IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 30, 2020, accepted October 24, 2020, date of publication October 27, 2020, date of current version November 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3034230

Real-Time Ultrasound Image Despeckling Using Mixed-Attention Mechanism Based Residual UNet

YANCHENG LAN AND XUMING ZHANG

Key Laboratory of Molecular Biophysics, Ministry of Education, School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Xuming Zhang (zxmboshi@hust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1303100, and in part by the National Natural Science Foundation of China under Grant 61871440.

ABSTRACT Ultrasound imaging has been widely used for clinical diagnosis. However, the inherent speckle noise will degrade the quality of ultrasound images. Existing despeckling methods cannot deliver sufficient speckle reduction and preserve image details well at high noise corruption and they cannot realize real-time ultrasound image denoising. With the popularity of deep learning, supervised learning for image denoising has recently attracted considerable attention. In this paper, we have proposed a novel residual UNet using mixed-attention mechanism (MARU) for real-time ultrasound image despeckling. In view of the signal-dependent characteristics of speckle noise, we have designed an encoder-decoder network to reconstruct the despeckled image by extracting features from the noisy image. Furthermore, a lightweight mixed-attention block is proposed to effectively enhance the image features and suppress some speckle noise during the encoding phase by using separation and re-fusion strategy for channel and spatial attention. Besides, we have graded the speckle noise levels with a certain interval and designed an algorithm to estimate the noise levels for despeckling real ultrasound images. Experiments have been done on the natural images, the synthetic image, the image simulated using Field II and the real ultrasound images. Compared with existing despeckling methods, the proposed network has achieved the state-of-the-art despeckling performance in terms of subjective human vision and such quantitative indexes as peak signal to noise ratio (PSNR), structural similarity (SSIM), equivalent number of looks (ENL) and contrast-to-noise ratio (CNR).

INDEX TERMS Ultrasound image, speckle noise, supervised learning, mixed-attention mechanism, residual UNet.

I. INTRODUCTION

Ultrasound imaging has become one popular medical imaging technology due to its non-invasive, inexpensive and real-time advantages. However, the coherent nature of ultrasound imaging results in inherent speckle noise in the ultrasound image [1]. The presence of speckle noise reduces the resolution and contrast of the image, and adversely affects subsequent image processing and analysis tasks such as image segmentation, image registration, image feature extraction and recognition [2]. Therefore, speckle noise reduction from medical ultrasound images is highly important for improving image quality.

The distribution of ultrasound speckle noise is signal dependent and is governed by Fisher-Tippett distribution [3] or Gamma distribution [4], which is represented as:

$$u(x, y) = v(x, y) + v(x, y)^{\gamma} \eta(x, y) \tag{1}$$

where (x, y) is the pixel location, $v(x, y)$ is the noise-free image, $u(x, y)$ is the noisy image, $\eta(x, y)$ is Gaussian noise distributed with zero-mean and variance σ^2, and the factor γ is related to ultrasound devices and additional processing. Extensive studies indicate that $\gamma = 0.5$ can be used to model speckle noise in the ultrasound image [2].

Up to now, various methods have been proposed for ultrasound image despeckling. In general, the existing despeckling methods can be classified as frequency domain based methods and spatial domain based ones. As for the frequency domain based methods, the popular techniques are the wavelet based methods [5]–[8]. These methods work by transforming speckle noise into additive noise and then removing it within the wavelet domain. However, the despeckling performance of these methods is affected

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

because speckle noise in the real ultrasound images is not purely multiplicative noise and the artifacts related to the choice of mother wavelet may be introduced.

