De-Noising of Photoacoustic Microscopy Images by Attentive Generative Adversarial Network

Da He, Jiasheng Zhou, Xiaoyu Shang, Xingye Tang, Jiajia Luo, Member, IEEE, and Sung-Liang Chen

Abstract—As a hybrid imaging technology, photoacoustic microscopy (PAM) imaging suffers from noise due to the maximum permissible exposure of laser intensity, attenuation of ultrasound in the tissue, and the inherent noise of the transducer. De-noising is an image processing method to reduce noise, and PAM image quality can be recovered. However, previous de-noising techniques usually heavily rely on manually selected parameters, resulting in unsatisfactory and slow de-noising performance for different noisy images, which greatly hinders practical and clinical applications. In this work, we propose a deep learning-based method to remove noise from PAM images without manual selection of settings for different noisy images. An attention enhanced generative adversarial network is used to extract image features and adaptively remove various levels of Gaussian, Poisson, and Rayleigh noise. The proposed method is demonstrated on both synthetic and real datasets, including phantom (leaf veins) and in vivo (mouse ear blood vessels and zebrafish pigment) experiments. In the in vivo experiments using synthetic datasets, our method achieves the improvement of 6.53 dB and 0.26 in peak signal-to-noise ratio and structural similarity metrics, respectively. The results show that compared with previous PAM de-noising methods, our method exhibits good performance in recovering images quantitatively. In addition, the de-noising processing speed of 0.016 s is achieved for an image with 256 × 256 pixels, which has the potential for real-time applications. Our approach is effective and practical for the de-noising of PAM images.

Index Terms—Photoacoustic microscopy, de-noising, generative adversarial network, deep learning.

I. INTRODUCTION

PHOTOACOUSTIC (PA) imaging (PAI), based on the PA effect, is a new non-invasive imaging technology, which images the object according to the PA signal generated upon laser illumination of the object [1], [2]. The advantage of PAI is that it combines the highlights of optical and ultrasonic imaging with good performance in terms of contrast and penetration. PAI can be implemented as PA computed tomography (PACT) and PA microscopy (PAM). The former forms images by image reconstruction algorithms, while the latter performs raster scanning without complicated image reconstruction [2]. PAI has potential applications in vascular biology, ophthalmology, dermatology, neurology, etc. [1]. High signal-to-noise ratio (SNR) is the key to high-quality PA images, which are essential for in vivo animal studies and clinical applications. A pulsed laser with pulse duration of a few nanoseconds is typically used for efficient PA conversion. The maximum PA signal amplitude is in part limited by allowable laser fluence on the tissue surface and is capped by the ANSI safety limit. In addition, the demand for large imaging depth and the expectation of an inexpensive hardware system limit noise suppression. PA contrast agents, as exogenous absorbers, are a common solution to enhance the PA signal amplitude [3]. Development of PA contrast agents for in vivo and clinical applications usually requires cumbersome effort. On the other hand, in PA signal and image acquisition, noise is generated from laser illumination to signal detection [4], [5], [6]. The noise in PA signals and images arises from several different factors, and building an accurate PA noise model can be complicated. Overall, the PA noise can be categorized into source-related noise (e.g., fluctuations in laser illumination) and system-related noise (e.g., randomly distributed thermal and electronic noise) [5]. Among the different types of thermal noise, white Gaussian
noise is the most common one [5]. Besides boosting the PA signal amplitude with contrast agents, de-noising is an alternative and promising approach to enhance SNR.

Several PA de-noising methods have been demonstrated to enhance the SNR of PA signals and/or images [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. The de-noising has been studied for PACT [7], [8], [9], [10], [12], [16], [17] and PAM [11], [14], [15]. For most PACT de-noising, PA raw data (A-line signals) are usually de-noised first prior to image reconstruction [7], [9], [12]. For PAM, de-noising may be conducted for PA A-line signals [14], [15] or for PA images directly [11], [13]. De-noising of in vivo images was also tested in some of these works [8], [9], [10], [11], [13], [14], [15], [17]. Among these PA de-noising techniques, signal averaging is the most commonly used one [9], yet acquisition of multiple signals is needed, which is time-consuming and results in low temporal resolution. Further, researchers have made effort to enhance SNR using various methods to handle the PA raw A-line signals, such as wavelet-based algorithms [7], [12], empirical mode decomposition (EMD) [14], [15], and sparsity-based methods [8], [16]. Alternatively, methods applied in the image domain, such as K-means singular value decomposition (KSVVD) and non-local means (NLM) [11], [13], can be used to de-noise PA images directly. However, these methods suffer one or some of the issues: (i) Prior information about the noise property is needed, which is a significant challenge [17]; (ii) time-consuming computation is required, especially for those based on iterative optimization; (iii) some parameters such as the noise level are needed to be manually specified for different input signals and images.

In recent years, deep learning-based methods have been extensively used for medical imaging [18]. There are many applications to medical imaging such as classification between lesions and non-lesions, classification of lesion types, detection of lesions, etc. [18]. Deep learning has also been used in PAI, such as reflection artifact removal in PACT images, PACT image reconstruction with sparse data, and PAM imaging with sparse data [19], [20], [21]. Besides, deep learning has shown promise in de-noising natural images [22], [23] as well as medical images including computed tomography (CT) [24], ultrasound imaging (de-speckling) [25], and optical coherence tomography [26]. Especially, generative adversarial network (GAN), which utilizes a discriminator network to guide the distribution of generated samples of a generator network [31], was used and showed great performance in de-noising [22], [23], [24], [25], [26]. Recently, PA de-noising based on deep learning has been investigated [17], [27]. Although Hariri et al. applied CNN to PA de-noising [17], PACT images with relatively low resolution and sparse patterns (in contrast to PAM images) were studied. Sharma et al. studied resolution enhancement in acoustic-resolution PAM (AR-PAM) images using CNN [27]. In that work, a CNN architecture was developed mainly to improve out-of-focus lateral resolution of AR-PAM images, while background noise reduction was also observed. Zhao et al. demonstrated good de-noising performance for optical-resolution PAM (OR-PAM) images [28], yet the use of two wavelengths for PA excitation was required, which results in higher cost and more complexity of the imaging system (e.g., an optical parametric oscillator laser was used.).

In this work, we aim to de-noise PAM images based on the deep learning method. As mentioned above, GAN has been demonstrated to be powerful in de-noising natural and medical images [22], [23], [24], [25], [26], while GAN-based PA de-noising work is still limited. Inspired by these works, we propose a GAN-based CNN to de-noise PAM images with the following key contributions. First, we develop a CNN network with key features including: (i) Attention modules in the generator, which effectively capture the pattern relationships in 2D images and thus can distinguish signal pixels from background noise; (ii) a novel combined loss function, which excels in restoring fine features and thus enables faithful recovery from noisy images. By contrast, a single pixel-based loss function (e.g., L2 loss) tends to ignore high-frequency details. Secondly, we demonstrate that for effective de-noising of PAM images, the developed CNN network outperforms the existing methods using block-matching and 3D filtering (BM3D) [29], KSVDD [11], and weighted nuclear norm minimization (WNNM) [30]. This is evidenced by quantitative and qualitative comparison. Finally, we explore how well our CNN network (trained with only the synthetic dataset) de-noises real in vivo experimental datasets of mouse ear vasculature and zebrafish pigment. The results show that our CNN network is promising for in vivo PAM de-noising applications.

