images.

The potential of deep learning to enhance image quality motivates this work with a deep convolutional neural network. The goal is to map the low-fluence light photoacoustic images to corresponding high-fluence photoacoustic data. We demonstrated that deep learning can restore the images in low-fluence photoacoustic configuration with less computational cost than classical methods. Here, we first describe the proposed deep convolutional neural network and training details. We then demonstrate qualitative and quantitative phantom results. Finally, we show the capability our proposed model to image low concentrations of contrast agents *in vivo*.

2. Methods and materials

2.1. Training

The main aim of this study is to train the proposed convolutional network (MWCNN) to transform the low-fluence photoacoustic data into high-fluence images. We defined the high fluence images as the ground truth and then used TiO₂-based optical scatters to reduce the laser fluence (**Fig. 1(A)**). The laser fluence was 17, 0.95, 0.25, 0.065, and 0.016 mJ/pulse at a wavelength of 850 nm with 0, 4, 6, 8, and 10 mg/ml of TiO₂, respectively (**Fig. 7** in the Appendix). Importantly, we did not change the imaging setup in either the acoustic or optical path. We simply changed the disks without touching the transducer or the 3D samples. To further confirm that now changes were caused when switching the disks, we selected some landmarks on the first and last frames and compared those landmark among all the filters. There were no significant movements of these landmarks. Network training used a 3D pen print (Ginccley 3D Pen, AM3D Printers Inc.) to prepare a complicated 3-dimensional structure (2 cm x 2 cm x 3 cm) (**Fig. 1(B)**). These structures were placed in an agarose phantom and scanned (30 mm, 270 frames) with all five optical filters on top of them (**Fig. 1(C)**). We used 0 mg/ml which has the highest laser fluence (~17 mJ/pulse)

as a ground truth in the proposed network. **Figure 1(D)** shows that the signal-to-noise ratio will decrease by decreasing the laser fluence (increasing the TiO₂ concentration).

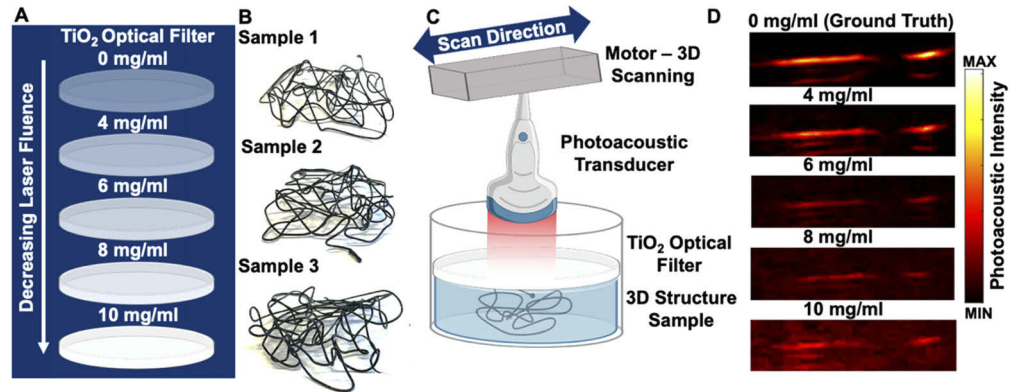


Fig. 1. Experimental training setup. **A)** TiO₂-based optical scattering gels reduce the laser fluence. The laser fluence at wavelength of 850 nm was 17, 0.95, 0.25, 0.065, and 0.016 mJ/pulse after using 0, 4, 6, 8, and 10 mg/ml of TiO₂, respectively. **B)** Three different complicated 3D structures were made using a 3D pen print to collect training data. **C)** Imaging setup—the 3D structures are placed in the agarose phantom with a TiO₂-based optical scatterer on top. We scanned the entire structure for each sample and acquired 270 frames; 850 nm was the illumination wavelength. **D)** B-mode photoacoustic images with different optical scatterer concentrations and thus laser fluence values show reduced SNR with increasing TiO₂ concentration (decreased laser fluence). Scale bars represent 1 cm.

In the training process, 85% of the frames (with fluence values of 17, 0.95, 0.25 mJ/pulse) were randomly selected as training set and the rest as test set. The training algorithm was implemented under the PyTorch platform using two NVIDIA GeForce GTX 1080 Ti GPUs. We used ADAM optimizer for our training algorithm with an initial learning rate of 1.024×10^{-4} . The training process completed 256 epochs in one day.

2.2. Testing

2.2.1. Low fluence laser source

To test the trained model, we printed term “UCSD” on a transparent film in black ink and placed it beneath the agarose hydrogel. The black ink is strongly absorbing and will produce photoacoustic signal. The TiO₂ optical scatters with the same concentrations as in the training section were used to test the model under different laser fluence values. We collected 270 frames, and each frame was individually used as an input in the model.

For a fair evaluation of our trained model, we need a testing test that does not share the same illumination condition as training. Such a test would quantify the extent of model’s operational scope around a set of training data. In that regard, testing was done at 0.065 and 0.016 mJ/pulse fluences to evaluate model scalability in illuminations lower than the training domain (0.95 and 0.25 mJ/pulse). This strategy minimizes the risk of overfitting the model and we evaluate the model based on actual sample features and beyond certain fluency ranges. All output frames were placed next to each other to generate the 3D volumetric data. Importantly, the trained model was totally blind to this new data set. We measured the PSNR and SSIM metrics on each letter: U, C, S, and D. We performed t-test statistical analysis, and p values < 0.05 were considered to be significantly different.

2.2.2. LED-based light source

We also tested the model with the LED-based photoacoustic imaging system but without any nanoparticle gel scatterers (LED system has inherently low fluence). We again printed “UCSD” and placed it beneath a transparent agarose hydrogel. The LED was operated at 1 K Hz and 2 K Hz for 40 and 80 $\mu\text{J}/\text{pulse}$ on the sample. We used all 180 frames as the input for the model. We placed all frames after each other to create a 3D volumetric image. The ground truth data were collected by operating the LED source at 4 K Hz (160 μJ of fluence). We used 20 rounds of averaging for each frame. We separated each letter 3D map and measured PSNR and SSIM metric on each letter.

