Research Article | Vol. 15, No. 8 / 1 Aug 2024 / Biomedical Optics Express 4390
Biomedical Optics EXPRESS

Unsupervised denoising of photoacoustic images based on the Noise2Noise network

YANDA CHENG,[1] WENHAN ZHENG,[1] ROBERT BING,[1] HUIJUAN ZHANG,[1] CHUQIN HUANG,[1] PEIZHOU HUANG,[1] LESLIE YING,[1,2] AND JUN XIA[1,*]

[1]*Department of Biomedical Engineering, University at Buffalo, The State University of New York, Buffalo, New York, USA*
[2]*Department of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, New York, USA*
[*]*junxia@buffalo.edu*

Abstract: In this study, we implemented an unsupervised deep learning method, the Noise2Noise network, for the improvement of linear-array-based photoacoustic (PA) imaging. Unlike supervised learning, which requires a noise-free ground truth, the Noise2Noise network can learn noise patterns from a pair of noisy images. This is particularly important for in vivo PA imaging, where the ground truth is not available. In this study, we developed a method to generate noise pairs from a single set of PA images and verified our approach through simulation and experimental studies. Our results reveal that the method can effectively remove noise, improve signal-to-noise ratio, and enhance vascular structures at deeper depths. The denoised images show clear and detailed vascular structure at different depths, providing valuable insights for preclinical research and potential clinical applications.

1. Introduction

Photoacoustic (PA) imaging is a biomedical imaging method that combines the advantages of high-contrast optical imaging and high-resolution ultrasonic imaging [1]. PA imaging is based on the PA effect, which involves the absorption of short-pulsed laser light by biomolecules such as hemoglobin, lipid, or melanin, causing local thermoelastic expansion and generating ultrasonic acoustic waves [2]. The acoustic signals are detected by ultrasound transducer arrays and then digitized for image reconstruction.

Endogenous contrasts, such as hemoglobin, are used in PA due to their high optical absorption coefficients at specific wavelengths. This allows PA imaging to visualize deep underlying structures, making it a valuable tool in various preclinical and clinical applications. The main advantage of PA imaging is its ability to provide high-resolution and high-contrast images with deep penetration, which makes it particularly useful for biomedical applications such as brain imaging, tumor detection, and foot ulcer imaging [3–5]. However, since the PA signals are much weaker than pulse-echo ultrasound, PA imaging also faces challenges in signal-to-noise ratios (SNR). The reduced SNR is particularly evident at increasing imaging depths due to light attenuation. In addition to thermal noise, electromagnetic interference (EMI) from the excitation laser might also affect the SNR [6]. These factors collectively influence the quality of PA images, resulting in challenges in clinical translation [7]. Several methods have been proposed to improve SNRs. Traditional approaches include median filtering, Gaussian filtering, and wavelet filtering [8–10]. However, these methods may not always be effective in removing all noise sources and may result in image quality losses [11,12].

Recently, machine-learning-based methods have also been widely used for denoising. These methods can be classified as either supervised learning or unsupervised learning. Supervised learning is a prevalent approach in the field of image processing and computer vision for denoising

#529253 https://doi.org/10.1364/BOE.529253
Journal © 2024 Received 6 May 2024; revised 29 May 2024; accepted 15 Jun 2024; published 2 Jul 2024

Biomedical Optics EXPRESS

tasks [13–15]. This approach involves training a neural network using a labeled dataset of noisy and clean image pairs [16,17]. The neural network is trained to predict a clean image as output by giving the noisy image as input. However, a primary challenge with supervised learning is the need for noise-free ground truth, which is impossible to obtain for in-vivo experiments [18,19]. Due to this limitation, most supervised denoising algorithms in PA utilized simulation data generated by the MATLAB-based K-wave toolbox [20]. The algorithm's success relies heavily on the accuracy of the simulation and the prior knowledge of noise patterns in the experimental system [18]. Therefore, the trained model performs poorly when the simulated noisy pattern cannot fully mimic the experimental noise [21].

On the other hand, unsupervised learning does not require clean experimental data for its training process [22–25]. It can be effectively trained using only experimental data, eliminating the need for simulated data generation. The Noise-2-Noise (N2N) network is an example of unsupervised learning. N2N uses neural networks to find patterns in pairs of noisy images, each with its own noise level or type. This approach allows for efficient noise removal and versatility in different scenarios, making it particularly useful for in vivo studies where the noise-free ground truth is absent [26–29]. The N2N network has been successfully applied in various imaging modalities, including Magnetic Resonance Imaging (MRI) [30,31], Computed Tomography (CT) imaging [32], and Ultrasound [32]. Most N2N studies need two images captured under identical conditions but with different noise levels [33]. This strategy is impractical for in vivo PA imaging due to body movement.

To resolve this issue, we developed a data generation method that can provide a pair of low-noise and high-noise images from a single in vivo dataset. The ultimate goal is to enhance the visibility of deep vessels in PA images, as light attenuation makes these vessels hard to differentiate from the background noise. To validate our method, we first trained a simulation model using simulated data incorporating experimental noise. As ground-truth information is available in the simulation dataset, we can thoroughly verify our approach. Then, another independent experimental model was trained using only in-vivo data, and we quantified the performance through SNR analysis. Our results demonstrate the high potential of the model for denoising PA in vivo images.

2. Methods

2.1. System setup

The photoacoustic images were acquired from the dual-scan foot photoacoustic imaging system [5]. The system consists of a customized linear-array transducer (Imasonic SAS, France) with 128 elements and a 2.25 MHz central frequency. A portable Nd: YAG laser (Big Sky Laser) is utilized as the light source. The laser operates at a frequency of 10 Hz with a pulse duration of approximately 8 ns and an emission wavelength of 1064 nm. The PA signals are captured by a 256-channel data acquisition unit (PhotoSound Technologies Inc.) operating at a 40 MHz sampling rate. The synchronization between the data capturing and laser pulse is achieved through trigger signals from the laser. The entire mobile platform comprises a scanning head connected to a translation stage and driven by a stepper motor to perform linear scanning.

2.2. Deep-learning models

2.2.1. Noise2Noise model

In the image-denoising task, the general idea is to take a noisy image as $\gamma = \chi + \epsilon$, where γ is the noisy image, χ is the clean photoacoustic image, and ϵ is random variable noise. To estimate the unknown χ, we can collect a n set of noisy photoacoustic images $[\gamma_1, \gamma_2 \ldots, \gamma_n]$ from the same fixed region. In supervised learning, the most common way is to find the estimated average $\hat{\chi}$ by minimizing the expectation ($\mathbb{E}$) of deviation from the observations, as determined by a specific

Biomedical Optics EXPRESS

loss function (L). The loss function can be described below as Eq. (1):

$$\arg\min_{\hat{\chi}} \mathbb{E}_{\gamma}\{L(\gamma, \hat{\chi})\} \tag{1}$$

Here, L2 loss is the most commonly used loss function and can be defined as: $L(\gamma, \hat{\chi}) = (\chi - \gamma)^2$. Then, a supervised learning neural network f with network parameter θ can retrieve clean image χ from training pairs (γ_{i} , χ_{i}) by solving the point estimation problem [34]. The training process can be expressed in a mathematical way as Eq. (2):

$$\arg\min_{\theta} \mathbb{E}_{(\gamma,\ \chi).}\{\mathrm{L}(f_{\theta}(\gamma), \chi)\} \tag{2}$$

This method requires a clean ground truth χ for training, and it only works well when the dataset is large enough with well-labeled ground truth. However, if the noise ϵ, such as Gaussian noise, has an expectation with a mean of zero, Eq. (2). can be rewritten and decomposed into Eq. (3). The output result remains unchanged, as shown in Eq. (3).:

$$\arg\min_{\theta} \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (f_{\theta}(\chi_{\mathrm{i}} + \epsilon_{\mathrm{i}a}) - (\chi_{\mathrm{i}} + \epsilon_{\mathrm{i}b}))^2 \tag{3}$$

Here, χ_{i} is the unknown noise-free image of scene i, and $\epsilon_{\mathrm{i}a}$, and $\epsilon_{\mathrm{i}b}$ are two independent noises from the same scene. They may have different amplitudes, but their mean value should be zero. The network f_{θ} determines the average of all potential functions by minimizing the L2 loss in Eq. (3), enabling it to effectively separate the clean image χ from noisy image γ_{i} pairs [26]. Eventually, the network will approximate the target expectation $\hat{\chi}$, similar to the supervised method. This concept has been utilized for training deep denoising models when a clean target is absent [29].