II. METHOD

The proposed GAN-based CNN model consists of three main components: a generator network, a discriminator network, and a perceptual loss calculator (VGG network). Let \( I_N, I_C \), and \( \hat{I}_C \) (\( I_N, I_C, \hat{I}_C \in \mathbb{R}^{M \times N} \)) be the noisy PA image, the corresponding real clean image (if any), and the corresponding de-noised image, respectively, where \( M \) and \( N \) denote the height and width of the image, respectively. Then, our objective is to train a generator \( G \) that maps \( I_N \) to \( \hat{I}_C \) (\( G : I_N \rightarrow \hat{I}_C \)). Meanwhile, the discriminator \( D \) tries to distinguish the real clean PA image (\( I_C \)) from the de-noised PA image (\( \hat{I}_C \)). The training of the generator network against the discriminator forms the adversarial min-max problem [31].

For the generator network, we incorporate a global context (GC) attention module into a modified U-Net style structure to generate the de-noised image as output. In order to optimize the generator network, we elaborate a combined loss function comprising the perceptual loss, the pixel-level loss, and the adversarial loss.

In the following subsections, first, the noise model used to produce synthetic noisy PA images is discussed. Secondly, the architecture of our CNN model is detailed, including the attention module and the combined loss function. Thirdly, the dataset, network implementation, and evaluation metrics in this study are presented. Finally, other de-noising methods for comparison with our proposed CNN-based method are introduced.

A. PA Image Noise Model

Typically, noise in the PAI process can be simply modeled as the combination of Gaussian noise, Poisson noise, and
Rayleigh noise. Gaussian noise is a basic noise model to account for a few types of noise including thermal, amplifier, and read noise. Gaussian noise is independent of the signal variation and thus can be added (i.e., additive noise) to any other noise that might be intrinsic to the system. In PA de-noising works, Gaussian noise is commonly assumed and modeled [8], [11], [16]. On the other hand, the signal-dependent part of the noise can be modeled as Poisson noise, which can be introduced in the process of signal conversion and transmission relating to the fluctuation and digitalization of particles in electronic devices [32]. Rayleigh noise may be introduced by reflection and scattering of PA waves in tissue transmission [33].

In real experimental PA images, the noise is usually complicated and non-uniformly distributed over the entire image. Therefore, there is no clear clue how the noisy PA images and their latent clean PA images are interrelated, which makes it difficult to de-noise PA images using traditional methods. Alternatively, one can use a CNN model trained on a noisy PA image dataset, e.g., using a synthetic dataset through introducing the noise into clean PA images, to de-noise real noisy PA images.

**B. Network Architecture**

1) **Generator Network:** As indicated in Fig. 1, the generator framework employs the U-Net shape architecture [50]. U-Net shape networks have shown high performance in medical image processing due to multi-scale feature fusion and lightweight parameters. In Fig. 1, the encoder is the left part of the generator with decreasing feature map sizes, including five standard units, five GC blocks [34], and four max-pooling layers. On the other hand, the decoder is the right part of the generator with increasing feature map sizes, including four transposed convolutional layers and four standard units. The proposed standard unit block consists of two $3 \times 3$ convolutional layers, each followed by an instance normalization layer [35] and a leaky ReLU (LReLU) function. All nine standard unit blocks in the generator have 32, 64, 128, 256, 512, 256, 128, 64, and 32 filters, respectively.

To efficiently capture features and distinguish information with varying importance, an attention mechanism is applied in our network. Unlike regular CNNs that may treat all information equally, attention blocks additionally introduce attention weights for different feature channels or spatial positions. Specifically, the proposed method utilizes the attention block, i.e., the GC block, to enhance the attention to long-range dependencies and thus better handle unexpected noise instead of focusing on signal pixels. The detailed structure of the GC attention block includes $1 \times 1$ convolutions and layer normalizations as described in Fig. 1. In the generator, GC blocks are placed after each standard unit block of the encoder.

Then, in the decoding process, the generator performs direct up-sampling of the previously extracted high-level feature maps, combining global and local features, until the size (in pixels) of the original image is restored. The max-pooling layers and the transposed convolutional layers are used for down-sampling and up-sampling in the encoder and decoder parts, respectively. The generator takes the noisy image as input and outputs its de-noised version.

Overall, the baseline architecture (i.e., U-Net style architecture) has been widely proved effective for biomedical imaging modalities. We further apply the attention mechanism to improve information filtering in the network, which is quite effective for the de-noising task because the informative signal and the background noise can be distinguished in certain convolutional channels.

2) **Discriminator Network:** The discriminator illustrated in Fig. 1 is an 8-layer CNN with the number of filters as 64, 64, 128, 128, 256, 256, 512, and 512, respectively. Each of the convolutional layers has a kernel size of $3 \times 3$ and is subsequently equipped with a LeReLU and a batch normalization (BN). In the end, there are two fully-connected (FC) layers with 1024 outputs and a single output, respectively. The input of the discriminator is the de-noised image or its corresponding
C. Combined Loss Function

The de-noising results are expected to show effective noise-free performance while maintaining clear and informative signals. To achieve this goal, we propose a combined loss function to help network training with several advantages, as detailed in the following. The combined loss function for our model consists of three parts: (i) perceptual loss [36] \( L_{\text{perceptual}} \), (ii) smooth L1 loss [37] \( L_{\text{smoothL1}} \), and (iii) adversarial loss \( L_{\text{GAN}} \). As a variant of the pixel-wise loss function, smooth L1 loss contributes to the convergence of training efficiently and effectively. Facing the fact that the pixel-wise loss function might lead to over-smoothed patterns [21], we apply the perceptual loss function to get clear pattern edges. Note that only the perceptual loss is introduced in [21], which aims at the recovery from an OR-PAM image with sparse data. On the other hand, noise may resemble some subtle patterns in PAM images. To better distinguish informative signals from noise, we utilize the GAN loss function to help distinguish between real and fake signals, which contributes to noise removal while maintaining informative signals. The GAN loss function thus improves the fidelity of the de-noising results. As a result, the novel combined loss function serving different purposes is expected to be effective for the PM de-noising task. Our combined loss function is expressed as:

\[
L = k_1L_{\text{perceptual}} + k_2L_{\text{smoothL1}} + k_3L_{\text{GAN}},
\]

where \( k_1, k_2 \) and \( k_3 \) are hyperparameters to control the trade-off among all components.