We also evaluated the model with an LED-based system and a different configuration. We placed pencil lead (0.5 mm HB, Newell Rubbermaid, Inc., IL, USA) at depths of 2.5, 7.5, 12.5, 17.5, and 22.5 mm in 2% intralipid (20%, emulsion, Sigma-Aldrich Co, MO, USA) mixed with agarose. We used intralipid to mimic biological tissue. We collected a single frame with the LED system at 1 K Hz and 2 K Hz to have 40 and 80 $\mu\text{J}/\text{pulse}$ on the surface of the sample. To measure the CNR for both input (Noisy) and output (MWCNN model) images, μ_{object} and $\mu_{\text{background}}$ were defined as the average (5 different areas) of the mean photoacoustic intensity on the pencil lead (ROI of 1 X 1 mm^2) and the background area (ROI of 3 X 3 mm^2), respectively. Term $\sigma_{\text{background}}$ is the average of all five standard deviations of background intensity.

2.2.3. In vivo performance

We also evaluated our trained model in its ability to enhance the contrast agent detection in vivo. Here, the murine tissue reduces the fluence. We purchased nine nude mice (8-10 weeks) from the University of California San Diego Animal Care and Use Program (ACP). All animal experiments were performed in accordance with NIH guidelines and approved by the Institutional Animal Care and Use Committee (IACUC) at the University of California, San Diego. The mice were anesthetized with 2.5% isoflurane in oxygen at 1.5 L/min. Methylene blue (MB) (Fisher Science Education Inc., PA, USA) was purchased and dissolved in distilled water. MB concentrations of 0.01, 0.05, 0.1, 1, and 5 mM were injected intramuscularly in a murine model (n=3). The Vevo LAZR (VisualSonic Inc.) system was used for this in vivo experiment. We monitored the injection procedure using both ultrasound and photoacoustic images. The location of injected MB was confirmed using photoacoustic spectral data. We measured the CNR for both input (Noisy) and output (MWCNN model) images. For this calculation, μ_{object} and $\mu_{\text{background}}$ were defined as the average (5 different areas) of mean values of photoacoustic intensity at the injected area (ROI of 1 X 1 mm^2) and around the injected area (ROI of 3 X 3 mm^2), respectively. Term $\sigma_{\text{background}}$ is average of all five standard deviations of background intensity.

2.3. Multi-level wavelet-CNN

We used a multi-level wavelet-CNN (MWCNN) model with low receptive field, low computational cost, and high adaptivity for PA imaging in multi-frequency space [62]. The model is based on a U-Net architecture and consists of a contracting sub-network followed by an expanding subnetwork. The contracting subnetwork uses discrete wavelet transform (DWT) instead of pooling operations. This substitution allows high-resolution restoration of image features through inverse wavelet transform (IWT) within the expanding subnetwork.

The model takes a 512×512 noisy image as the input and transfers that to a 512×512 denoised output image (Fig. 2). The model processes the image in one channel as a heatmap. The input is a 2D cross-sectional framelet from a PA imagery set. The model attempts to reduce the noise in the image while preserving the signal. This model expands the feature dimension of the input from 1 to 1024 channels and then contracts the feature maps back into 1 channel as the output. The convolution blocks may contain multiple convolutional layers. Each convolution layer is followed by a ReLU activation function.

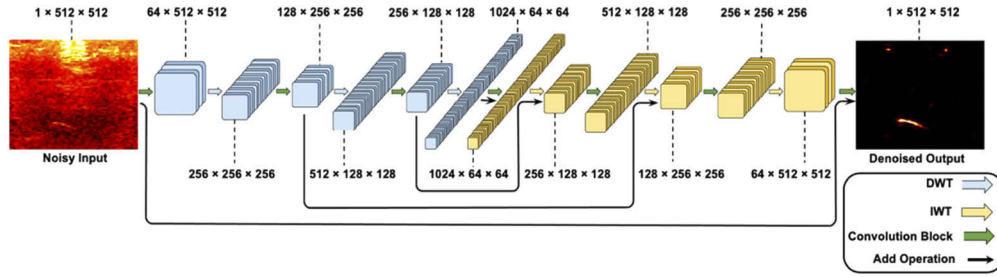


Fig. 2. MWCNN model architecture. Contracting subnetwork features are extracted in wavelet space. In expanding the subnetwork, features expand into the image space while preserving high resolution details. This model takes a 512×512 noisy image as the input and transfers that to a 512×512 denoised output image. Add operations directly feed the contracting feature maps to expanding feature maps to preserve image details and avoid blur effects.

In the contracting subnetwork of the model, the image features go through multiple convolutions with intermittent DWT blocks. Our model uses a Haar wavelet transform based on the following orthogonal filters:

$$f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad f_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \quad f_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \quad f_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (1)$$

The DWT blocks transform feature maps into four sub-bands. Due to the biorthogonal property of this operation, the original feature map can be accurately retrieved by an inverse Haar wavelet transform. The IWT blocks are then placed in between convolution blocks of the expanding subnetwork. For more details on the properties this transform, readers are referred to the original work [62].

Other CNN methods mostly use U-Net based architecture utilizing pooling operating in between convolutions—the average pooling in these systems can cause information loss in the feature maps. The MWCNN architecture benefits from DWT and IWT as a safe down-sampling and up-sampling processes where the feature maps can be transmitted with no information loss throughout the model. The objective of the training process is to optimize the model parameters Θ with the goal of minimizing the MSE loss function:

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^N \|y_i - F(x_i; \Theta)\|^2 \quad (2)$$

The training set is $\{(x_i, y_i)\}_{i=1}^N$. In this equation, x_i is the low fluence (noisy) input image, y_i is the corresponding high energy ground truth, and $F(x_i; \Theta)$ is the model output.

In PA imaging, the absolute magnitude of the signal and noise is dependent on multiple factors like light illumination, acoustic detection, and the experimental setup. Training a model based on the relative magnitude of the signal and noise might limit the model to specific types of samples and settings. Here, we focus the training on the shape features of the signal rather than the magnitude because such a model can be more generic and scalable. To minimize the model's reliance on the signal magnitude, we normalize the pixel values of the input between zero and one. In this setting, the model is inclined to distinguish noise from signal based on the *shape* features. The model is trained in a supervised manner to transform low energy inputs into outputs as close as possible to the ground truth frames.