The traditional spatial domain based methods, such as Frost filter [9], Kuan filter [10], squeeze box filter (SBF) [11] and speckle reducing anisotropic diffusion filter (SRAD) [12] are based on local comparison of pixels. The shortcoming of these methods is that they cannot adequately reduce noise while preserving image details effectively. To address this problem, the non-local means (NLM) method has been proposed by Buades *et al.* [13]. This method explores the self-similarities between image patches instead of individual pixels, and restores each image pixel by the weighted average of all pixels in a search window. Despite the success in additive noise reduction, the NLM method by its very nature is unsuitable for speckle noise reduction. To overcome this drawback, several modified NLM approaches have been proposed for despeckling. Coupe *et al.* [14] have introduced the optimized Bayesian non-local means (OBNLM) filter which determines the similarity between two image patches based on the Pearson distance derived by the Bayesian framework instead of the Euclidean distance in the traditional NLM method. Yang *et al.* [15] have presented a hybrid despeckling approach which combines the NLM with the local statistics of noise. In this method, the local statistics of speckle noise is used to pre-filter the ultrasound image and the non-local similarity is computed based on the pre-filtered image. Santos *et al.* [16] have derived the new stochastic distances for the Fisher-Tippett distribution based on well-known statistical divergences, and used them as patch distance measures in a modified version of the BM3D algorithm for despeckling log-compressed ultrasound images. Yu *et al.* [2] have proposed the PCANet based NLM method, in which the intrinsic features of image patches extracted by the PCANet instead of the pixel intensities are introduced to determine the nonlocal similarity of ultrasound images.

The above-mentioned despeckling methods cannot deliver sufficient noise reduction while preserving image details especially at high speckle corruption. Meanwhile, most of these methods generally cannot realize real-time ultrasound image despeckling due to the involved complicated operations. The deep learning, as a popular algorithm in the field of machine learning, provides a possible and valuable solution to real-time and effective ultrasound image despeckling because it can automatically learn the intrinsic features from the training data, and can facilitate highly efficient image denoising.

The various deep learning models such as deep belief network (DBN) [17], stacked auto-encoder (SAE) [18], convolutional neural network (CNN) and recurrent neural network (RNN) have been proposed. Among these models, the CNN is very popular and many CNN-based models have been successfully applied to such image processing and analysis tasks as classification [19], [20], super-resolution [21], [22], segmentation [23]–[25] and image denoising [26]–[30]. For the image denoising task, Zhang *et al.* [26] have proposed a deep CNN denoiser (DCNND) for image restoration in which the dilated convolution is used to produce larger receptive field and the residual learning is adopted to maintain a noise map corresponding to the input noisy image. Later, they have taken one step forward by investigating the construction of feed-forward denoising convolutional neural network (DnCNN) [27] to embrace the progress in very deep architecture, learning algorithm and regularization method into image denoising. The two models perform well in removing Gaussian noise by using residual learning strategy, but they cannot work well for speckle noise removal because they cannot accurately estimate the residual term $v(x, y)^{\gamma}\eta(x, y)$ in equation (1). Chierchia *et al.* [28] have proposed a CNN for synthetic aperture radar (SAR) image despeckling using a residual learning strategy to recover the speckle component. Wang *et al.* [29] have proposed a SAR image despeckling CNN (ID-CNN) by integrating the convolutional layers involving batch normalization (BN) and the rectified linear unit (ReLU) with a componentwise division residual layer. The two methods work on the purely multiplicative noise model and use division or logarithmic transformation for residual learning. However, such a model cannot represent characteristics of speckle noise in the real ultrasound image.

In this paper, we have proposed a novel mixed-attention mechanism based residual UNet (MARU) for real-time speckle noise reduction. Furthermore, we have made improvements on the non-local neural network [31] and GCNet [32] and proposed a lightweight mixed-attention block which can maintain both channel and spatial attention using separation and re-fusion strategy with very little additional memory and time consumption. In order to cope with the complex noise situation in the real ultrasound images, we have graded the noise levels and designed an algorithm to estimate the levels of speckle noise for MARU based image despeckling. The experiments on the natural images, the synthetic image, the image generated by Field II and the real ultrasound images demonstrate the advantage of the proposed network over several traditional despeckling methods and deep learning based methods.

The remainder of this paper is structured as follows. The proposed despeckling method is detailed in Section II. Then we will make an analysis of the proposed network in Section III. In Section IV, experimental results of the proposed method and other compared methods on different test images are provided. Finally, conclusion and future research directions are given in Section V.