1) Perceptual Loss: Most de-noising algorithms, including deep learning-based methods, aim to minimize the mean square error (MSE) between the de-noised image (\( \hat{I}_C \)) and the ground truth (IC). However, using only MSE loss may produce blurred images with loss of details and faithfulness [21]. Note that pixel-wise loss functions (e.g., MSE) directly compare differences based on all pixels and ultimately yield a single value for the whole image, which does not focus on recovery of a sharp edge and could lead to an over-smoothed pattern. In contrast, by virtue of the convolutional filters (including edge filters) in the perceptual loss, fine features are more likely to be restored. This phenomenon has been widely demonstrated previously (e.g., [21], [36]). Thus, to better restore images, we adopt the perceptual loss, which contains various convolutional filters to extract fine textures and focuses on the high-level features to make the restoration performance more consistent with the perception of the human visual system. Specifically, the high-level features are extracted from the pre-trained VGG-19 model [38]. The perceptual loss can be formulated as follows:

\[
L_{\text{perceptual}}(\hat{I}_C, IC) = \frac{1}{HWC} \sum_{i,j,k}(VGG_{i,j,k}(\hat{I}_C) - VGG_{i,j,k}(IC))^2,
\]

where \( H, W \), and \( C \) are the height, width, and channel size of the input tensor.

2) Smooth L1 Loss: Smooth L1 loss can avoid the defects of the regular L1 loss and L2 loss, alleviating the gradient explosion of L2 loss for large errors and also improving the stability of L1 loss for small errors. The applied smooth L1 loss is expressed as follows:

\[
L_{\text{smoothL1}}(\hat{I}_C, IC) = \frac{1}{HWC} \sum_{i,j,k}L(\hat{I}_{C,i,j,k} - IC_{i,j,k}),
\]

\[
L = \begin{cases} 
0.5(\hat{x} - x)^2, & \text{if } |\hat{x} - x| < 1 \\
|\hat{x} - x| - 0.5, & \text{otherwise}
\end{cases}
\]

where \( |\hat{x} - x| \) is the pixel difference between the restored image patch and the clean one.

3) GAN Loss: A generative component is added to our GAN to force the restored image to be as realistic as the clean image, so that the restored image can deceive the discriminator. The adversarial loss over a batch of training samples is expressed as follows:

\[
L_{\text{GAN}}(G) = \sum_{i=1}^{BS}[-\log D(G(I_N))],
\]

where \( BS \) is the batch size, and \( D(G(I_N)) \) is the estimated probability that the de-noised image \( G(I_N) \) (i.e., \( \hat{I}_C \)) is a real clean PA image.

D. Dataset

Supervised de-noising training is expected to receive numerous image pairs with corresponding noisy and clean images; however, obtaining paired biomedical PAM images in experiments is difficult and time-consuming. To handle the limitation, a synthesis operation is applied to construct the training dataset. Besides, since the training dataset employs an inclusive noise model that considers three types of noise, the CNN model trained from the dataset is expected to perform well on synthetic datasets as well as real testing datasets.

We first introduce the preparation of clean PAM images of leaf veins and mouse ear blood vessels as the clean images used in this work. Then, we describe the dataset of synthetic noisy images used for training. Finally, we present the acquisition of the dataset of real noisy images. The details about the PAM experimental setup and image acquisition can be found in our recent paper [21]. Note that in PAM images, 2D maximum amplitude projection (MAP) images (projection along the axial direction) are typically used. Therefore, in this work, we aim to restore PAM MAP images with good de-noising effect. The animal experiment was conducted in conformity with the laboratory animal protocol approved by Institutional Animal Care and Use Committee of Shanghai Jiao Tong University.

1) Raw Clean Leaf Vein Data: We adopted the full-scanning PAM data, consisting of bodhi and magnolia leaf veins, acquired in our previous work [21]. The size of the raw images is 256 \times 256 \times 180, where 180 is the number of pixels along the axial direction. Thus, the size of the 2D MAP images is 256 \times 256. In total, there are 260 leaf vein PAM images as the clean leaf vein dataset used in this work.
2) Raw Clean Blood Vessel Data: We experimentally acquired many PAM images of mouse ear blood vessels in vivo. In total, 165 image patches with size of 250 \times 250 \times 180 were prepared as the clean blood vessel dataset. Compared with PAM images of leaf veins, those of blood vessels have relatively rich details and complex structure. Thus, the de-noising of PAM images of blood vessels is expected to be more challenging.

3) Synthetic Noisy Image Dataset: The noise model described in subsection II-A can be used to synthesize noisy images. The raw clean data mentioned above (subsections II-D.1 and II-D.2) were used before adding noise. All the three types of noise (i.e., Gaussian, Poisson, and Rayleigh noise) were randomly generated by Python built-in functions, with a series of random noise levels ranging from mild to severe. To better mimic the noise generation during PAM image acquisition, instead of adding noise directly to the 2D MAP images, noise was added to PA A-line signals, and then MAP was applied to obtain noisy 2D images. In other words, 3D raw clean data were used before adding noise. Among the clean leaf vein dataset, 236 images were used to construct clean-noisy image pairs for training, and 24 images were used for testing. As for the clean blood vessel dataset, similarly, 149 images were used for training and 16 images for testing.

For instance, the inclusion of Gaussian noise during the synthesis operation is elaborated as follows. As mentioned in subsection II-A, Gaussian noise is an additive noise and is independent of the signal variation. Its noise level depends on a particular experimental system. Therefore, to ensure that our method can well handle different and unknown levels of Gaussian noise from various systems, we generated random Gaussian noise levels during the synthesis operation. Specifically, for each input clean image, we randomly selected 5 to 12 standard deviation values of the Gaussian model (corresponding to 5 to 12 Gaussian noise levels), and the synthetic noisy images with different Gaussian noise levels can be obtained. Moreover, with the proposed methodology, the range of standard deviation values of the Gaussian model can be expanded accordingly to account for more different systems.

4) Real Noisy Image Dataset: In PAM imaging, the higher the excitation light energy used, the better the SNR obtained. Therefore, a real noisy PAM image dataset with low SNR was obtained by using relatively low excitation light energy. In total, we acquired 16 real noisy images of leaf veins with size of 256 \times 256 and 21 real noisy images of blood vessels with size of 250 \times 250 to test the performance of our CNN model. Further, to test the generalization ability of our CNN model, we also used real noisy images of zebrafish pigment acquired in our previous work [39].

5) Image Data Acquired Under Low to High Excitation Light Energy: For more comparison, 16 leaf vein image pairs under low excitation light energy and high excitation light energy were experimentally acquired. Besides, two sets of in vivo mouse ear blood vessel images under low to high excitation light energy were experimentally acquired. The detailed results are described later.