2.4. Photoacoustic imaging system

Two different commercially available pre-clinical photoacoustic imaging systems were used in this study. Both systems can physically translate the transducer to generate three-dimensional (3D) images. Model training used the Vevo LAZR (VisualSonic Inc.), which utilizes laser excitation integrated into a high frequency linear array transducer (LZ-201, Fc = 15 MHz) with optical fibers integrated to both sides of the transducer [63]. For optical excitation, this system uses a Q-switched Nd:YAG laser (4-6 ns pulse width) with a repetition rate of 20 Hz (frame rate of 6 Hz) followed by an optical parametric oscillator (tunable wavelength 680-970 nm). The laser fluence was 17.06 ± 0.82 mJ using a laser pyroelectric energy sensor (PE50BF-C, Ophir LLC, USA).

To modulate the intensity of the laser, we placed the sample in different concentrations of TiO₂ nanoparticles. These nanoparticles are well-known scatterers that decrease the fluence on the sample (when placed between the source and the sample). We measured the fluence with different concentrations of nanoparticles using same energy sensor mentioned above.

Test data used a scanner lower fluence than the laser (LED excitation; AcousticX; CYBERDYNE Inc.) [28]. This system is equipped with a 128-element linear array ultrasound transducer with a central frequency of 10 MHz and a bandwidth of 80.9% fitted with two 690 nm LED arrays. The repetition rate of these LEDs is tunable between 1, 2, 3, and 4 K Hz. The pulse width can be changed from 50 ns to 150 ns with a 5-ns step size. The LED fluence at 1 K Hz and 2 K Hz was 20 and 40 μ J/pulse, respectively.

2.5. Image evaluation metrics

2.5.1. Peak signal-to-noise ratio (PSNR)

We used the PSNR metric to evaluate the model in terms of noise cancellation. The PSNR is described in decibel (dB) and calculated based on square differences between the model output and ground truth images as:

$$PSNR = 20 \log_{10} \left(\frac{I_{max}}{\sqrt{MSE}} \right) \quad (3)$$

where,

$$MSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{GT}(m, n) - I_{MWCNN}(m, n))^2 \quad (4)$$

I_{GT} and I_{MWCNN} are the ground truth and model output images, respectively. Term I_{max} represents the maximum possible value in the given images [59].

2.5.2. Structural similarity index measurement (SSIM)

The SSIM evaluates image quality in terms of structural similarity; it is represented on a scale of 1. A higher SSIM shows better structural similarity of an output model image with the ground truth data.

$$SSIM = \frac{(2\mu_{GT}\mu_{MWCNN} + k_1)(2\sigma_{cov} + k_2)}{(\mu_{GT}^2 + \mu_{MWCNN}^2 + k_1)(\sigma_{GT}^2 + \sigma_{MWCNN}^2 + k_2)} \quad (5)$$

Here, μ_{GT} (σ_{GT}) and μ_{MWCNN} (σ_{MWCNN}) are the mean (variance) of the ground truth and MWCNN output images, respectively; σ_{cov} shows the covariance between these two data. k_1 and k_2 are used to stabilize the division with a weak denominator [59].

2.5.3. Contrast-to-noise ratio (CNR)

CNR determines the image quality and is described in decibel (dB) via the following equation:

$$CNR = 20 \log_{10} \frac{(\mu_{object} - \mu_{background})}{\sigma_{background}} \quad (6)$$

Here, μ_{object} and $\mu_{background}$ are mean of object and background noise, respectively; $\sigma_{background}$ represents standard deviation of background intensity in the image [64].

3. Results

3.1. Low fluence laser source

Laser fluence decreases when passing through a scattering media such as biological tissue. The photoacoustic signal intensity is proportional to the laser fluence, and thus the quality of acquired images will be affected. We intentionally decreased the laser fluence and improved the acquired images using MWCNN.