II. THE PROPOSED DESPECKLING METHOD

A. THE PROPOSED DEEP RESIDUAL UNet

The proposed network architecture is shown in Fig. 1(b), which is a deep residual UNet [33], [34] with mixed-attention mechanism. We have utilized two-stage down/up sampling and stacked two residual blocks [35], [36] as shown in Fig. 1(a) at each stage. Here we have replaced the original ReLU with LeakyReLU [37]. During the encoding phase, the input image will firstly pass through a convolutional layer

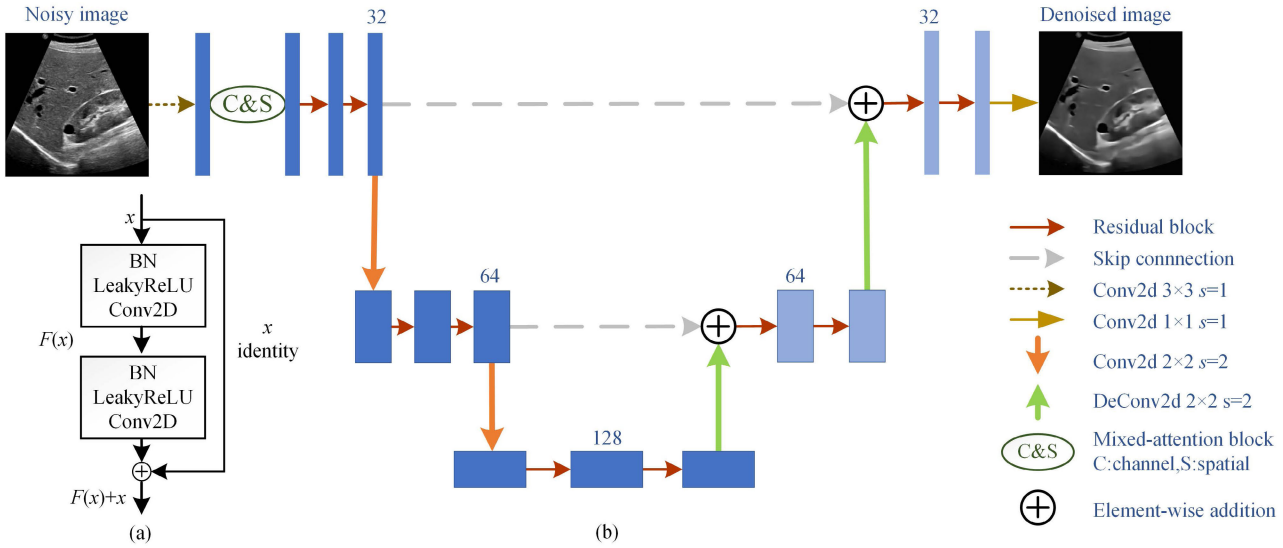


FIGURE 1. The framework of the proposed deep residual UNet with mixed-attention block. (a) The standard residual block with two pre-activated units, i.e. BN-LeakyReLU-Conv2D; (b) The architecture of the proposed network.

and a mixed-attention block. Leaving aside max-pooling or average-pooling, we will use a convolutional layer with stride of 2 to halve the size of feature maps [38] and double the number of channels to avoid the representational bottleneck mentioned in [39]. During the decoding phase, the deconvolution layer [40] will be used for upsampling and then the feature maps will be added with the skip connection between the encoding and decoding phases. Finally, a convolutional layer is used to fuse all feature maps into a despeckled image.

In order to further improve the despeckling performance, we have proposed a lightweight mixed-attention block based on the non-local network and GCNet and introduced it into our method. This block imitates the visual characteristics of the human eye and assigns weights to each pixel in each channel, that is, attention mechanism. The attention mechanism helps to enhance image features while suppressing noise. Here, the non-local network has been proposed by He *et al.* [31]. Inspired by the idea of non-local means, He *et al.* [31] first introduced this idea into the deep learning field and defined a generic non-local operation in the deep neural network as:

$$z_i = \frac{1}{C(u)} \sum_{\forall j} f(u_i, u_j) g(u_j) \tag{2}$$

where j is the index that enumerates all possible positions. u is the input image and z is the output image with the same size to u. A pairwise function f computes a scalar which represents the relationship between i and all j. The function g computes a representation of the input signal at the position j. The response is normalized by a factor $C(u)$.