E. Network Implementation

To evaluate the performance of our CNN model, we trained two models using different synthetic noisy image datasets (including leaf veins and mouse ear blood vessels, as mentioned in III.D). We adopted Adam optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ [40]. We set the learning rate as 0.0001, the batch size as 8, and the iteration number as 60,000. During training, the coefficients of different loss terms in Eq. (1) were tuned based on experimental results. The $k_2$ in the combined loss was initially 1.0 for efficient convergence and finally decreased to 0, while $k_1$ and $k_3$ increased from 0 to 1.0 and $10^{-3}$, respectively, for stable training and high-quality results. As mentioned previously, smooth L1 loss is a pixel-wise loss function that is efficient and effective for training convergence. Thus, we applied smooth L1 loss during the early stages to facilitate the robust convergence of the network training. However, smooth L1 loss may lead to over-smoothed patterns. This is why we turned to using perceptual loss and GAN loss to tune fine details in the final training stage. All the models were implemented using TensorFlow (v. 1.13.1) based on Python (v. 3.5.2). The training of each model took about 8 hours on an NVIDIA TITAN RTX GPU with 24 GB memory.

F. Evaluation Metrics

1) Full-Reference Metrics: For the synthetic noisy image dataset with ground truth, to quantitatively evaluate de-noising performance, we used the image similarity metrics of peak SNR (PSNR) and structural similarity (SSIM) [42]. High-quality images are expected to have high PSNR or high SSIM values. Specifically, PSNR is defined as:

$$MSE = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} (I_C(i, j) - \hat{I}_C(i, j))^2,$$

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

where $MAX$ is the maximum possible pixel value (e.g., 255 for 8-bit images). Meanwhile, SSIM is defined as:

$$SSIM = \frac{(2\mu_{I_c}\mu_{\hat{I}_c} + \epsilon_1)(2\sigma_{I_c\hat{I}_c} + \epsilon_2)}{(\mu_{I_c}^2 + \mu_{\hat{I}_c}^2 + \epsilon_1)(\sigma_{I_c}^2 + \sigma_{\hat{I}_c}^2 + \epsilon_2)},$$

where $\mu$, $\sigma$, $\sigma_{I_c\hat{I}_c}$ indicate the mean value, the variance, and the covariance, respectively. Small constants $\epsilon_1$ and $\epsilon_2$ are applied to maintain the stabilization.

G. No-Reference Metrics

By contrast, for the real noisy image dataset without ground truth, the de-noising performance was evaluated by the no-reference metrics of SNR and contrast-to-noise ratio (CNR). To calculate these metrics, representative signal regions and background regions from each sample should be selected. SNR and CNR can be calculated as follows, respectively [5]:

$$SNR = 20 \log \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\mu_i}{\sigma_b} \right),$$
\[ CNR = 20 \log \left( \frac{1}{n} \sum_{i=1}^{n} \frac{|\mu_i - \mu_b|}{\sqrt{\sigma_i^2 + \sigma_b^2}} \right), \]  

(10)

where \(\mu_i\) and \(\sigma_i\) denote the mean and standard deviation of the \(i\)th signal region, respectively; \(\mu_b\) and \(\sigma_b\) are the mean and standard deviation of the background region; \(n\) is the number of signal regions, which was set as 4 in our evaluation. The signal and background regions are illustrated as blue and green boxes, respectively, as shown in Fig. 2.

These metrics (PSNR, SSIM, SNR, and CNR) have been widely used in published articles regarding biomedical image restoration and imaging (e.g., [5], [24], [28]). Besides, our choice of these four metrics allows a thorough evaluation of the de-noising performance with different focuses. PSNR focuses on pixel-wise corrections, while SSIM may be more suitable for the human visual system. SNR is calculated by the mean value of the signal region and the standard deviation of the background noise. CNR further considers the difference between the mean values of the signal region and background noise. Therefore, by considering the informative region (i.e., the signal region) and the background noise separately, high SNR tends to obtain even backgrounds (accounted by \(\sigma_b\) in Eq. (9)), while high CNR further expects an obvious margin between informative signal intensities and background noise intensities (accounted by \(|\mu_i - \mu_b|\) in Eq. (10)) [49]. In other words, high CNR expects good separability between the signal and the background noise. Both higher SNR and higher CNR indicate better image quality.

H. Comparison With Other Methods

In this work, we compared our proposed method with three common de-noising algorithms, which are based on an image non-local self-similarity (NSS) model named BM3D [29], a sparse model named KSVD [41], and a low-rank model named WNNM [30]. It should be suitable and fair to use these three methods for comparison with our method because they have already been verified on both natural images and related biomedical images (e.g., ultrasound images and PA images [6], [45], [46]).

III. RESULTS

A. Synthetic Leaf Vein Dataset

As described previously, 260 leaf vein PAM images were used to implement the de-noising experiment. These clean PAM images were regarded as the ground truth while their corresponding synthetic noisy images were used as input, forming a supervised learning manner to train the proposed de-noising network. Specifically, after randomly adding three kinds of noise, there are 808 image pairs for training and 83 image pairs for testing. After training, the de-noising model was evaluated using the testing set. For fair comparison, all the following quantitative and qualitative results were generated based on the testing set only.

Because the synthetic experiment includes both the noisy input and the ground truth of PAM images, full-reference metrics of PSNR and SSIM can be used to provide a good measure of the de-noising performance (i.e., the similarity between the de-noised result and the ground truth). Table I shows the statistical results of the testing set.

According to Table I, our method outperforms other methods with the obvious margin in terms of both PSNR and SSIM. After adding various types and levels of noise, the noisy images in the testing set have the average PSNR and SSIM of 21.79 dB and 0.34, respectively, which indicate that the image quality is quite low. Compared to the noisy input, our method achieves around 5 dB and 147% improvement in these two metrics, respectively. A high SSIM of 0.84 indicates that the restored image is almost the same as the ground truth. By contrast, the other three methods (i.e., BM3D, KSVD, and WNNM) show very limited performance in the quantitative comparison. Especially for the SSIM, all the three methods show almost no improvement.

The PAM images would provide visual and qualitative comparison. Fig. 3 shows two representative results from the testing set, where the drawbacks of the three traditional methods (BM3D, KSVD, and WNNM) can be identified. First, in some PAM images, including those recovered by BM3D (in the bottom row) and KSVD (in both rows), the noise cannot be effectively removed. The traditional methods rely on rigid mathematical priors, which are associated with an accurate noise model and certain proper parameters. When the noise model is complex and/or certain parameters are not chosen properly, the de-noising performance would be highly degraded. In addition, the background regions of the PAM images recovered by the three traditional methods are found to be noisier (e.g., scattered noise observed in the cases of BM3D and KSVD) or not dark enough (e.g., a uniform red background observed in the case of WNNM (the full image in the bottom row), leading to low signal-to-background contrast) compared to those recovered by our CNN method. By contrast, our method is free from these drawbacks and thus produces the best de-noising results.

B. Real Noisy Leaf Vein Dataset

As mentioned in subsection II-D, the three types of noise added in PA A-line signals can well simulate the noise generated during PAM imaging. Therefore, we expect that the model trained on the synthetic dataset (subsection II-E) would also perform well for real noisy images.