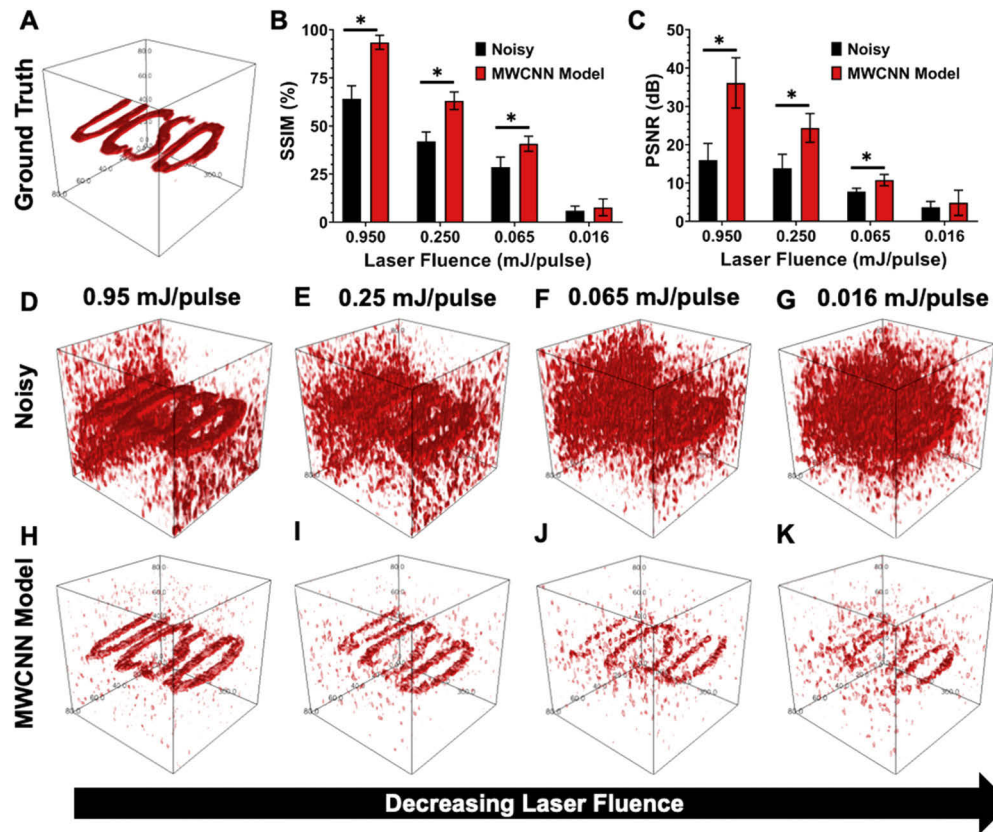


Fig. 3. Low fluence laser source evaluation. A) Ground truth 3D image of UCSD sample with full laser fluence of 17 mJ/pulse. We used this image as a reference for measuring image quality metrics. B) SSIM of noisy (input) and MWCNN data vs laser fluence. The results show that the SSIM is significantly improved by 1.45, 1.5, and 1.62 at laser fluence values of 0.95, 0.25, 0.065 mJ/pulse, respectively. The model failed to improve the structural similarity at fluence of 0.016 mJ/pulse. C) PSNR of both noisy and MWCNN data vs laser fluence—the PSNR is significantly improved with a factor of 2.25, 1.84, and 1.42 for 0.95, 0.25, 0.065 mJ/pulse, respectively. However, the MWCNN cannot significantly improve the image quality with a laser fluence of 0.016 mJ/pulse. In both B and C, the error bars represent the standard deviation of SSIM and PSNR among the four letters in “UCSD”. * indicates $p < 0.05$. D, E, F, and G) Noisy (input) images with 0.95, 0.25, 0.065, and 0.016 mJ/pulse laser fluence, respectively. H, I, J, and K) MWCNN model (output) images with 0.95, 0.25, 0.065, and 0.016 mJ/pulse laser fluence, respectively.

Figure 3(A) illustrates the ground truth 3D image of “UCSD” using the full laser fluence (17 mJ/pulse) from the Vevo LAZR system. **Figure 3(B), (C)** shows the SSIM and PSNR vs laser fluence for both noisy (input) and MWCNN model (output) images. This model improved the SSIM with a factor of 1.45, 1.5, and 1.62 for laser fluence values of 0.95, 0.25, and 0.065 mJ/pulse, respectively. The PSNR was enhanced by 2.25-, 1.84-, and 1.42-fold for laser fluence values of 0.95, 0.25, and 0.065 mJ/pulse, respectively. The trained MWCNN model cannot significantly improve SSIM and PSNR ($p > 0.05$) at 0.016 mJ/pulse. **Figure 3(D), (E), F, and G** shows the noisy (input) 3D image of the UCSD object captured with laser fluence values of 0.95, 0.25, 0.065, and 0.016 mJ/pulse, respectively. The output of MWCNN for laser fluence of 0.95, 0.25, 0.065, 0.016 mJ is represented in **Fig. 3(H), (I), J, and K**, respectively.

3.2. LED source

We next examined our MWCNN model with LEDs as a low fluence source. **Figure 4(A)** shows the ground truth 3D photoacoustic image using the LED-based photoacoustic imaging system. After collecting all the noisy data with 40 and 80 $\mu\text{J}/\text{pulse}$ as LED fluence on the sample, we noted an improvement in SSIM by a factor of 2.2 and 2.5 for 40 and 80 $\mu\text{J}/\text{pulse}$, respectively (**Fig. 4(B)**). **Figure 4(C)** demonstrates a 2.1- and 1.9-fold increase in PSNR on MWCNN model (output) versus noisy (input) for both 40 and 80 $\mu\text{J}/\text{pulse}$. **Figure 4(D)** and **(F)** shows the noisy 3D photoacoustic image using the LED-based imaging system with fluence values of 80 and 40 $\mu\text{J}/\text{pulse}$. **Figure 4(E)** and **(G)** are MWCNN 3D results with 80 and 40 $\mu\text{J}/\text{pulse}$, respectively.

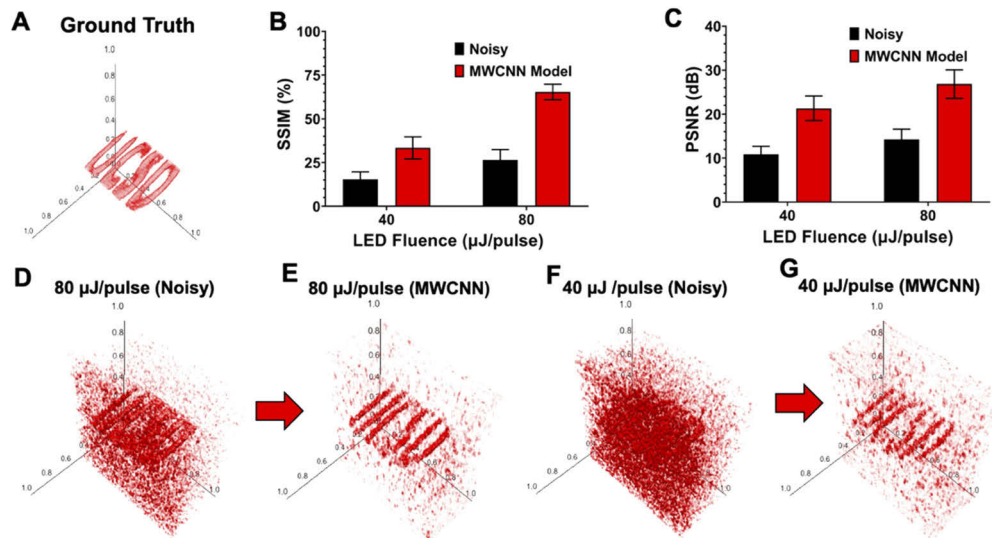


Fig. 4. LED light source evaluation. **A)** Ground truth 3D image of UCSD word using the LED-based photoacoustic imaging system. The ground truth data were collected by operating the LED source at 4 K Hz with a fluence of 160 μJ and 20 rounds of averaging for each frame. **B)** SSIM results of both noisy (input) and MWCNN model (output) in two different LED fluences of 40 and 80 $\mu\text{J}/\text{pulse}$. An improvement of 2.2- and 2.5-fold is observed for 40 and 80 $\mu\text{J}/\text{pulse}$, respectively. **C)** PSNR of noisy (input) and MWCNN model (output) at 40 and 80 $\mu\text{J}/\text{pulse}$. MWCNN improved the PSNR by 2.1 and 1.9 for 40 and 80 $\mu\text{J}/\text{pulse}$, respectively. In both B and C, the error bars represent the standard deviation of SSIM and PSNR among the four letters in “UCSD”. **D)** The 3D noisy (input) photoacoustic image used 80 $\mu\text{J}/\text{pulse}$. **E)** 3D MWCNN mode (output) photoacoustic image using 40 $\mu\text{J}/\text{pulse}$. **F)** 3D Noisy (input) photoacoustic image with 40 $\mu\text{J}/\text{pulse}$. **G)** 3D MWCNN mode (output) photoacoustic image with 40 $\mu\text{J}/\text{pulse}$.