The architecture of the non-local network (NLNet) is shown in Fig. 2(a). It is indeed a great idea, but the memory and time consumption will increase dramatically with the increasing image size. This is very disadvantageous for

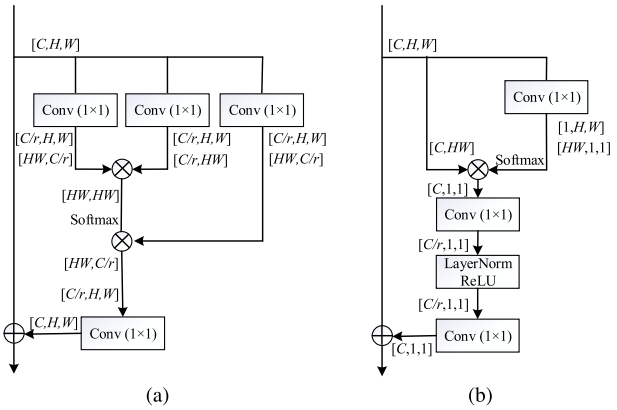


FIGURE 2. The architecture of the NLNet and the GCNet. (a)NLNet; (b)GCNet. Here r is the compression coefficient.

the real-time denoising of ultrasound images due to their relatively large size. To address this issue, Yue *et al.* [32] have proposed the global context network (GCNet) as shown in Fig. 2(b) based on the observation that the global contexts modeled by the non-local network are almost same for different query positions within an image. This means that we can use the global context of one point to represent all points of an image to greatly reduce the calculation. Correspondingly, equation (2) can be simplified as:

$$z_1 = \frac{1}{C(u)} \sum_{\forall j} f(u_1, u_j) g(u_j) \tag{3}$$

where the output z_1 is the global context of each point in the input u.

Based on the above researches, we have proposed the mixed-attention block with both channel and spatial attention using separation and re-fusion strategy. The architecture of

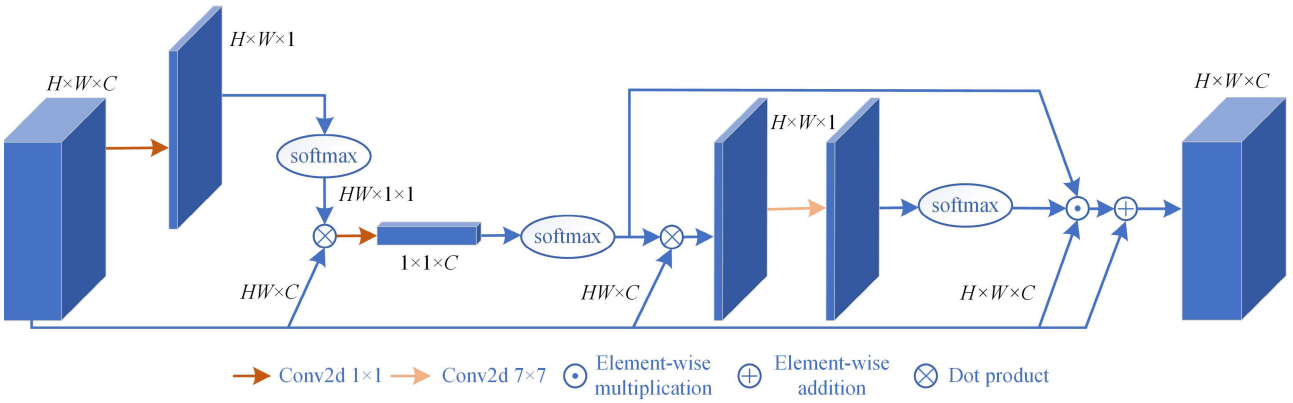


FIGURE 3. The architecture of the proposed mixed-attention block. The operations of adjusting the shape of tensor are omitted for simplicity.

this block is shown in Fig. 3. Firstly, the global context of each channel is modeled as done in the GCNet, and then channel attention is obtained after interaction between channels and softmax activation. Secondly, all channels are fused using the obtained channel weights, and then the spatial attention will be obtained after the interaction between pixels and softmax activation. Finally, the channel weights, spatial weights and inputs are multiplied for each point to produce the scaled feature maps. Besides, residual connection is added for identity mapping. This attention module is modified as:

$$F' = (1 + M_c \cdot M_s) \cdot F \tag{4}$$

where M_c is the channel attention matrix with the shape of $[1, 1, C]$, and M_s is the spatial attention matrix with the shape of $[H, W, 1]$; F and F' are the input feature maps and the output ones, respectively.