As described earlier, 16 noisy leaf vein PAM images were experimentally acquired (subsection II-D.4) using relatively
TABLE I
QUANTITATIVE COMPARISON RESULTS (MEAN ± STANDARD DEVIATION) USING THE SYNTHETIC LEAF VEIN DATASET. P-VALUES ARE CALCULATED BASED ON THE COMPARISON WITH OUR RESULTS. THE SYMBOLS ***, **, AND * INDICATE P-VALUES < 0.001, < 0.01, AND < 0.05, RESPECTIVELY, WHICH IS USED THROUGHOUT THE PAPER.

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy input</th>
<th>BM3D</th>
<th>KSVD</th>
<th>WNNM</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>21.79±4.08***</td>
<td>21.97±3.91***</td>
<td>21.88±3.93***</td>
<td>22.15±4.04***</td>
<td>26.81±1.93</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.34±0.14***</td>
<td>0.34±0.14***</td>
<td>0.34±0.14***</td>
<td>0.35±0.14***</td>
<td>0.84±0.08</td>
</tr>
</tbody>
</table>

![Ground truth | Noisy input | Recovered by BM3D | Recovered by KSVD | Recovered by WNNM | Recovered by ours](image1)

Fig. 3. Representative results from the testing set of the synthetic leaf vein dataset. The top row is from a bodhi leaf and the bottom row comes from a magnolia leaf. Scale bar: 500 µm. All images, excluding zoom images, share the same scale bar. The values in the colorbar indicate relative PA intensity.

TABLE II
QUANTITATIVE COMPARISON RESULTS (MEAN ± STANDARD DEVIATION) USING THE REAL NOISY LEAF VEIN DATASET. P-VALUES ARE CALCULATED BASED ON THE COMPARISON WITH OUR RESULTS.

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy input</th>
<th>BM3D</th>
<th>KSVD</th>
<th>WNNM</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR (dB)</td>
<td>29.08±7.96***</td>
<td>36.23±5.36***</td>
<td>29.20±6.58***</td>
<td>35.53±15.65***</td>
<td>90.73±15.86</td>
</tr>
<tr>
<td>CNR (dB)</td>
<td>4.80±2.59**</td>
<td>5.37±2.09**</td>
<td>5.53±2.46**</td>
<td>5.72±2.25*</td>
<td>7.63±1.77</td>
</tr>
</tbody>
</table>

![Real noisy input | Recovered by BM3D | Recovered by KSVD | Recovered by WNNM | Recovered by ours](image2)

Fig. 4. Representative results from the experimentally-acquired noisy leaf vein dataset. Both rows are from bodhi leaves. Scale bar: 500 µm. All images, excluding zoom images, share the same scale bar. The values in the colorbar indicate relative PA intensity.

According to Table II, the proposed method also achieves the best quantitative results in terms of both SNR and CNR metrics. For SNR, compared to the noisy input, BM3D and KSVD produce little improvement (<8 dB), and WNNM enables appreciable improvement (~26 dB). As a comparison, our method results in the highest SNR improvement (~62 dB).

low excitation light energy. The model trained on the synthetic dataset (subsection III-A) was used without further fine-tuning. Table II shows the quantitative results, and the no-reference metrics of SNR and CNR are used to compare the various denoising methods. Similarly, two representative results of PAM images (out of 16) are shown in Fig. 4.
TABLE III
QUANTITATIVE COMPARISON RESULTS (MEAN ± STANDARD DEVIATION) USING THE SYNTHETIC BLOOD VESSEL DATASET. P-VALUES ARE CALCULATED BASED ON THE COMPARISON WITH OUR RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy input</th>
<th>BM3D</th>
<th>KSVD</th>
<th>WNNM</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>23.14±5.88***</td>
<td>24.75±5.82***</td>
<td>23.44±5.54***</td>
<td>25.17±6.03***</td>
<td>29.67±4.48</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.37±0.23***</td>
<td>0.47±0.21***</td>
<td>0.36±0.18***</td>
<td>0.50±0.22***</td>
<td>0.63±0.12</td>
</tr>
</tbody>
</table>

TABLE IV
QUANTITATIVE COMPARISON RESULTS (MEAN ± STANDARD DEVIATION) USING THE REAL NOISY BLOOD VESSEL DATASET. P-VALUES ARE CALCULATED BASED ON THE COMPARISON WITH OUR RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy input</th>
<th>BM3D</th>
<th>KSVD</th>
<th>WNNM</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR (dB)</td>
<td>26.64±5.03***</td>
<td>36.54±2.50***</td>
<td>27.53±4.37***</td>
<td>42.33±6.03***</td>
<td>47.24±6.60</td>
</tr>
<tr>
<td>CNR (dB)</td>
<td>9.93±1.82**</td>
<td>11.17±2.62</td>
<td>9.98±1.70**</td>
<td>11.07±2.60</td>
<td>12.37±2.66</td>
</tr>
</tbody>
</table>

Fig. 5. Representative results of the mouse ear blood vessel dataset acquired by in vivo experiment. Top row: a representative sample from the synthetic noisy dataset; bottom row: a representative sample from the real noisy dataset. Scale bar: 250 µm. All images, excluding zoom images, share the same scale bar. The values in the colorbar indicate relative PA intensity.

**C. Blood Vessel Dataset: In Vivo Experiment**

To further demonstrate that our method can be used for in vivo imaging applications, de-noising of PAM images of mouse ear blood vessels was studied. Compared with leaf veins, blood vessels usually have relatively complex structure and low image quality, including (i) tortuous pattern, (ii) low SNR due to the laser safety limit in tissue, and (iii) artifacts (e.g., serrated boundaries) due to animal breathing and undesired absorbers in tissue. Therefore, it is expected to be more challenging to de-noise PAM images of blood vessels.

Our in vivo model was trained on the synthetic dataset (the clean blood vessel dataset (subsection II-D.2) and the corresponding synthetic noisy blood vessel dataset). Specifically, there are 1603 image pairs for training and 128 image pairs for testing. Similar to the synthetic leaf vein experiment, the model was first evaluated using the synthetic testing set (i.e., the 128 image pairs), and the results are shown in Table III. From Table III, the proposed method still outperforms other methods with noteworthy improvement for the in vivo dataset.

As mentioned in subsection II-D.4, 21 real noisy blood vessel images were experimentally acquired. Table IV shows the quantitative results to compare the de-noising performance by different methods. Without ground truth, the no-reference metrics of SNR and CNR are used. The model trained on the synthetic blood vessel dataset was used here without additional transfer learning. According to Table IV, our method still achieves the best de-noising performance in terms of quantitative metrics of SNR and CNR.