2.2.2. Unet model

For comparison, we also trained a CNN model with a U-Net structure specifically designed for medical image denoising [35–37]. This supervised learning model requires labeled datasets for training. For experimental results, since we do not have the noise-free ground truth, we use the reconstructed images as targets and noise-enhanced images as inputs. Here, the noise-ehanced images were obtained by adding experimental noise to the reconstructed data, similar to the method proposed in [38].

2.3. Model training

We developed models based on the N2N concept using PyTorch, training two separate models for simulation and experimental datasets, respectively. Both models used the same training strategy, each adapted to its respective dataset type. As shown in Fig. 1, the network has five convolution pooling layers, which have the same size as the convolution up-sampling layers [39–41]. The construction of a deep regressor between the input and output is achieved by utilizing the U-Net architecture structure [42,43]. We configured the input and output channels to handle single-channel image data (grey-scale image) with dimensions of $400 \times 860 \times 1$. Here 400 refers to the pixel number along axial direction while 860 refers to the pixel number along elevation direction. No batch normalization or dropout techniques are employed during training, but skip connections are utilized to improve the resolution of the final output images [44,45].

LeakyReLU is preferred over ReLU in our model because it allows for a small gradient when the neuron's activation is negative, preventing neurons from dying—a common issue with ReLU, where neurons stop learning [46–48]. MaxPool2d is used for its downsampling capability, which reduces computational complexity and helps the network extract and focus on the most salient

Biomedical Optics EXPRESS

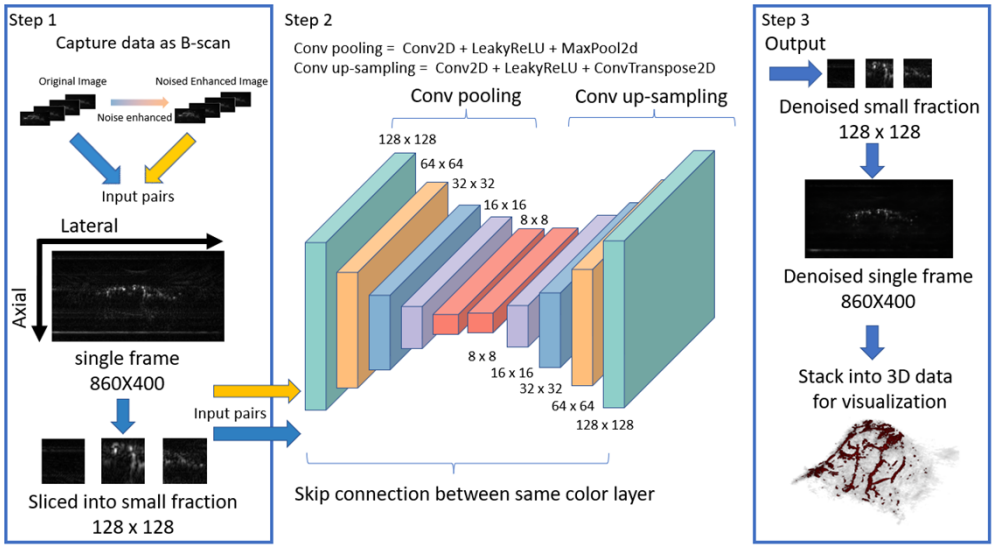

Fig. 1. Overall data generation and training structure of the Noise2Noise Network. Step 1: Prepare the input pairs from the original B-scan images. Step 2: Feed image pairs into the trained network model. Step 3: Output the result and stack images for 3D visualization.

features, which is essential for distinguishing noise from the signal. ConvTranspose2D serves to upsample the feature maps, increasing their spatial resolution for the reconstruction of the denoised image, effectively reversing the downsampling effect of MaxPool2d. Skip connections in the U-Net architecture, as shown in Fig. 1, enhance image resolution by facilitating the flow of information from earlier to later layers within the network. It provides a direct connection between the down-sampled and up-sampled feature maps, resulting in improved resolution and performance.

The model does not include batch normalization or any other regularization techniques, despite their proven effectiveness in supervised learning tasks like classification and segmentation. This is because in studies involving super-resolution and denoising, where noise and signal intensity are similar, including a batch normalization layer can sometimes result in undesirable outcomes [49]. Specifically, it may fail to improve the final output and instead blend the noise and signal together, thereby diminishing the effectiveness of the training process. Meanwhile, batch normalization will increase the training time significantly.

For unsupervised learning in our model, these layers enable the network to learn robust feature representations from noisy data without requiring clean targets, facilitating the reconstruction of high-quality images. The combination of LeakyReLU with MaxPool2d and ConvTranspose2D allows the network to handle a wide range of noise patterns and restore intricate details, improving the overall denoising process. For PA imaging, the major two noise types are EMI and thermal noise. The later is presented as Gaussian noise.

We set the learning rate as 0.001 for training the simulation dataset and the in-vivo dataset. The training was conducted in a workstation with AMD Ryzen 9 3950X CPU, 128 GB RAM, and two NVIDIA GTX3090 Ti graphics cards.

Figure 1 illustrates the overall workflow of the Noise2Noise network. The process starts with the captured B-scan images, which are then processed to generate noise-enhanced input pairs. These pairs undergo a series of convolutional operations, including down-sampling and up-sampling steps linked via skip connections to preserve detail across layers. The final output is stacked to create a 3D volume, which highlights the vascular structure.

Biomedical Optics EXPRESS

2.4. Dataset

2.4.1. Simulation dataset

First, we trained a model based on simulation data. The ground-truth from simulation allowed us to validate our training strategy before training another model for the experimental dataset.

In simulation dataset generation, the three-dimensional vascular matrices were generated by the Insight Segmentation and Registration Toolkit (ITK) [50]. After setting the number of branches, nodes, direction, and vessel volume, the toolkit generates the formation of vessels in 3D space with varying diameters and densities. This 3D image is used as the input for K-wave simulation, and we assume that the pixel amplitude of each pixel is proportional to the optical absorption coefficient. To create PA data from the vascular patterns, we utilized the K-Wave toolbox, a MATLAB-based tool for photoacoustic simulation [51]. The arc-shaped transducer is assumed to scan along the elevational direction with step size of 0.1 mm. The B-scan image is generated by assembling A-lines sequences. Detailed simulation processes can be found in previous work [52]. The transducer's height and pitches are 15 mm and 0.67 mm, respectively; the acoustic focus of the transducer is set at 40 mm; the transducer bandwidth is 65%; the sampling rate is 9 MHz; the central frequency of the transducer is 2.25 MHz; and the transducer scanning step size is 0.1 mm. These simulation parameters are similar to those of the experimental transducer. In this study, we generated 20 volumetric vessel data, data was simulated with size of $50 \times 50 \times 50$ mm at a pixel resolution of 0.1 mm. Then, we sliced the 3D data into cross-sectional views. A total of 10,000 (500×20) cross-sectional images were generated for the training. Figure 2(a) shows the maximum amplitude projection (MAP) of one noise-free data, which serves as the ground truth.