We also evaluated the performance of the MWCNN model in depth phantoms using the LED system and scattering media. **Figure 5(A)** and **(B)** shows noisy images for LED fluence at 40 and 80 $\mu\text{J}/\text{pulse}$. We observed a significant CNR improvement both qualitatively (**Fig. 5(C)** and **(D)**) and quantitatively (**Fig. 5(E)**). We measured an average of 4.3- and 4.1-fold enhancement in MWCNN model versus noisy images at different depths for 40 and 80 $\mu\text{J}/\text{pulse}$. **Figure 5(E)** shows that the MWCNN also enhanced the linearity of CNR vs depth from $R^2 = 0.84$ and 0.85 to $R^2 = 0.97$ and 0.95 for both LED fluence values.

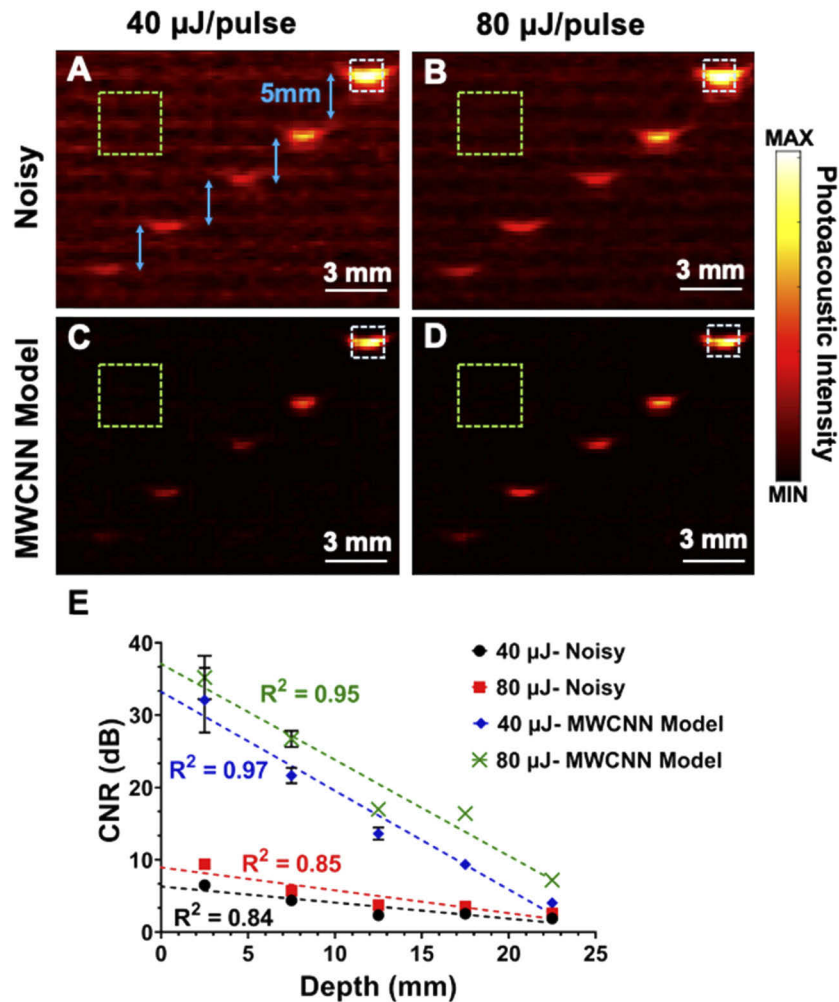


Fig. 5. Penetration depth evaluation using an LED. A) B-mode noisy (input) photoacoustic image using LED at a fluence of 40 $\mu\text{J}/\text{pulse}$. Pencil leads were placed at 2.5, 7.5, 12.5, 17.5, and 22.5 mm in 2% intralipid. B) B-mode noisy (input) photoacoustic images at a fluence of 80 $\mu\text{J}/\text{pulse}$ with similar experimental setup as described in A. C, D) B-mode MWCNN model (output) photoacoustic image for 40 and 80 $\mu\text{J}/\text{pulse}$. E) CNR versus depth for 40 and 80 $\mu\text{J}/\text{pulse}$ in both noisy and MWCNN model. Dotted green and white rectangles represent the ROI used to measure mean values and standard deviations of background (ROI size: $3 \times 3 \text{ mm}^2$) and object (ROI size: $1 \times 1 \text{ mm}^2$). We observed an average of 4.3- and 4.1-fold enhancement in the MWCNN model versus noisy data at different depths for both LED values.

3.3. *In vivo* performance

Image enhancement methods become more valid when validated *in vivo*. The detection of exogenous contrast agents using photoacoustic imaging technique can be a challenge due to low fluence due to scattering by biological tissue. Here, we injected different concentrations of MB intramuscularly in mice (**Fig. 6(A)**) and analyzed the acquired images using the MWCNN model and compared the CNR with and without the model. **Figure 6(B)** represents the CNR for both noisy (input) and MWCNN model (output) for all 5 different concentrations. We observed a significant improvement between noisy and model output. There was an improvement of 1.55, 1.76, 1.62, and 1.48 in CNR for 0.05, 0.1, 1.0, and 5.0 mM, respectively. The MWCNN failed to improve the CNR for 0.01 mM. The signal intensity of 0.01 mM MB was so low that the model considered it to be noise.