The de-noising performance can also be qualitatively observed in Fig. 5. The top row of Fig. 5 shows the de-noising results of a representative sample from the synthetic noisy testing set, while the bottom row shows the de-noising results of a representative sample from the real noisy dataset. In general, the results in Fig. 5 are consistent with Tables III and IV. Our method produces clear blood vessels, while the other methods suffer from discontinuities (e.g., in the case of WNNM (the zoom image in the bottom row)) and severe interference by the
background (e.g., in the cases of BM3D and KSVD). Although the model was only trained on the synthetic dataset, the noise in real in vivo PAM images can still be effectively removed, which may be attributed to (i) the powerful image feature extraction and processing in the GAN-based model, and (ii) the inclusive noise model (i.e., three types of noise considered) for the training dataset.

D. Self-Adaptability Demonstration

To illustrate the advantage of the self-adaptability of our method, further comparison was demonstrated using the testing sets of both the synthetic leaf vein dataset and the synthetic mouse ear blood vessel dataset. Specifically, each testing set was split into three groups (low noise level, middle noise level, and high noise level) based on the noise level of the synthetic noisy input images. Each group had a similar number of images. Different de-noising methods were applied to each group, respectively, for comparison. The results are shown in Fig. 6, where PSNR and SSIM are compared. The superiority of our method can be found in three aspects.

First, the three conventional methods adopt the manually tuned parameters to produce optimized results for different noise levels, while our method achieves automatic processing. Therefore, the proposed deep learning-based method would facilitate practical de-noising applications for PAM. Here, automatic processing refers that our model works well for various noise levels of noisy input images without any manual settings of parameters; by contrast, in conventional de-noising methods or non-end-to-end deep learning methods, manual settings of parameters according to different noise levels of noisy input images are required. Secondly, it can be seen that our method achieves the highest PSNR and SSIM in all the three groups in both testing sets. Besides, among the three different noise levels, our method produces the best absolute performance (i.e., the highest values of PSNR and SSIM) for the low noise level group. This is reasonable because the de-noising performance is typically better given the noisy input images with low noise level. Thirdly, the margin of PSNR and SSIM between our method and the conventional methods becomes larger as the noise level of the noisy input image increases in both testing sets. For example, in the in vivo testing set (Figs. 6(c) and 6(d)), our method outperforms WNNM by $\sim 2.3$ dB and $\sim 0.05$ for PSNR and SSIM, respectively, for the group of low noise level, while the corresponding margins become $\sim 6.1$ dB and $\sim 0.17$ for the group of high noise level. That is, among the three different noise levels, interestingly, our method achieves the best relative improvement (i.e., the largest difference of PSNR and SSIM values compared with the noisy input) for the high noise level group. Compared with traditional methods, our method leads to the best absolute performance as well as relative improvement for all the three different noise levels, which indicates that our method has the good adaptive nature without the need to manually adjust the parameters. The above statements are true for both PSNR (Figs. 6(a) and 6(c)) and SSIM (Figs. 6(b) and 6(d)).

In Fig. 6, p-values are also checked based on the comparison with our results. Except for the case of WNNM vs. ours for the SSIM metric of the low noise level group of the in vivo dataset (in Fig. 6(d)), the results show significant differences in all the other comparison: (i) P-value $< 0.05$ for WNNM vs. ours for the PSNR metric of the low noise level group of the in vivo dataset (in Fig. 6(c)); (ii) p-value $< 0.01$ for WNNM vs. ours for the PSNR metric of the low noise level group of the phantom dataset (in Fig. 6(a)) and for BM3D vs. ours for the PSNR and SSIM metrics of the low noise level group of the in vivo dataset (in Figs. 6(c) and 6(d)); (iii) p-value $< 0.001$ for the others.

In addition, we tested our method for real noisy images with much larger size and with much different patterns (zebrafish pigment vs. blood vessels). The results show that our method is robust. As shown in Fig. 7, without any transfer learning, our model trained on the synthetic blood vessel dataset can be used to de-noise wide-field PAM images. The two real noisy input images in Fig. 7 are of $1500 \times 600$ pixels (corresponding physical size: $7.5 \text{ mm} \times 3.0 \text{ mm}$) for mouse ear blood vessels (Fig. 7(a)) and $1448 \times 280$ pixels (corresponding physical size: $6.2 \text{ mm} \times 1.2 \text{ mm}$) for zebrafish pigment (Fig. 7(b)). The pixels in Fig. 7 is much more than those in the training set ($250 \times 250$ pixels). Furthermore, although the pattern of zebrafish pigment differs a lot from the pattern of blood vessels used for training (e.g., circle-like patterns in the zoom image around the zebrafish brain region), our model achieves excellent de-noising performance for zebrafish pigment, demonstrating the effectiveness of our comprehensive synthesis of three types of noise and the robust adaptability of our model to some extent.

For better comparison, Fig. 7(d) shows representative one-dimensional (1D) profiles along the dotted lines in the left zoom images of mouse ear blood vessels (Fig. 7(c)), which presents clear comparison between before and after de-noising (i.e., noisy input vs. recovered by ours). Similarly, Fig. 7(e)
Fig. 7. Real noisy images with much larger size and with much different patterns. (a) Mouse ear blood vessels before and after de-noising (scale bar: 1 mm), shown in the left and right figures, respectively. (b) Zebrafish pigment before and after de-noising (scale bar: 500 µm), shown in the left and right figures, respectively. The white boxes indicate the regions for gCNR calculation. (c) Representative zoom images of mouse ear blood vessels in the blue boxes (corresponding to the blue boxes from left to right in (a)) and zebrafish pigment in the green boxes (corresponding to the green boxes from left to right in (b)); the zoom images before de-noising are shown in the left 6 figures, and the zoom images after de-noising are shown in the right 6 figures. The values in the colorbar indicate the relative PA intensity. (d) 1D profiles along the dotted lines in the left zoom images of mouse ear blood vessels in (c) (the black line before de-noising and the cyan line after de-noising) for clear comparison between before and after de-noising. (e) 1D profiles along the dotted lines in the left zoom images of zebrafish pigment in (c) (the gray line before de-noising and the green line after de-noising) for clear comparison between before and after de-noising.

shows representative 1D profiles for the case of zebrafish pigment. As can be seen in Figs. 7(d) and 7(e), the edges after de-noising are similar or slightly sharper compared with those before de-noising (e.g., indicated by the arrow in Fig. 7(d)). That is, the spatial resolution can be regarded as similar or slightly improved after de-noising. Further, in Figs. 7(d) and 7(e), some small fluctuations in the signal region and in the background region are perfectly removed after de-noising (e.g., indicated by the arrows in Fig. 7(e)), which is consistent with the improved SNR and CNR values for the recovered PAM images (shown in Figs. 7(a) and 7(b)). The de-noising performance is evident by side-by-side comparison of zoom images in Fig. 7(c) and 1D profiles in Figs. 7(d) and 7(e) using our model. Note that in the above comparison, we mainly refer to zoom images (Fig. 7(c)), rather than wide-field images (Figs. 7(a) and 7(b)). This is because the comparison of de-noising performance is much clearer using zoom images, as can also be observed in Fig. 5.