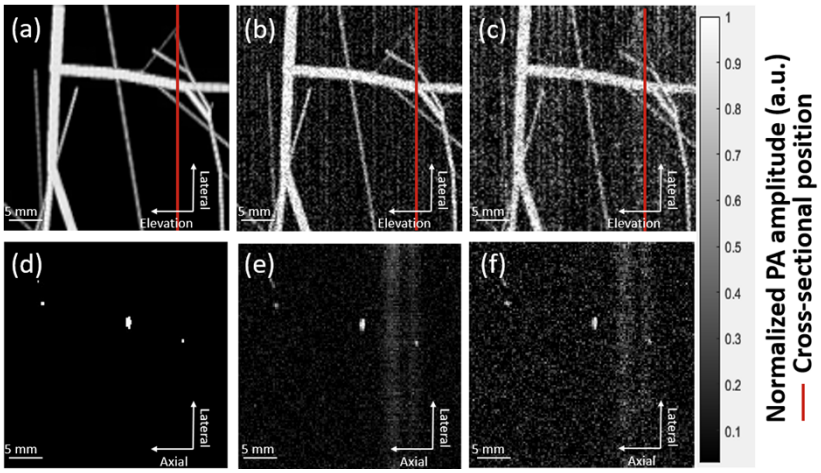

Fig. 2. Stimulated training dataset. (a) The ground truth data in MAP view, (b) Training data with minimal experimental noise in MAP view (c) Training data with enhanced experimental noise in MAP view. (d)-(f) are axial-lateral cross-sectional views along the red line in a-c, respectively.

For training, we generated a pair of noisy images. To mimic the experimental condition, noise captured from the experimental setup was randomly added to the simulated 3D data. To ensure that the two images have different levels of noise, we applied different weighting factors to the experimental noise before adding them to the simulated data. To better mimic the low SNR at deep tissue, we further supplemented the data with Gaussian noise, using MATLAB functions (awgn) [53,54]. Exemplary input pairs are shown in Figs. 2(b) and 2(c), and the corresponding cross-section views are shown in Figs. 2(e) and 2(f), respectively.

Biomedical Optics EXPRESS

In this study, we reserved 10% of the simulation data as the testing set to evaluate the performance of the simulation model. The 10% division was chosen because unsupervised learning focuses on evaluating the model's ability to discover patterns and relationships within the data rather than making predictions. Using a smaller testing set, compared to the typical 20% required for supervised learning, has a significant advantage when the in-vivo data size is limited.

2.4.2. Experimental dataset

After the model was validated in simulation testing dataset, we further trained another model using experimental dataset only, where ground truth was not available. All experimental data was captured by the dual-scan photoacoustic foot imaging system [5]. We selected data from 20 subjects and 5 phantoms. All subject data was captured at 2.7W, while the phantom data was captured using five different optical fluence maximum power outputs, ranging from level 5 to level 1: 2.7W, 2.56W, 1.44W, 1.01W, and 0.65W.

The pencil lead phantom was constructed using a mixture of 4% agar gel and 96% water, designed to mimic human soft tissue's acoustic and optical properties [55]. Within this phantom, four pencil leads, each with a diameter of 0.5 mm, were embedded in the gel to simulate the vessels beneath the human skin. These pencil leads were spatially separated by 10 mm along both elevation and axial directions, as shown in Fig. 3.

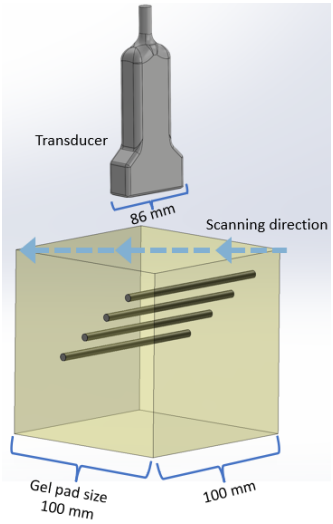

Fig. 3. Pencil lead phantom with each lead spaced 10 mm apart along the elevation and axial directions. The ultrasound transducer scans the phantom from the top.

In this study, our network training begins with experimental datasets and their corresponding noisy pairs. Here, the original image refers to the initial B-scan image captured from the system, and the noisy pair is generated by adding noise to the original image. To ensure the network's robustness in handling diverse noise patterns encountered in practical scenarios, we incorporated data with different noise levels to generate noisy reconstructed data. All in-vivo data reconstruction starts at a depth right above the skin. After the scan, we utilized back projection to reconstruct each B-mode image. Then, the reconstructed images were stacked along the elevation direction to form a 3D image for maximum amplitude project (MAP). For denoising purposes, we denoised each reconstructed B-mode image before stacking them into the 3D form [56]. Each 3D image has a matrix size of 400 (axial) $\times$ 860 (lateral) $\times$ 500 (elevational).

The complete data generation process is illustrated in Fig. 4. First, we randomly selected experimental noise from different regions of the original data. Then, we applied a random

Biomedical Optics EXPRESS

weighting factor to the noise and incorporated it into the data to generate the noisy pair. The random weighting factor further enhances the diversity of our dataset. In the last step, we reconstructed the original image and the corresponding noisy pair before feeding them into the network for training. Because the noisy pair was generated by adding noise to the original data, we can ensure it has a lower SNR than the original one. The whole dataset is made up of 12,500 (500×25) unique B-scans, and the noisy pair matches this with another 12,500 B-scans, leading to a total of 25,000 images as training pair for model. It should be noted that both the original data and its noisy pair contain experimental noises. Unlike supervised learning, there are no noise-free images in our training input.

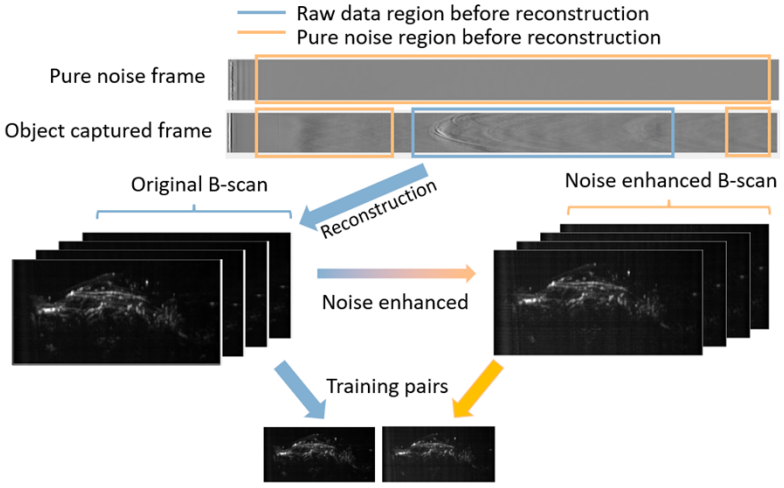

Fig. 4. Dataset generation flowchart. The top row shows the raw PA data, where the object-containing region (blue) is combined with the pure noise region (orange) to generate the noise-enhanced image. The bottom row shows the reconstructed PA images before and after noise enhancement.