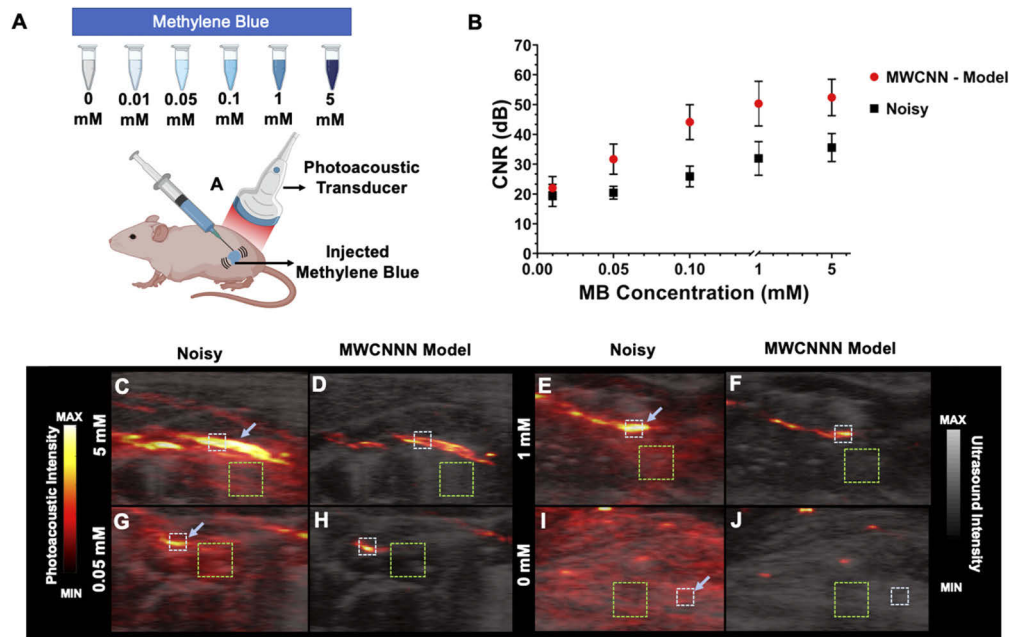


Fig. 6. In vivo evaluation of MWCNN model. A) Experimental schematic for *in vivo* evaluation of MWCNN model. Five different concentrations of MB (0.01, 0.05, 0.1, 1, and 5 mM) were injected intramuscularly. B) CNR versus injected MB concentrations for both noisy and MWCNN model. We noted 1.55-, 1.76-, 1.62-, and 1.48-fold improvement of CNR for 0.05, 0.1, 1.0, and 5.0 mM, respectively. Error bars represent the CNR among three different animals. For CNR calculation, μ_{object} and $\mu_{background}$ were defined as the average of five different areas of mean values of photoacoustic intensity at the injected area (ROI of $1 \times 1 \text{ mm}^2$) and around the injected area (ROI of $3 \times 3 \text{ mm}^2$), respectively. Term $\sigma_{background}$ is the average of all five standard deviations of background intensity. Panels C, E, G, and I are B-mode noisy photoacoustic images for 5.0, 1.0, 0.05, 0 mM, respectively. These images are overlaid on ultrasound data. D, F, H, and J) B-mode MWCNN photoacoustic images for 5, 1, 0.05, 0 mM, respectively. Dotted green and white rectangles represent the used ROIs for background and object. Blue arrows show the MB injection area.

4. Discussions and conclusions

CNNs have been widely utilized in computer vision, image processing, and medical imaging. However, deep learning has utility beyond image segmentation, object detection, and object

tracking. Here, we proposed a deep learning model that can learn to restore PA images at different low fluence configurations and samples. To ensure the scalability of our solution, we built our model based on a limited training process and evaluated it with different illumination sources on other sample types and materials.

We observed quantitative and qualitative enhancement results. The proposed model was completely blind to our test data. We could achieve up to 1.62- and 2.2-fold improvement in SSIM (**Fig. 3(B)** and **4(B)**) for low fluence laser source and LED, respectively. The model improved the PSNR by a factor of up to 2.25 and 2.1 for low fluence laser and LED, respectively. A higher number of training datasets can lead to improvement factors (SSIM, PSNR) that will be significantly higher. PSNR and SSIM calculations require a ground truth image. However, having this data is not feasible in most cases. To show that our proposed method can enhance other image quality metrics, we used the CNR to evaluate the penetration depth and *in vivo* data. The ground truth is not required with this metric, and the CNR will be measured using just a single frame. We showed that the MWCNN can improve the contrast as well (**Fig. 5(E)**, **6(B)**). Finally, we showed that this contrast improvement has value *in vivo* with contrast improved up to 1.76-fold.

Like other deep learning methods, our solution gains most of its computational cost at the training stage. The training cost can scale up as the training set grows. At runtime, the model can process each frame at 0.8 seconds, which is relatively faster than classical methods like BM3D (3.33 seconds). It is also similar to DL methods like low-dose CT CNN (2.05 seconds) [45].

Our training included a small set of frames from a laser source within a specific range of illumination fluencies. Such a small training set can facilitate a model that trains fast for practical solutions. On the other hand, normalizing the training data made the model independent of signal magnitude in the input. This independence guided the model to generically learn important spatial features of samples in PA images beyond the settings and configurations. Future work may also evaluate any improvements achieved via increasing numbers of training sets.

The tests introduced lower fluencies of illumination from different sources. The observations suggest that the model learned features to distinguish signal from noise regardless of the input image quality. Comparable results between laser and LED based inputs also suggests the utility of our solution among various imaging systems.

The next step of this work will focus on training and testing processes on *in vivo* samples including actual blood vessels and other exogenous contrast agents. We will also expand the model from a 2D framework to 3D data. In that regard, we can train models based on a stack of PA images to potentially improve the consistency of results along the axis and reduce the noise in 3D results as well as cross-sectional images. In such setting, we need to upgrade the structure of samples in our training data to better represent the complex 3D features. Also, increasing the dimensions of the model will inherently increase the amount of training data required to develop the models but may facilitate even more advanced *in vivo* imaging.

Appendix

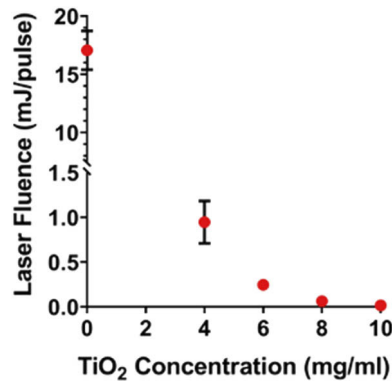


Fig. 7. Laser fluence vs TiO₂-based optical scattering concentrations.

Funding

National Institutes of Health (DP2 HL137187, OD S10021821, R21 AG06577, R21 DE029025); National Science Foundation (1842387, 937674).