In addition, generalized CNR (gCNR) is one of the well-designed metrics for evaluating the separability of the target from the background [49], especially for very noisy regions. By selecting the very noisy parts of the zebrafish in Fig. 7(b) (e.g., the fin and the tail) as the target regions, as indicated by the white boxes in Fig. 7(b), and the top right part as the background region, as indicated by the white dashed box in Fig. 7(b), the gCNR value is improved from 0.6836 to 0.8553 after de-noising, showing that our model can also improve the separability of the target from the background.

E. Validation of the Network Architecture

To further validate the effectiveness of the proposed method, we also conducted comparative experiments with previous deep learning architectures and the ablation counterpart of our method. Specifically, we trained DnCNN [43], U-Net [50], Residual U-Net [44], and our network without GC blocks using the synthetic mouse ear blood vessel dataset.
TABLE V
QUANTITATIVE COMPARISON RESULTS (MEAN ± STANDARD DEVIATION) BY DIFFERENT DEEP LEARNING METHODS AND THE ABLATION COUNTERPART OF OUR METHOD USING THE SYNTHETIC BLOOD VESSEL DATASET (IN PSNR AND SSIM) AND THE REAL NOISY BLOOD VESSEL DATASET (IN SNR AND CNR). P-VALUES ARE CALCULATED BASED ON THE COMPARISON WITH OUR RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy input</th>
<th>DnCNN</th>
<th>U-Net</th>
<th>Residual U-Net</th>
<th>Ours w/o GC blocks</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>23.14±3.88***</td>
<td>26.44±4.33***</td>
<td>27.80±2.97***</td>
<td>25.94±3.38***</td>
<td>29.08±4.28</td>
<td>29.67±4.48</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.37±0.23***</td>
<td>0.56±0.18***</td>
<td>0.66±0.14</td>
<td>0.61±0.15</td>
<td>0.69±0.09***</td>
<td>0.63±0.12</td>
</tr>
<tr>
<td>SNR (dB)</td>
<td>26.64±2.03***</td>
<td>31.55±2.69***</td>
<td>40.92±3.58***</td>
<td>37.28±5.86***</td>
<td>43.78±15.94</td>
<td>47.24±6.60</td>
</tr>
<tr>
<td>CNR (dB)</td>
<td>9.93±1.82**</td>
<td>11.05±1.65</td>
<td>12.34±2.43</td>
<td>12.02±3.42</td>
<td>12.51±2.68</td>
<td>12.37±2.66</td>
</tr>
</tbody>
</table>

TABLE VI
QUANTITATIVE COMPARISON RESULTS (MEAN±STANDARD DEVIATION) USING 16 LEAF VEIN IMAGE PAIRS ACQUIRED UNDER HIGH EXCITATION LIGHT ENERGY (AS GROUND TRUTH) AND LOW EXCITATION LIGHT ENERGY (AS NOISY INPUT). P-VALUES ARE CALCULATED BASED ON THE COMPARISON WITH OUR RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>Noisy input</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>19.25±0.93***</td>
<td>22.25±0.83</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.17±0.01***</td>
<td>0.74±0.01***</td>
</tr>
</tbody>
</table>

DnCNN [43] is one of the early deep learning de-noising methods with high performance. U-Net [50] architecture has been widely and effectively applied in various biomedical imaging applications, while Residual U-Net [44] is one of the variants of U-Net. The results are shown in Table V.

According to Table V, our method outperforms all the compared deep learning-based methods in terms of the PSNR and SNR metrics. For the SSIM and CNR metrics, our method achieves similar or slightly better performance compared with other methods. Therefore, our developed GAN-based model with sophisticated design is generally better suited to the PAM de-noising task than the other methods in Table V.

F. Comparison of Image Data Acquired Under Different Excitation Light Energy

The experiments were conducted with low excitation light energy to obtain real noisy PAM images and high excitation light energy to obtain the corresponding ground truth. Sixteen leaf vein image pairs (i.e., 32 images in total) were acquired. Note that the excitation light energy used to acquire the ground truth image was 12.5 times higher than that used to acquire the real noisy input image. Fig. 8 shows the representative results from two image pairs. The de-noising effect is obvious in the recovered images by our method compared with the corresponding noisy input images. Table VI shows that our method achieves good performance with an average PSNR and SSIM improvement of 3.30 dB and 0.57, respectively.

Moreover, the in vivo experiments using low to high excitation light energy to obtain real noisy and clean PAM images of mouse ear blood vessels were conducted. Two image sets of blood vessels were acquired, each set under three different excitation light energy levels (80 nJ, 200 nJ, and 400 nJ for one set; 80 nJ, 200 nJ, and 320 nJ for the other set). Fig. 9 shows the acquired PAM images as noisy or clean input (the top row) and the corresponding recovered images by our method (the bottom row). For the input images, as expected, image quality (specifically, noisy vs. clean) is obviously improved from the low to high excitation light energy. By contrast, for the recovered images, the best image quality is achieved by the high excitation light energy, and satisfactory image quality is also obtained by the low excitation light energy (i.e., 80 nJ). Besides, by comparing between the input and recovered images, image quality is much improved for the case of lower excitation light energy (e.g., 80 nJ), while image quality looks similar or slightly improved for the case of higher excitation light energy (e.g., 400 nJ and 320 nJ). The improvement by comparing the images mentioned above is overall consistent with that by comparing the SNR and CNR values (indicated in Fig. 9) between the input and recovered images. Therefore, the in vivo results show that our CNN method would benefit noisy input images (significantly) as well as clean input images (to some degree), which also shows the potential of our CNN model for real clinical applications.

IV. DISCUSSION

The safety limitation of laser, the low signal intensity for deep tissues, the expectation of a cheap and low-intensity light source, and the electronic nature of ultrasound transducer are some of the reasons for the relatively apparent noise in PAM images, as detailed in [28]. One solution is to implement a powerful de-noising algorithm for PAM imaging modality, which is of necessity and importance for in vivo and clinical applications. However, existing PAM de-noising methods usually suffer from low speed, limited performance, and the requirement of manual adjustment. Therefore, in this work, we aim at developing a new de-noising method that can overcome the above issues.

In this study, we proposed a PAM de-noising network with attention mechanism (GC block) and the combined...
loss. The proposed deep learning-based method demonstrated excellent de-noising performance compared to existing methods (BM3D, KSVD, and WNNM) on both phantom and in vivo datasets. Specifically, quantitative metrics and qualitative observation of PAM images recovered by different methods were compared, and our method achieved the best results. The difference of our method using a GAN-based model with sophisticated design from other deep learning methods can be summarized in two aspects: (i) Improved U-Net shape architecture with attention blocks (detailed in subsection II-C) and (ii) the combined loss function (detailed in subsection II-C). In self-comparison using the synthetic in vivo blood vessel dataset, the model without GC blocks (while with other parts the same) leads to the decreases of PSNR and SSIM by 0.59 dB and 0.04 (Table V). Besides, our network has other advantages, as elaborated in the following.