During the training phase, we first normalized and upsized each B-scan image to 1024×1024 pixels. Then, these images were partitioned into smaller segments, each with 128×128 pixels. To ensure continuity, we intentionally overlapped these segments by 50% to prevent any discontinuities in the result. Furthermore, we used an adaptive histogram algorithm to improve the visualization of our image outputs. This technique was especially helpful for enhancing the visibility of deeper tissue vessels, which are typically harder to distinguish due to their darker appearance.

2.5. Evaluation metric for simulation and experimental data

To quantitatively evaluate the performance of the proposed N2N network with the simulation dataset, we used the Structural Similarity Index (SSIM), Mean Absolute Error (MAE), and Peak Signal-to-Noise Ratio (PSNR) as shown in Table 1 [6]. The mean squared error (MSE) metric measures the average difference between the actual and predicted pixel values of the denoised images. The lower the MSE, the more accurate the network's denoising ability. The PSNR is calculated between the simulated ground truth, noisy input image, Gaussian denoised output, and network denoised output result [6]. The PSNR function is described as:

$$PSNR = 10 \times \log\left(\frac{MAX^2}{MSE}\right)$$

Biomedical Optics EXPRESS

where MAX is the maximum possible pixel value of the image, and PSNR is expressed in decibels (dB). Third, SSIM quantifies the similarity between original and processed images, assessing structural integrity to evaluate denoising effectiveness. The result is represented on a scale of 1 [57–59].

As both the in-vivo and phantom data were captured using the same experimental setup, we assume they exhibit similar noise patterns. For the experimental dataset, we have 20 human subject data and 5 phantom data as the training dataset. Data from 2 more subjects and 1 phantom were used as the testing dataset. For the phantom test data, we treat the maximum power captured results as the ground truth and the low power captured as noisy input. The performance of the network was evaluated using full width at half maximum (FWHM), PSNR, SSIM, and MSE. For the in-vivo test data, where ground truth is unavailable, we evaluate our results by measuring the SNR across various depths within selected regions of interest (ROI).

3. Results

3.1. Validation with simulation data

We first evaluated the performance of the simulation-data-trained model. Figure 5(a) displays a noise-free image, which serves as the ground truth. Figure 5(b) is the corresponding noisy image input. The noise intensity is occasionally higher than the vessel signal, which creates a challenge for traditional denoising methods, such as the Gaussian filter [28,60]. As shown in Fig. 5(c), the small vessels were blurred, and the difference between the noise and signal could not be distinguished. In contrast, our network was able to successfully detect the differences between the vessel signals and the noisy signals and generate a clean denoised image, as shown in Fig. 5(d).

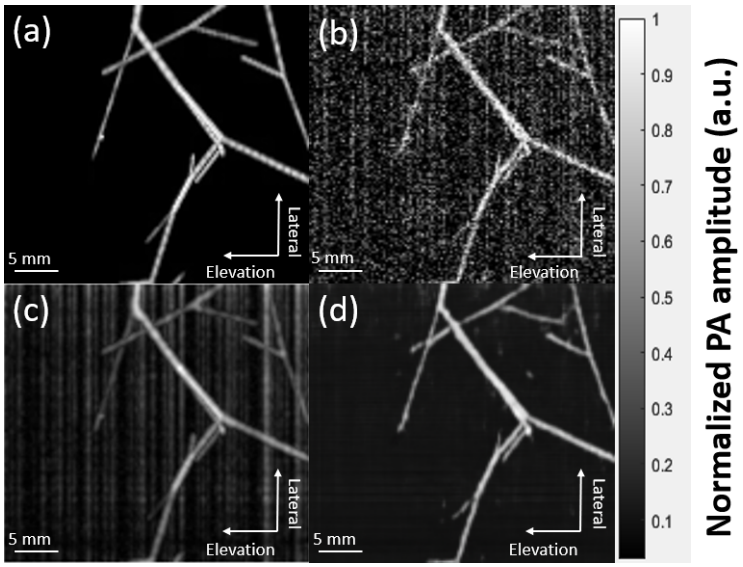

Fig. 5. Simulation testing dataset. (a): MAP of the noise-free data as ground truth. (b): The noisy image was generated from the ground truth image by adding experimental noise. (c): Gaussian filter denoised result. (d) Noise2Noise denoised result.

Comparison between the two denoising outputs demonstrates that the N2N network outperforms Gaussian denoising filters in terms of both PSNR and SSIM. This performance met our expectations as the network was trained using different levels of experimental noise, while the Gaussian filter simply smoothed the image.

Biomedical Optics EXPRESS

As Table 1 shows, the PSNR value of the N2N-denoised image is significantly higher than those of the Gaussian filter-denoised images and the noisy input. The results demonstrate that the network's denoising performance significantly surpasses traditional methods. The N2N method has an SSIM of 0.841. This shows a major improvement in keeping the structure and texture clear in denoised images. Higher SSIM values mean the denoised image is more similar to the original. In terms of MSE, the N2N network's MSE value is 13 times lower than that of Gaussian filtering. This significantly lower MSE indicates a closer resemblance of the denoised image to the original, demonstrating a more effective denoising process.

Table 1. Evaluation metrics: PSNR, SSIM, and MSE of simulation results

Image	PSNR	SSIM	MSE
Noisy Input	5.37	0.045	0.033
Gaussian Denoised	6.65	0.506	0.093
N2N Denoised	15.0	0.841	0.007

3.2. Phantom results

After validating the performance of the simulation-data-trained model, we proceeded to test the performance of our experimental-data-trained model, starting with the pencil lead phantom data. In the pencil-lead phantom study, Fig. 6(a) shows the MAP result along the axial-lateral plane. The image was acquired with maximum laser power at 2.7W. Figure 6(b) shows the corresponding result along the elevation-lateral plane. It can be seen that the pencil leads are spaced 10 mm apart, spanning depths from 50 mm to 80 mm, with the fourth lead (red color) positioned at the 80 mm depth. During the experiment, in order to visualize the deepest pencils at max power, we increased the laser power, which resulted in a slight signal saturation in the top pencil lead (blue). The noisy image, obtained with lower laser power at 1.83W, is shown in Figs. 6(c) and 6(d). The noisy image is not saturated and revealed a less blurry image for the top pencil lead. This power output is chosen since it allows the fourth pencil lead (orange) to be barely visible due to strong EMI noise (white arrows) in the background. The unet-denoised images in Figs. 6(e) and 6(f) demonstrate some improvement. However, the noises are not completely removed. The N2N denoised images in Figs. 6(g) and 6(h) exhibit marked improvements in clarity. The pencil lead at the 80 mm depth becomes notably more visible, with a considerable noise reduction.

Table 2. Network performance analysis of pencil lead at 50 and 80 mm imaging depths.