B. SPECKLE NOISE GRADING AND ESTIMATION

Considering that σ is unknown for the real ultrasound images, We have graded the noise levels and designed an estimation algorithm. Because it is unrealistic to train the model for each σ value, it is necessary to grade the noise levels according to a certain interval. We have set the interval of grading as 0.25 based on the comparative experimental results in Section III(C).

When it comes to noise level estimation, we can get the $\eta(x, y)$ item by inverting equation (1). The problem is that there are not noise-free images for real ultrasound images. However, the mean value of the smooth area before and after adding noise can be considered to be almost same due to Gaussian distribution characteristics of $\eta(x, y)$. Based on the above consideration, we will divide the whole image into many approximately uniform sub-areas and calculate their mean standard deviation by inverting equation (1) as:

$$\sigma = \frac{1}{N} \sum_{i=1}^{N} std((u_i - \overline{u_i})/\sqrt{\overline{u_i}}) \tag{5}$$

where u_i is the sub-area, $\overline{u_i}$ is the mean of u_i, N is the number of sub-areas and std denotes the standard deviation. Through experiments, we have determined the size of sub-areas to be 6×6 pixels. Since there is zero-filled region in the real ultrasound images, we will only select the areas whose mean value is greater than zero. Besides, this estimation process is only required once at the beginning of a clinical inspection.

C. TRAINING OF THE PROPOSED NETWORK

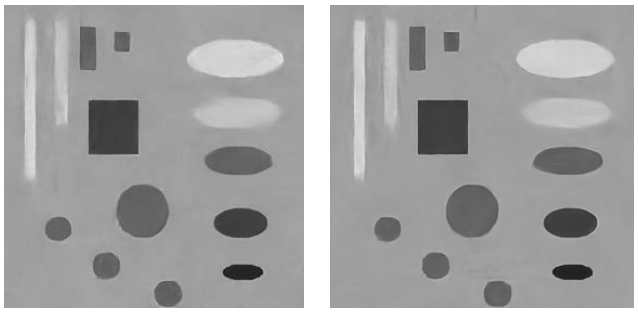
To train our network, we will use the Berkeley segmentation dataset (BSD400) [27] consisting of 400 images of size 180×180 for training. Considering that the total receptive field during the encoding phase is no more than 64, we will set the patch size as 64×64 and accordingly obtain about 25600 patches after data augmentation such as scale, flip and rotation for training. Then we will add speckle noise to these patches to produce the noisy image patches according to equation (1).

The proposed network is realized with Python based on Keras 2.2.4 on a Ubuntu 16.04, and it is run on a computer with a Core I7-6950X CPU and 96G RAM. The NVIDIA GTX 1080Ti GPU with CUDA 10.1 is used for acceleration. The despeckling network will be trained using the Adam optimizer with default setting and mean square error (MSE) loss with total variation (TV) regularization [29] for 100 epochs. The loss function is defined as:

$$L = ||v - v'||_2^2 + \lambda_{TV}(||\nabla_h v'||_2^2 + ||\nabla_v v'||_2^2) \tag{6}$$

where v' is the output and v is the label, and ∇_h/∇_v denotes the gradient operator along the horizontal/vertical direction. λ_{TV} is a regular coefficient set as 0.05.

Here, it should be noted that although the MSE loss has shown to work well on many image restoration tasks, it may result in various artifacts on the final estimated image. To overcome this problem, the TV regularization is utilized here to maintain the smoothness of the image and attenuate the artifacts that may be caused by image denoising. As shown in Fig. 4, we can see that the despeckled result using



(a) (b)

FIGURE 4. The denoised result of the proposed method tested on a synthetic image with or without TV regularization. (a) MSE loss with TV regularization, (b) MSE loss without TV regularization.