Acknowledgements

JVJ acknowledges funding from the National Institutes of Health under grants R21 AG065776, R21 DE029025, and DP2 HL137187. We acknowledge NSF funding under grants 1842387 and 1937674. Infrastructure for this work was supported under NIH grant OD S10021821. Figures 1(A), 1(C), and 6A were created with BioRender.com.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. M. Xu and L. V. Wang, "Photoacoustic imaging in biomedicine," *Rev. Sci. Instrum.* **77**(4), 041101 (2006).
2. P. Beard, "Biomedical photoacoustic imaging," *Interface Focus* **1**(4), 602–631 (2011).
3. Y. Mantri, B. Davidi, J. E. Lemaster, A. Hariri, and J. V. Jokerst, "Iodide-doped Precious Metal Nanoparticles: Measuring Oxidative Stress in vivo via Photoacoustic Imaging," *Nanoscale* (2020).
4. T. J. Allen and P. C. Beard, "Pulsed near-infrared laser diode excitation system for biomedical photoacoustic imaging," *Opt. Lett.* **31**(23), 3462–3464 (2006).
5. L. V. Wang, *Photoacoustic Imaging and Spectroscopy* (CRC Press, 2017).
6. K. Wilson, K. Homan, and S. Emelianov, "Biomedical photoacoustics beyond thermal expansion using triggered nanodroplet vaporization for contrast-enhanced imaging," *Nat. Commun.* **3**, 1–10 (2012).
7. S. Telenkov, R. Alwi, A. Mandelis, and A. Worthington, "Frequency-domain photoacoustic phased array probe for biomedical imaging applications," *Opt. Lett.* **36**(23), 4560–4562 (2011).
8. Y. Wang, D. Xing, Y. Zeng, and Q. Chen, "Photoacoustic imaging with deconvolution algorithm," *Phys. Med. Biol.* **49**(14), 3117–3124 (2004).
9. M. Haltmeier, O. Scherzer, and G. Zangerl, "A reconstruction algorithm for photoacoustic imaging based on the nonuniform FFT," *IEEE Trans. Med. Imaging* **28**(11), 1727–1735 (2009).
10. M. Mozaffarzadeh, A. Hariri, C. Moore, and J. V. Jokerst, "The double-stage delay-multiply-and-sum image reconstruction method improves imaging quality in a led-based photoacoustic array scanner," *J. Photoacoust.* **12**, 22–29 (2018).
11. P. Omid, M. Zafar, M. Mozaffarzadeh, A. Hariri, X. Haung, M. Orooji, and M. Nasirivanaki, "A novel dictionary-based image reconstruction for photoacoustic computed tomography," *Appl. Sci.* **8**(9), 1570 (2018).
12. A. de La Zerda, Y. M. Paulus, R. Teed, S. Bodapati, Y. Dollberg, B. T. Khuri-Yakub, M. S. Blumenkranz, D. M. Moshfeghi, and S. S. Gambhir, "Photoacoustic ocular imaging," *Opt. Lett.* **35**(3), 270–272 (2010).

13. W. Liu and H. F. Zhang, "Photoacoustic imaging of the eye: a mini review," *J. Photoacoust.* **4**(3), 112–123 (2016).
14. A. Hariri, J. Wang, Y. Kim, A. Jhunjhunwala, D. L. Chao, and J. V. Jokerst, "In vivo photoacoustic imaging of chorioretinal oxygen gradients," *J. Biomed. Opt.* **23**(03), 1 (2018).
15. S. Mallidi, G. P. Luke, and S. Emelianov, "Photoacoustic imaging in cancer detection, diagnosis, and treatment guidance," *Trends Biotechnol.* **29**(5), 213–221 (2011).
16. A. Agarwal, S. Huang, M. O'donnell, K. Day, M. Day, N. Kotov, and S. Ashkenazi, "Targeted gold nanorod contrast agent for prostate cancer detection by photoacoustic imaging," *J. Appl. Phys.* **102**(6), 064701 (2007).
17. M. Mehrmohammadi, S. Joon Yoon, D. Yeager, and S. Y. Emelianov, "Photoacoustic imaging for cancer detection and staging," *Curr. Mol. Imaging* **2**(1), 89–105 (2013).
18. B. Zabihian, J. Weingast, M. Liu, E. Zhang, P. Beard, H. Pehamberger, W. Drexler, and B. Hermann, "In vivo dual-modality photoacoustic and optical coherence tomography imaging of human dermatological pathologies," *Biomed. Opt. Express* **6**(9), 3163–3178 (2015).
19. J. Kim, Y. Kim, B. Park, H. M. Seo, C. Bang, G. Park, Y. Park, J. Rhie, J. Lee, and C. Kim, "Multispectral ex vivo photoacoustic imaging of cutaneous melanoma for better selection of the excision margin," *Br. J. Dermatol.* **179**(3), 780–782 (2018).
20. A. Hariri, F. Chen, C. Moore, and J. V. Jokerst, "Noninvasive staging of pressure ulcers using photoacoustic imaging," *Wound Rep. Reg.* **27**(5), 488–496 (2019).
21. T. N. Erpelding, C. Kim, M. Pramanik, L. Jankovic, K. Maslov, Z. Guo, J. A. Margenthaler, M. D. Pashley, and L. V. Wang, "Sentinel lymph nodes in the rat: noninvasive photoacoustic and US imaging with a clinical US system," *Radiology* **256**(1), 102–110 (2010).
22. L. Song, C. Kim, K. Maslov, K. K. Shung, and L. V. Wang, "High-speed dynamic 3D photoacoustic imaging of sentinel lymph node in a murine model using an ultrasound array," *Med. Phys.* **36**(8), 3724–3729 (2009).
23. K. H. Song, C. Kim, K. Maslov, and L. V. Wang, "Noninvasive in vivo spectroscopic nanorod-contrast photoacoustic mapping of sentinel lymph nodes," *Eur. J. Radiol.* **70**(2), 227–231 (2009).
24. A. Hariri, A. Fatima, N. Mohammadian, S. Mahmoodkalayeh, M. A. Ansari, N. Bely, and M. R. Avanaki, "Development of low-cost photoacoustic imaging systems using very low-energy pulsed laser diodes," *J. Biomed. Opt.* **22**(7), 075001 (2017).
25. L. Zeng, G. Liu, D. Yang, and X. Ji, "Portable optical-resolution photoacoustic microscopy with a pulsed laser diode excitation," *Appl. Phys. Lett.* **102**(5), 053704 (2013).
26. T. Wang, S. Nandy, H. S. Salehi, P. D. Kumavor, and Q. Zhu, "A low-cost photoacoustic microscopy system with a laser diode excitation," *Biomed. Opt. Express* **5**(9), 3053–3058 (2014).
27. P. K. Upputuri and M. Pramanik, "Pulsed laser diode based optoacoustic imaging of biological tissues," *Biomed. Phys. Eng. Express* **1**(4), 045010 (2015).
28. A. Hariri, J. Lemaster, J. Wang, A. S. Jeevarathinam, D. L. Chao, and J. V. Jokerst, "The characterization of an economic and portable LED-based photoacoustic imaging system to facilitate molecular imaging," *J. Photoacoust.* **9**, 10–20 (2018).
29. Y. Zhu, G. Xu, J. Yuan, J. Jo, G. Gandikota, H. Demirci, T. Agano, N. Sato, Y. Shigeta, and X. Wang, "Light emitting diodes based photoacoustic imaging and potential clinical applications," *Sci. Rep.* **8**(1), 1–12 (2018).
30. R. S. Hansen, "Using high-power light emitting diodes for photoacoustic imaging," in *Medical Imaging 2011: Ultrasonic Imaging, Tomography, and Therapy*, 7968 (International Society for Optics and Photonics, 2011), 79680Y.
31. Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. R. Soc. London, Ser. A* **460**(2046), 1597–1611 (2004).
32. Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt. Data Anal.* **01**(01), 1–41 (2009).
33. S. R. Messer, J. Agzarian, and D. Abbott, "Optimal wavelet denoising for phonocardiograms," *Microelectron. J.* **32**(12), 931–941 (2001).
34. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. on Image Process.* **9**(9), 1532–1546 (2000).
35. C. V. Sindelar and N. Grigorieff, "An adaptation of the Wiener filter suitable for analyzing images of isolated single particles," *J. Struct. Biol.* **176**(1), 60–74 (2011).
36. G. Redler, B. Epel, and H. J. Halpern, "Principal component analysis enhances SNR for dynamic electron paramagnetic resonance oxygen imaging of cycling hypoxia in vivo," *Magn. Reson. Med.* **71**(1), 440–450 (2014).
37. T. Yoshida, R. Miyamoto, M. Takada, and K. Suzuki, "Adaptive noise canceller," (Google Patents, 1995).
38. R. Manwar, M. Hosseinzadeh, A. Hariri, K. Kratkiewicz, S. Noei, N. Avanaki, and R. Mohammad, "photoacoustic Signal enhancement: towards utilization of Low energy laser diodes in Real-time photoacoustic imaging," *Sensors* **18**(10), 3498 (2018).
39. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012), 1097–1105.
40. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
41. A. Arafati, P. Hu, J. P. Finn, C. Rickers, A. L. Cheng, H. Jafarkhani, and A. Kheradvar, "Artificial intelligence in pediatric and adult congenital cardiac MRI: an unmet clinical need," *Cardiovasc. Diagn. Ther.* **9**(S2), S310–S325 (2019).