Pencil lead depth (mm)	Image	FWHM (mm)	FWHM Diff %	PSNR (dB)	SSIM	MSE
50	Max Power	1.25	None	None	None	None
50	N2N Denoised	1.17	6.40%	27.3	0.81	0.043
50	Unet Denoised	1.12	10.40%	26.8	0.73	0.046
50	Noisy	1.01	19.2%	19.0	0.35	0.112
80	Max Power	2.31	None	None	None	None
80	N2N Denoised	2.14	7.36%	24.3	0.80	0.061
80	Unet Denoised	2.65	14.72%	23.5	0.73	0.067
80	Noisy	2.77	19.91%	18.0	0.57	0.127

In addition, as shown in Table 2, the dimension of pencil leads also closely matches the maximum power result based on FWHM quantification. The FWHM of the 50 mm pencil lead is slightly larger at max power due to the saturation effect we mentioned earlier. At depths of 80 mm, the FWHM value of the N2N denoised image is close to that observed in the max power

Biomedical Optics EXPRESS

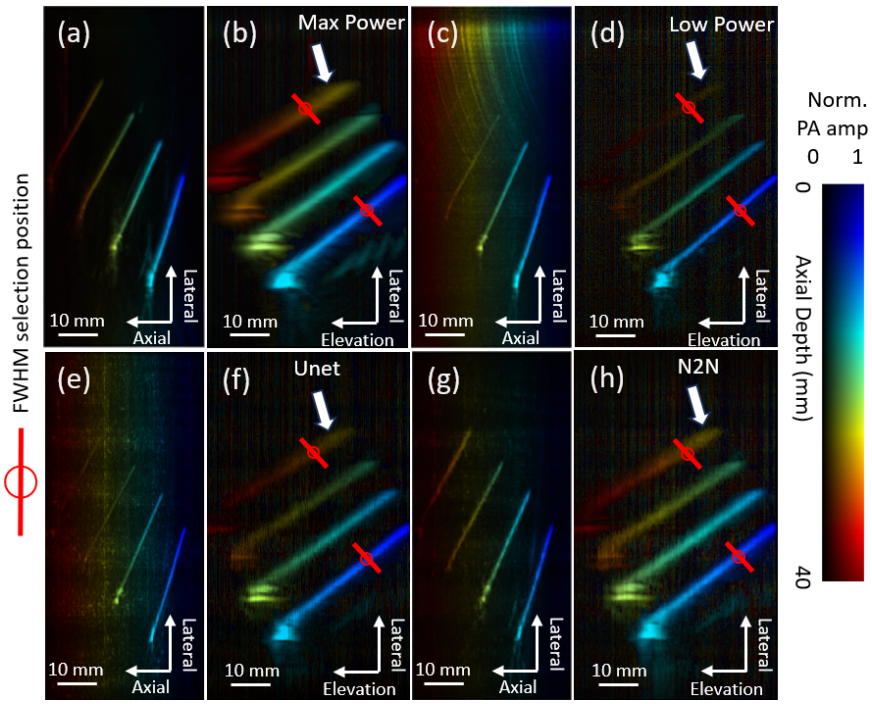

Fig. 6. Phantom imaging results. Maximum laser power result of the pencil lead in (a) axial-lateral projection and (b) elevation-lateral projection. (c) and (d) represent corresponding images acquired at low laser power. (e) and (f) represent Unet-denoised images of (c) and (d), respectively. (g) and (h) represent N2N denoised images of (c) and (d), showcasing significant contrast improvement at deeper depths.

image, as shown in Fig. 6(b). This indicates a significant denoising performance from the noisy images, demonstrating that the denoising process effectively aligns the results more closely with the ideal max power outcomes.

Furthermore, the PSNR, SSIM, and MSE values collectively assess the denoising performance. The PSNR improved by approximately 39%, signifying significant noise reduction. The 85% improvement in SSIM indicates a significant improvement in the denoised image's structural integrity. Meanwhile, the approximately 57% decrease in MSE indicates that the denoised image is much closer to the original in terms of pixel values. These findings highlight the effectiveness of our denoising approach in reducing noise, enhancing structural similarity, and minimizing image distortion post-denoising.

3.3. In-vivo foot result

After validation of the pencil lead results, we further tested the model on in-vivo data acquired from two human subjects. Since there is no ground truth for in-vivo data, we evaluated the performance based on SNR over different imaging depths.

Figures 7 and 8 demonstrate the in-vivo result of two human subjects. Figures 7(a) and 8(a) demonstrate cross-sectional views of the feet for subjects 01 and 02, respectively, while Figs. 7(b) and 8(b) show the MAP views of their feet. These images reveal the presence of experimental noise, particularly EMI and Gaussian-like noise, which causes noticeable discontinuity in the vascular structures.

Biomedical Optics EXPRESS

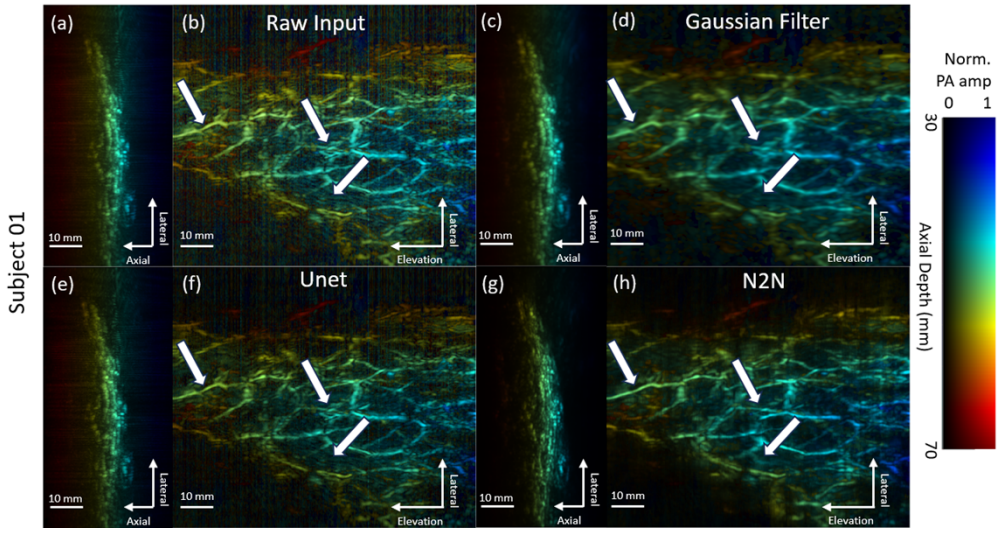

Fig. 7. In vivo human foot imaging results of Subject 01. (a) Axial-lateral projection image from Subject 1. (b) Elevation-lateral projection image from Subject 1. (c) Gaussian filter denoised result of (a). (d) Gaussian filter denoised result of (b). (e) Unet denoised result of (a). (f) Unet denoised result of (b). (g) Network denoised result of (a). (h) N2N denoised result of (b). All images are depth encoded with color.

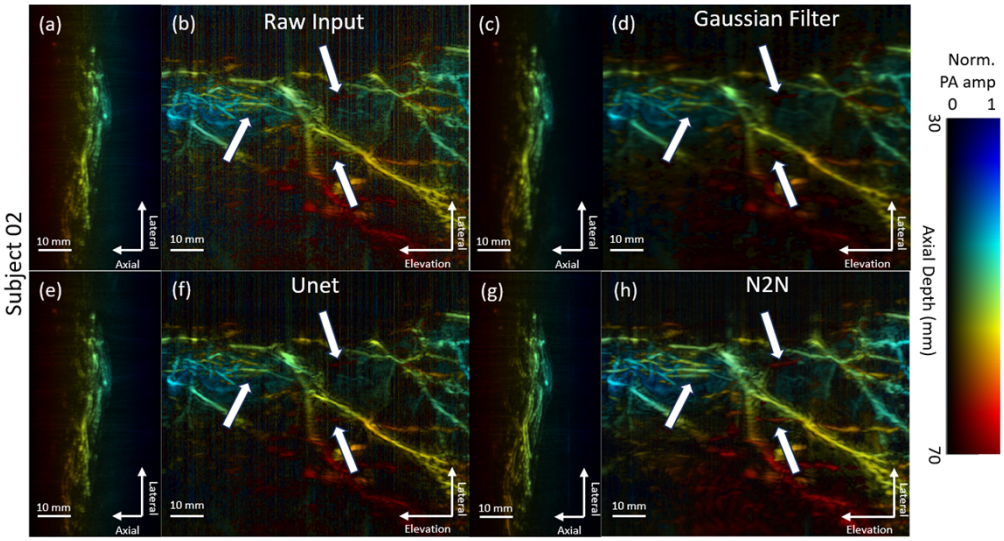

Fig. 8. In vivo human foot imaging results of Subject 2. (a) Axial-lateral projection image from Subject 2. (b) Elevation-lateral projection image from Subject 2. (c) Gaussian filter denoised result of (a). (d) Gaussian filter denoised result of (b). (e) Unet denoised result of (a). (f) Unet denoised result of (b). (g) Network denoised result of (a). (h) N2N denoised result of (b). All images are depth encoded with color.