TV regularization is smoother and involves less artifact than that resulting from only using MSE loss.

D. APPLICATION OF THE MARU METHOD

When the proposed MARU model is trained, it can be used for image despeckling. Fig. 5 shows the application framework of the proposed MARU method. For a real ultrasound image, we will firstly estimate its noise standard deviation and determine the noise level using the speckle noise estimation method in Section II(B). Secondly, we will select the corresponding trained MARU model for despeckling according to the determined noise level.

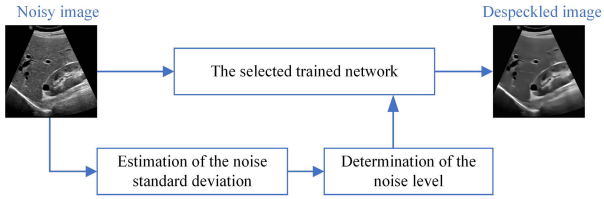


FIGURE 5. The application framework of the proposed MARU method.

III. ANALYSIS OF THE PROPOSED NETWORK

A. RESIDUAL CONNECTION AND NETWORK DEPTH

To verify the role of residual connection, we will conduct several comparative tests with or without residual connections on a test dataset containing 68 natural images from BSD68 [43]. Besides, we will also conduct several comparative tests with different sampling stages of residual UNet to determine the proper network depth.

In order to quantitatively measure the despeckling performance, two well-known evaluation indexes such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [41] are used for performance appreciation, which are defined as:

$$PSNR = 10\lg\left(\frac{255^2}{\frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}(v'(i,j)-v(i,j))^2}\right) \quad (7)$$

$$SSIM = \frac{(2\mu_{v'}\mu_v + C_1)(2\delta_{v'v} + C_2)}{(\mu_{v'}^2 + \mu_v^2 + C_1)(\delta_{v'}^2 + \delta_v^2 + C_2)} \quad (8)$$

where W and H represent the width and the height of the image, respectively. v is the noise-free image and v' is the denoised image. μ_v and $\mu_{v'}$ are the mean intensity of images v and v', respectively. δ_v and $\delta_{v'}$ are the standard deviation of images v and v', respectively. $\delta_{v'v}$ is the covariance between images v and v'. C_1 and C_2 are the small constants to stabilize SSIM.

Fig. 6(a) and Table 1 show that the network performance improves as the number of sampling stages increases. However, each additional sampling stage will increase the number of parameters by more than four times, which will lead to more training time and prediction time and is more likely to cause overfitting. The observation from the validation curve of the network at different depths in Fig. 6(a) shows that the network has already involved overfitting when the sampling stage is 3. Considering that low-level features are more conducive to protecting image details for denoising task, we have determined the sampling stage as 2. Besides, Fig. 6(b) shows that residual connection can help the network converge better and faster. Thus, we have introduced the residual connection into our network.

B. MIXED-ATTENTION BLOCK

In order to prove the validity of this mixed-attention block, we will insert it and GCNet respectively behind the first convolutional layer of the baseline (i.e., residual UNet) where the backward transmission of noise information can be suppressed more effectively. The comparative test results on BSD68 are shown in Table 2. It can be seen that the average PSNR value of baseline+mixed-attention block is 0.21dB higher than that of baseline and 0.11dB higher than that of baseline+GCNet. Clearly, the mixed-attention block provides the performance improvement with very little extra parameters and time consumption.

C. NOISE LEVEL GRADING AND ESTIMATION

To verify the reasonableness of the grading interval of 0.25 and the effectiveness of the estimation algorithm, we have conducted a series of comparative experiments. Fig. 7 shows the despeckled results of the synthetic image corrupted with different speckle noise using the proposed model. From Fig. 7, we can see that a specific model using higher noise levels than the real ones can ensure good restoration results and it will produce residual noise when the noise levels used for the network model are lower than the real ones. This observation demonstrates that it is preferable to choose a model whose noise levels should not be lower than the real ones. Meanwhile, the comparison between Fig. 7(a) and Fig. 7(c) shows that although the former is a little smoother and the latter has slightly clearer boundaries, the difference between the two images is very small. The comparison indicates that such an interval of 0.25 is reasonable.