42. A. Arafati, "Segmentation and Tracking of Echocardiograms Using Deep Learning Algorithms," (UC Irvine, 2019).
43. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 3431–3440.
44. E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.* **44**(10), e360–e375 (2017).
45. H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomed. Opt. Express* **8**(2), 679–694 (2017).
46. W. Yang, H. Zhang, J. Yang, J. Wu, X. Yin, Y. Chen, H. Shu, L. Luo, G. Coatrieux, and Z. Gui, "Improving low-dose CT image using residual convolutional network," *IEEE Access* **5**, 24698–24705 (2017).
47. E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denoising for low-dose CT via wavelet residual network," *IEEE Trans. Med. Imaging* **37**(6), 1358–1369 (2018).
48. Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imaging* **37**(6), 1348–1357 (2018).
49. J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imaging* **36**(12), 2536–2545 (2017).
50. X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *J Digit Imaging* **31**(5), 655–669 (2018).
51. H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imaging* **36**(12), 2524–2535 (2017).
52. F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122 (2015).
53. P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in *2018 IEEE winter conference on applications of computer vision (WACV)*, (IEEE, 2018), 1451–1460.
54. M. A. L. Bell, N. Kuo, D. Y. Song, and E. M. Boctor, "Short-lag spatial coherence beamforming of photoacoustic images for enhanced visualization of prostate brachytherapy seeds," *Biomed. Opt. Express* **4**(10), 1964–1977 (2013).
55. B. Pourebrahimi, S. Yoon, D. Dopsa, and M. C. Kolios, "Improving the quality of photoacoustic images using the short-lag spatial coherence imaging technique," in *Photons Plus Ultrasound: Imaging and Sensing 2013*, (International Society for Optics and Photonics, 2013), 85813Y.
56. E. R. Hill, W. Xia, M. J. Clarkson, and A. E. Desjardins, "Identification and removal of laser-induced noise in photoacoustic imaging using singular value decomposition," *Biomed. Opt. Express* **8**(1), 68–77 (2017).
57. A. Reiter and M. A. L. Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in *Photons Plus Ultrasound: Imaging and Sensing 2017* (International Society for Optics and Photonics, 2017), 100643J.
58. D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Trans. Med. Imaging* **37**(6), 1464–1477 (2018).
59. E. M. A. Anas, H. K. Zhang, J. Kang, and E. Boctor, "Enabling fast and high quality LED photoacoustic imaging: a recurrent neural networks based approach," *Biomed. Opt. Express* **9**(8), 3852–3866 (2018).
60. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse Probl. Sci. Eng.* **27**(7), 987–1005 (2019).
61. E. M. A. Anas, H. K. Zhang, J. Kang, and E. M. Boctor, "Towards a Fast and Safe LED-Based Photoacoustic Imaging Using Deep Convolutional Neural Network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer, 2018), 159–167.
62. P. Liu, H. Zhang, W. Lian, and W. Zuo, "Multi-level wavelet convolutional neural networks," *IEEE Access* **7**, 74973–74985 (2019).
63. S. E. Bohndiek, S. Bodapati, D. Van De Sompel, S.-R. Kothapalli, and S. S. Gambhir, "Development and application of stable phantoms for the evaluation of photoacoustic imaging instruments," *PLoS One* **8**(9), e75533 (2013).
64. J. Yan, J. Schaefferkoetter, M. Conti, and D. Townsend, "A method to assess image quality for low-dose PET: analysis of SNR, CNR, bias and image noise," *Cancer Imaging* **16**(1), 26 (2016).