Biomedical Optics EXPRESS

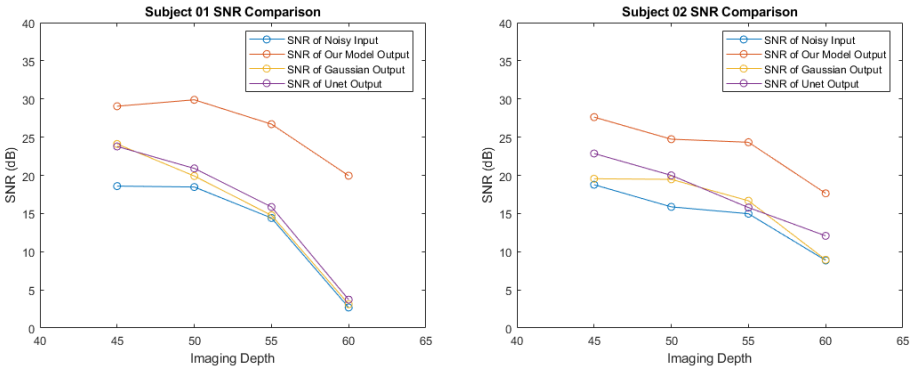

Fig. 9. Average SNR over depth between the noisy raw input, Gaussian filter result, Unet, and N2N denoised result.

The Gaussian denoising results are displayed in Figs. 7(c) and 8(c) for the cross-sectional views and Figs. 7(d) and 8(d) for the MAP views for subjects 01 and 02, respectively. The noise has been removed, but the vessel boundaries have become blurry. The Unet denoised results for subjects 01 and 02 are shown in Figs. 7(e) and 8(e) for the cross-sectional views, and Figs. 7(f) and 8(f) for the MAP views, respectively. The noise remains strong, and the deeper vessels cannot be revealed. Finally, the N2N network-denoised results for subjects 01 and 02 are depicted in Figs. 7(g) and 8(g) for the cross-sectional views, and Figs. 7(h) and 8(h) for the MAP views, respectively. It can be seen that most of the noise within the vessels and in the background has been successfully removed. In the output images, there is a clear enhancement in the visibility of deep vessels, along with a decrease in the discontinuity of vessels originally caused by noise. Key aspects of the performance are pointed out in the figures.

In Figs. 7(b) and 8(b), as indicated by the arrows, the vessels blend in the background and are hard to identify. Additionally, some deeper vessels become invisible, and the boundary of the vessels is obscured by noise. After applying the N2N denoising network, Figs. 7(h) and 8(h) display markedly clearer vascular structures. The vessels are distinctly separated, each with clean and well-defined boundaries. Following noise removal, the deep vessels become visible. Additionally, issues of discontinuity within the vessel structures are effectively resolved.

We further conducted a thorough evaluation of the denoising performance by comparing the original input and the denoised output, focusing on how the network denoising influenced the SNR over various imaging depths. This assessment covered multiple layers, each with a thickness of 5 mm, to ensure a comprehensive analysis across different depths. For each layer, we quantified the SNR of a vessel, and the results are shown in Fig. 9. Subject 01 data had an average SNR improvement of 11.39 dB across 45 mm to 60 mm. Meanwhile, subject 02 demonstrated an average SNR enhancement of 8.96 dB. This results in an overall average SNR improvement of 10.18 dB. On the other hand, the overall SNR improvement achieved by the Gaussian filter is 1.99 dB, and the overall SNR improvement achieved by the Unet is 3.04 dB. The low performance of Gaussian and Unet denoising is due to the struggle to enhance the SNR at deeper tissues due to the similarity in intensity between the noise and vessel signals. The absence of a clear ground truth for Unet training also causes its low performance. In contrast, our in-vivo model effectively removes noise from deeper tissues, resulting in improved signal quality. We treat the SNR at 60 mm in subject 01 as an outlier because the vessel intensity was either equivalent to or even lower than the noise. These results further highlighted the network's performance in improving PA vascular imaging.

Biomedical Optics EXPRESS

4. Discussion

In PA imaging, the SNR decreases with increasing imaging depth, posing a challenge for deep-tissue imaging [61,62]. Traditional PA systems relied on increasing laser power to improve the signal intensity at deeper depths [63,64]. However, this limitation is hard to overcome due to device constraints [3,65]. This limitation becomes particularly critical in clinical imaging, where noise sources such as EMI and Gaussian noise can significantly compromise the visibility of vessels in human subjects [17,35].

The proposed unsupervised approach, based on the Noise2Noise network, has demonstrated its effectiveness in removing complex structured noise in both phantom and human subjects. In contrast to supervised learning, our method does not need labeled ground truth data. This simplifies the training process and ensures that deep vessel structures remain undistorted during denoising since it focuses on learning fundamental noise patterns [66,67]. Our method is also superior to Gaussian and supervised learning Unet denoising. As can be seen in Fig. 7(d) and Fig. 8(d), the Gaussian denoising filter could not remove the EMI noise and struggled at deeper tissues as it often blurs vessel edges. As can be seen in Fig. 7(f) and Fig. 8(f), the supervised-learning Unet approach cannot handle the noise when noise-free ground truth is absent. Another significant advantage of our proposed network is its transferability to different PA imaging systems. As an unsupervised learning technique, our method does not require any modeling of the imaging setup and a pair of noisy inputs can be easily generated from a single dataset [45]. In addition, as the training pair utilized noise from the experimental data itself, we do not need to identify diverse and complex noise patterns in the PA systems [68]. Therefore, the data generation and training procedure proposed in this study can be easily translated to other PA imaging systems to improve the SNR and imaging depth.

A notable limitation of our study is the relatively small dataset size, encompassing 5 phantoms in the simulation study and 20 human subjects in the clinical study. It might be necessary to enhance the design of the training dataset by increasing its complexity. By employing methods like flipping, cropping, or rotation during pre- and post-image processing, we can effectively increase the size of our dataset [69]. This expansion, including more diverse noise types captured from the system, might enhance the training process of our model [70]. In future studies, we intend to extend our Noise2Noise model to 3D applications [71,72]. We will explore the model performance against other semi-supervised and unsupervised models, including attention-based models like transformers and GAN networks. This expansion from 2D plane assessments will provide a broader understanding of how these models perform in complex 3D scenarios, enriching our comparative analysis of denoising techniques in photoacoustic imaging [73].

5. Conclusion

In this study, we developed and validated a Noise-2-Noise-based unsupervised training strategy for denoising photoacoustic images. The denoised image produced by our approach demonstrates a significant improvement in SNR, vessel connectivity, and vessel visibility over the depth. Unlike traditional supervised learning methods, the Noise2Noise approach does not require a clean image dataset for network training, making it more flexible and capable of handling in vivo imaging results. We have trained two models using simulation data and experimental data, respectively, and quantitative assessments from the two models have demonstrated the effectiveness of our method in improving the SNR across different imaging depths. Our proposed method holds great potential for clinical translation of PA.