According to Tables I-IV, our method shows the best mean values of the metrics, including PSNR and SSIM in Tables I and III and SNR and CNR in Tables II and IV, on all the four datasets. In addition, by checking p-values based on the comparison with our results, the results by all the other three methods are significantly different for both metrics (either “PSNR and SSIM” or “SNR and CNR”) for the three datasets in Tables I-III. For the dataset in Table IV, the results by KSVD are significantly different for both metrics (SNR and CNR), while the results by BM3D and WNNM also show significantly different for one metric SNR. The analysis of mean values and p-values of the metrics shows that our method evidently outperforms the other three methods for all the four datasets.

Conventional de-noising methods (e.g., BM3D, KSVD, and WNNM in this work) always require manually-set parameters to ensure the de-noising performance. For example, if the parameters of the assumed noise level are not carefully chosen, the de-noising performance could be greatly degraded. To ensure the optimal results using the conventional de-noising methods, it is often necessary to test different parameters of noise levels, and the best visual results need to be evaluated by humans. Specifically, we applied the parameter grid search to optimize the performance of the three methods for each image of the testing set. The procedure to find the best parameter of noise level is time-consuming due to the need for grid searching loop and human evaluation, making the conventional algorithms inconvenient and impractical to use. The intrinsic limitations of the compared methods may be due to the following reasons. First, the three methods are usually built based on the simplified noise priors (e.g., white Gaussian noise), yet the real-world noise is more likely the complex combination of various types of noise. Secondly, some pre-set parameters are usually needed (e.g., noise level as well as the number of iterations), making the three methods challenging to optimize the de-noising performance for different images.

By contrast, our method has strength in self-adaptability. Once the training dataset is well prepared without serious data distribution errors and the training is effectively achieved, the trained model can be applied to images with a wide range of noise levels without the need for manual settings of parameters. Therefore, in our demonstrations (section III), our method does not require prior information of the noise level of the noisy input images and can still produce reasonable de-noised images without strange artifacts. This shows that our method can adapt to various noise levels.

Furthermore, the time cost is another clear advantage over conventional methods. In our experiment, the three conventional methods (BM3D, KSVD, and WNNM) take from a few seconds to a few minutes to de-noise an input image with 256 × 256 pixels using a 3.20 GHz Intel Core i7-8700 CPU. By contrast, our method takes only 0.58 s to de-noise an input image of the same size using the same CPU. As CNNs are more suitable for GPU computing, our speed can be improved to 0.016 s/image using an Nvidia Titan RTX GPU. Hence, given real-time image acquisition (i.e., hardware acquisition), the proposed method is capable of online real-time processing, which is essential for clinical PAM applications.

Joseph et al. reported applying GAN for PA image enhancement with different purposes [47]. Compared with that...
work [47], our work has obvious differences and significances in several aspects. (i) In our work, we aim at de-noising in PAM modality. By contrast, Joseph et al. applied GAN to remove artifacts in PACT modality, which come from the limited bandwidth or view angle in PACT, but did not focus on de-noising or handling different types of noise. In principle, PAM does not have such artifacts because of different image reconstruction mechanisms from PACT. Therefore, the model developed in [47] cannot be directly applied to the de-noising of PAM images. (ii) In our work, targeting at the specific noise features, the attention mechanism and the combined loss function are introduced in the GAN model architecture to better handle the noise features in PAM images. By contrast, in [47], almost the original pix2pix GAN architecture was directly used without further modification or optimization tailored for specific purposes. (iii) In our work, a thorough description of PAM noise, the technical details of the GAN model architecture, the demonstrations of de-noising performance by various ex vivo and in vivo images, and in-depth analyses and suggestions are provided. By contrast, technical details, demonstrations, results analyses, and discussion in [47] are obviously less than those in our work. (iv) For the training dataset, in this work, the noisy PAM images are obtained by experimentally-acquired clean images with synthetic noise added afterward, while in [47], the PACT images were generated using the k-wave toolbox. In this regard, our work is expected to perform well because of the small domain gap between the synthetic noisy data and the real noisy data.

Currently, the three types of noise (Gaussian noise, Poisson noise, and Rayleigh noise) are considered in our synthesis operation. These three types of noise may be introduced into PAM images during the procedure of image acquisition according to corresponding physical principles. It would be challenging and almost impossible to consider all the real-world noise. On the other hand, based on our knowledge of the relevant PAI work [4], [8], [9], [14], [16], [48], we have considered most of the noise types. In addition, we tested our model using experimentally-acquired PAM data, and the de-noising results of real noisy images demonstrate that our method works well to handle the real-world noise.

Most commonly, PAM has been extensively utilized to visualize vasculature (e.g., [51], [52]). Besides, we conducted extended tests using the zebrafish with much different patterns and obtained good de-noising results (Fig. 7), showing the generality of the trained model to some extent. On the other hand, even if the current trained model may not work well for the patterns different from the training set, our model can be further developed to suit certain patterns as long as the corresponding datasets could be collected for training.

In this work, the trained model is validated on OR-PAM images. OR-PAM suffers shallow imaging depth, which may restrict some potential clinical applications. As AR-PAM and PACT enable deeper penetration than OR-PAM, further development and validation of our model for de-noising AR-PAM images and PACT images is of great value for clinical applications. Toward clinical translation, PA images from more types of specimens, including human tissues, can be collected as new datasets to refine our model. The abovementioned different types of PA images (AR-PAM images, PACT images, PA images from more types of specimens) could have quite different patterns and noise from those in OR-PAM images. Thus, the current model trained on OR-PAM images may not be directly applied to the different types of PA images for de-noising. As PA images have some common hardware (e.g., ultrasound transducer) and share some common noise, transfer learning should work to some extent. Besides, for transfer learning, an additional training dataset can be prepared with a reasonable amount of images (i.e., experimentally-acquired clear images), which makes the extension feasible to implement.

V. Conclusion

A deep learning-based de-noising method for PAM images was proposed and developed. The CNN effectively extracted multi-level image features and utilized attention mechanisms to improve the de-noising performance. The training schedule using GAN with the combined loss resulted in clean output. Three types of noise in PAM systems were considered when generating the synthetic noisy images for training and testing the model. Besides, real noisy images from both phantom and in vivo datasets were used to validate the models trained on synthetic datasets. The results show that our proposed method performs well in de-noising PAM images compared with other methods (BM3D, KSD, WNNM and other deep learning-based models). By comparing PAM images recovered by different methods, our method has advantages in effective noise removal, well-suppressed background, and few artifacts. With the self-adaptability to various noise levels and the fast speed for de-noising processing, the proposed method for PAM de-noising is promising for clinical applications.

APPENDIX

The codes of this work are available at https://github.com/Da-He/PAM_de-noising.

REFERENCES