Funding. National Institutes of Health (R01EB028978, R01EB029596).

Disclosures. Dr. Jun Xia is the founder of Sonioptix, LLC, which, however, did not support this work. All other authors declare no conflicts of interest.

Biomedical Optics EXPRESS

Data availability. The data supporting the findings of this study are available upon request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. L. V. Wang, "Prospects of photoacoustic tomography," Med. Phys. **35**(12), 5758–5767 (2008).
2. J. Xia, J. Yao, and L. V. Wang, "Photoacoustic tomography: principles and advances," Prog. Electromagn. Res. **147**, 1–22 (2014).
3. Y. Gu, Y. Sun, X. Wang, *et al.*, "Application of photoacoustic computed tomography in biomedical imaging: A literature review," Bioeng. Translational Med. **8**(2), e10419 (2023).
4. Y. Wang, Y. Zhan, L. M. Harris, *et al.*, "A portable three-dimensional photoacoustic tomography system for imaging of chronic foot ulcers," Quant. Imaging Med. Surg **9**(5), 799 (2019).
5. C. Huang, Y. Cheng, W. Zheng, *et al.*, "Dual-scan photoacoustic tomography for the imaging of vascular structure on foot," IEEE Trans. Ultrason., Ferroelect., Freq. Contr. **70**(12), 1703–1713 (2023).
6. R. Manwar, M. Zafar, and Q. Xu, "Signal and image processing in biomedical photoacoustic imaging: a review," Optics **2**(1), 1–24 (2020).
7. E. Zheng, H. Zhang, S. Goswami, *et al.*, "Second-generation dual scan mammoscope with photoacoustic, ultrasound, and elastographic imaging capabilities," Front. Oncol. **11**, 779071 (2021).
8. T. Duan, Y. Tang, F. Gao, *et al.*, "Application of mathematical morphological filter for noise reduction in photoacoustic imaging," Proc. SPIE **10878**, 1087854 (2019).
9. I. U. Haq, R. Nagoaka, T. Makino, *et al.*, "3D Gabor wavelet based vessel filtering of photoacoustic images," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2016).
10. T. Oruganti, J. G. Laufer, and B. E. Treeby, "Vessel filtering of photoacoustic images," Proc. SPIE **8581**, 85811W (2013).
11. S. H. Holan and J. A. Viator, "Automated wavelet denoising of photoacoustic signals for circulating melanoma cell detection and burn image reconstruction," Phys. Med. Biol. **53**(12), N227 (2008).
12. Z. Wang and I. Voiculescu, "Weakly supervised medical image segmentation through dense combinations of dense pseudo-labels," in *MICCAI Workshop on Data Engineering in Medical Imaging* (Springer, 2023).
13. H. Deng, H. Qiao, Q. Dai, *et al.*, "Deep learning in photoacoustic imaging: a review," J. Biomed. Opt. **26**(4), 040901 (2021).
14. W. Feng, W. Zhang, M. Meng, *et al.*, "A novel binary classification algorithm for carpal tunnel syndrome detection using LSTM," in *2023 IEEE 3rd International Conference on Software Engineering and Artificial Intelligence (SEAI)* (IEEE, 2023).
15. P. Kaur, G. Singh, and P. Kaur, "A review of denoising medical images using machine learning approaches," Current Med. Imaging **14**(5), 675–685 (2018).
16. D. Liu, Y. Cui, L. Yan, *et al.*, "Densernet: Weakly supervised visual localization using multi-scale feature aggregation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
17. W. Zheng, H. Zhang, C. Huang, *et al.*, "Deep-E enhanced photoacoustic tomography using three-dimensional reconstruction for high-quality vascular imaging," Sensors **22**(20), 7725 (2022).
18. H. Zhang, W. Bo, W. Xu, *et al.*, "Deep-learning-enhanced three-dimensional photoacoustic tomography of human breast," in *Frontiers in Optics + Laser Science 2022 (FIO, LS)* (Optica Publishing Group, 2022).
19. D. Zhang, F. Zhou, F. Albu, *et al.*, "Unleashing the power of self-supervised image denoising: a comprehensive review," arXiv (2023).
20. W. Zheng, H. Zhang, C. Huang, *et al.*, "Deep learning enhanced volumetric photoacoustic imaging of vasculature in human," Adv. Sci. **10**(29), 2301277 (2023).
21. M. Bernhardt, D. C. Castro, R. Tanno, *et al.*, "Active label cleaning for improved dataset quality under resource constraints," Nat. Commun. **13**(1), 1161 (2022).
22. D. Zheng, S. H. Tan, X. Zhang, *et al.*, "An unsupervised deep learning approach for real-world image denoising," in *International Conference on Learning Representations*, 2020.
23. P. Hermosilla, T. Ritschel, and T. Ropinski, "Total denoising: unsupervised learning of 3D point cloud cleaning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.
24. C. C. Tsai, X. Chen, S. Ahmad, *et al.*, "Robust unsupervised super-resolution of infant MRI via dual-modal deep image prior," in *International Workshop on Machine Learning in Medical Imaging* (Springer, 2023).
25. Z. Feng, M. Tu, R. Xia, *et al.*, "Self-supervised audio-visual representation learning for in-the-wild videos," in *2020 IEEE International Conference on Big Data (Big Data)* (IEEE, 2020).
26. D. Wu, K. Kim, and Q. Li, "Low-dose CT reconstruction with Noise2Noise network and testing-time fine-tuning," Med. Phys. **48**(12), 7657–7672 (2021).
27. J. Lehtinen, J. Munkberg, J. Hasselgren, *et al.*, "Noise2Noise: learning image restoration without clean data," arXiv (2018).
28. A. Kazakeviciute, C. J. H. Ho, and M. Olivo, "Multispectral photoacoustic imaging artifact removal and denoising using time series model-based spectral noise estimation," IEEE Trans. Med. Imaging **35**(9), 2151–2163 (2016).
29. A. F. Calvarons, "Improved Noise2Noise denoising with limited data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021).

Biomedical Optics EXPRESS

30. W. Jung, H.-S. Lee, M. Seo, *et al.*, "MR-self Noise2Noise: self-supervised deep learning–based image quality improvement of submillimeter resolution 3D MR images," Eur. Radiol. **33**(4), 2686–2698 (2022).
31. W. Bian, A. Jang, and F. Liu, "Improving quantitative MRI using self-supervised deep learning with model reinforcement: demonstration for rapid T1 mapping," Magnetic Resonance in Medicine (2024).
32. A. M. Hasan, M. R. Mohebbian, K. A. Wahid, *et al.*, "Hybrid-collaborative Noise2Noise denoiser for low-dose CT images," IEEE Trans. Radiat. Plasma Med. Sci. **5**(2), 235–244 (2021).
33. C. Chan, J. Zhou, L. Yang, *et al.*, "Noise to noise ensemble learning for PET image denoising," in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)* (IEEE, 2019).
34. V. Dutordoir, J. Hensman, M. van der Wilk, *et al.*, "Deep neural networks as point estimates for deep Gaussian processes," Advances in Neural Information Processing Systems **34**, 9443–9455 (2021).
35. O. Gulenko, H. Yang, K. Kim, *et al.*, "Deep-learning-based algorithm for the removal of electromagnetic interference noise in photoacoustic endoscopic image processing," Sensors **22**(10), 3961 (2022).
36. K.-T. Hsu, S. Guan, and P. V. Chitnis, "Comparing deep learning frameworks for photoacoustic tomography image reconstruction," Photoacoustics **23**, 100271 (2021).
37. M. Kuniyil Ajith Singh, K. Sivasubramanian, N. Sato, *et al.*, "Deep learning-enhanced LED-based photoacoustic imaging," Proc. SPIE **11240**, 1124038 (2020).
38. C. Dehner, I. Olefir, K. B. Chowdhury, *et al.*, "Deep-learning-based electrical noise removal enables high spectral optoacoustic contrast in deep tissue," IEEE Trans. Med. Imaging **41**(11), 3182–3193 (2022).
39. Y. Li, D. Liu, H. Li, *et al.*, "Convolutional neural network-based block up-sampling for intra frame coding," IEEE Transactions on Circuits and Systems for Video Technology **28**(9), 2316–2330 (2018).
40. C. Xu, J. Yang, H. Lai, *et al.*, "UP-CNN: un-pooling augmented convolutional neural network," Pattern Recognition Letters **119**, 34–40 (2019).
41. J. Wu, X. Ye, C. Mou, *et al.*, "FineEHR: refine clinical note representations to improve mortality prediction," in *2023 11th International Symposium on Digital Forensics and Security (ISDFS)* (IEEE, 2023).
42. B. Qiu, S. Zeng, X. Meng, *et al.*, "Comparative study of deep neural networks with unsupervised Noise2Noise strategy for noise reduction of optical coherence tomography images," J. Biophotonics **14**(11), e202100151 (2021).
43. N. Zhu, C. Liu, B. Forsyth, *et al.*, "Segmentation with residual attention u-net and an edge-enhancement approach preserves cell shape features," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, 2022).
44. Y. Cui, X. Liu, H. Liu, *et al.*, "Geometric attentional dynamic graph convolutional neural networks for point cloud analysis," Neurocomputing **432**, 300–310 (2021).
45. Z. Feng, D. Nie, L. Wang, *et al.*, "Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (IEEE, 2018).
46. Z. Cai, J. Chen, M. Liu, *et al.*, "Deep least-squares methods: An unsupervised learning-based numerical method for solving elliptic PDEs," J. Comput. Phys. **420**, 109707 (2020).
47. V. Shijo, T. Vu, J. Yao, *et al.*, "SwinIR for photoacoustic computed tomography artifact reduction," in *2023 IEEE International Ultrasonics Symposium (IUS)* (IEEE, 2023).
48. S. Liu, K. Wu, C. Jiang, *et al.*, "Financial time-series forecasting: towards synergizing performance and interpretability within a hybrid machine learning approach," arXiv, (2023).
49. J. Bjorck, C. Gomes, B. Selman, *et al.*, "Understanding batch normalization," Advances in Neural Information Processing Systems, **31**, 2018.
50. M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, *et al.*, "Improvement of a retinal blood vessel segmentation method using the Insight Segmentation and Registration Toolkit (ITK)," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, 2007).
51. B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," J. Biomed. Opt. **15**(2), 021314 (2010).
52. H. Zhang, W. Bo, D. Wang, *et al.*, "Deep-E: a fully-dense neural network for improving the elevation resolution in linear-array-based photoacoustic tomography," IEEE Transactions on Medical Imaging **41**(5), 1279–1288 (2022).
53. K. Dabov, A. Foi, V. Katkovnik, *et al.*, "Image denoising with block-matching and 3D filtering," in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning* (SPIE, 2006).
54. M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," IEEE Transactions on Image Processing **18**(1), 27–35 (2009).
55. Z. Xu, C. Li, and L. V. Wang, "Photoacoustic tomography of water in phantoms and tissue," J. Biomed. Opt. **15**(3), 036019 (2010).
56. D. Wang, Y. Wang, Y. Zhou, *et al.*, "Coherent-weighted three-dimensional image reconstruction in linear-array-based photoacoustic tomography," Biomed. Opt. Express **7**(5), 1957–1965 (2016).
57. G. Godefroy, B. Arnal, and E. Bossy, "Compensating for visibility artefacts in photoacoustic imaging with a deep learning approach providing prediction uncertainties," Photoacoustics **21**, 100218 (2021).
58. L. Zhang, L. Zhang, X. Mou, *et al.*, "FSIM: a feature similarity index for image quality assessment," IEEE Transactions on Image Processing **20**(8), 2378–2386 (2011).
59. X. Yu, L. Wu, Y. Lin, *et al.*, "Ultrafast Bragg coherent diffraction imaging of epitaxial thin films using deep complex-valued neural networks," npj Comput. Mater. **10**(1), 24 (2024).

Biomedical Optics EXPRESS

60. G. Guney, N. Uluc, A. Demirkiran, *et al.*, "Comparison of noise reduction methods in photoacoustic microscopy," Comput. Biol. Med. **109**, 333–341 (2019).
61. S. Zackrisson, S. Van De Ven, and S. Gambhir, "Light in and sound out: emerging translational strategies for photoacoustic imaging translational strategies for photoacoustic imaging," Cancer Res. **74**(4), 979–1004 (2014).
62. N. Nyayapathi, R. Lim, H. Zhang, *et al.*, "Dual scan mammoscope (DSM)—a new portable photoacoustic breast imaging system with scanning in craniocaudal plane," IEEE Trans. Biomed. Eng. **67**(5), 1321–1327 (2020).
63. S. Mallidi and S. Emelianov, "Photoacoustic technique to measure beam profile of pulsed laser systems," Rev. Sci. Instrum. **80**(5), 054901 (2009).
64. H. Yoon, G. P. Luke, and S. Y. Emelianov, "Impact of depth-dependent optical attenuation on wavelength selection for spectroscopic photoacoustic imaging," Photoacoustics **12**, 46–54 (2018).
65. Y. Wang, R. S. A. Lim, H. Zhang, *et al.*, "Optimizing the light delivery of linear-array-based photoacoustic systems by double acoustic reflectors," Sci. Rep. **8**(1), 13004 (2018).
66. A. Krull, T. Vičar, M. Prakash, *et al.*, "Probabilistic noise2void: Unsupervised content-aware denoising," Frontiers in Computer Science **2**, 5 (2020).
67. F. Zhu, S. Zhao, P. Wang, *et al.*, "Semi-supervised wide-angle portraits correction by multi-scale transformer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
68. R. Kikkawa, H. Sekiguchi, I. Tsuge, *et al.*, "Semi-supervised learning with structured knowledge for body hair detection in photoacoustic image," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (IEEE, 2019).
69. M. Prakash, A. Krull, and F. Jug, "Fully unsupervised diversity denoising with convolutional variational autoencoders," arXiv, (2020).
70. T. Tong, W. Huang, K. Wang, *et al.*, "Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data," Photoacoustics **19**, 100190 (2020).
71. Y. Zhang, S. Xu, H. Li, *et al.*, "Deep learning-based denoising in brain tumor CHO PET: comparison with traditional approaches," Appl. Sci. **12**(10), 5187 (2022).
72. J. R. Gambin, M. J. Tadi, J. Teuho, *et al.*, "Learning to denoise gated cardiac PET images using convolutional neural networks," IEEE Access **9**, 145886–145899 (2021).
73. X. Liu, Y. Hong, Q. Yin, *et al.*, "DNT: learning unsupervised denoising transformer from single noisy image," in *Proceedings of the 4th International Conference on Image Processing and Machine Vision*, 2022.