Besides, we have tested the estimation algorithm on BSD68 with different noise levels and the results are listed in Table 3. We can see that the proposed estimation algorithm can generally estimate the levels of speckle noise with

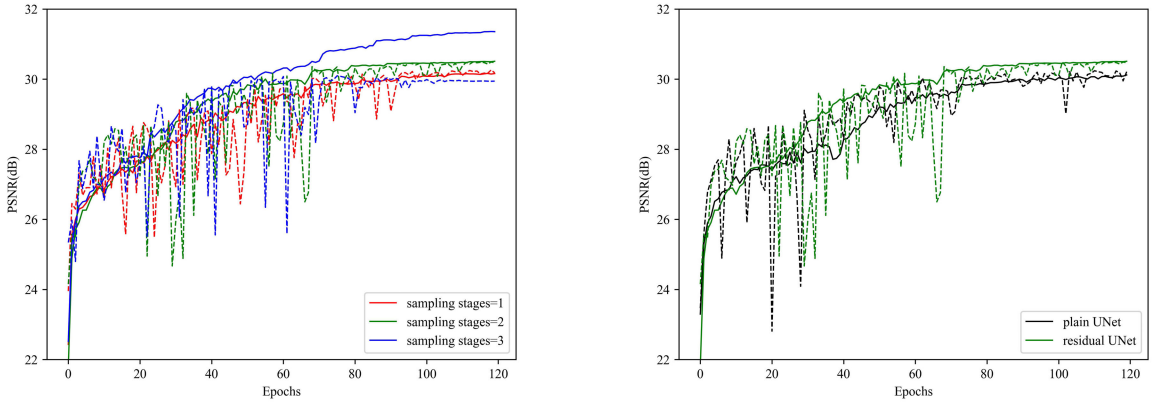


FIGURE 6. The training(solid)/validation(dotted) curves of different network architectures. (a) Residual UNet with sampling stages ranging from 1 to 3. (b) The curve of UNet with or without residual connection when the number of sampling stages is 2.

TABLE 1. The comparative results of different sampling stages and wether to utilize residual connection on BSD68 when $\sigma = 2.00$.

Sampling stages	PSNR(dB)	SSIM	Average time(ms)	Parameters(M)	Depth	Residual connection
1	31.02	0.86	30	0.24	16	Yes
2	31.10	0.87	37	1.05	26	Yes
2	30.99	0.87	35	1.05	26	No
3	30.79	0.86	41	4.27	36	Yes

TABLE 2. The PSNR, SSIM, parameters and time consumption of baseline structure and structures added with GCNet and mixed-attention block operating on BSD68 when $\sigma = 2.00$.

Structure	PSNR(dB)	SSIM	Average time(ms)	Parameters(M)
Baseline	31.10	0.87	37	1.05
Baseline+GCNet	31.20	0.87	38	1.05
Baseline+mixed-attention block	31.31	0.87	38	1.05

TABLE 3. The estimated and real noise levels for the BSD68 corrupted with different speckle noise.

Real noise levels	Mean of estimated σ	Variance of estimated σ	Mean estimation error	Chosen noise levels
σ=2.00	2.12	0.05	0.12	2.25
σ=3.00	3.09	0.05	0.09	3.25
σ=4.00	3.99	0.06	0.01	4.00
σ=5.00	4.86	0.08	0.14	5.00

acceptable errors in that the estimated noise levels are slightly higher than the real low noise ones and slightly lower than the real medium and high level ones.

IV. EXPERIMENTAL RESULTS

To demonstrate the superiority of the proposed network in terms of despeckling performance, it will be compared with such traditional well-known despeckling algorithms as SBF, SRAD, OBNLM and PCA-NLM methods and such deep learning based methods as ID-CNN, DCNND and DnCNN. Experiments have been done on the BSD68, the synthetic image, the simulated image by Field II and the real ultrasound images.

The datasets used in this paper are listed in Table 4. The Set25600 is cropped from BSD400 after data augmentation such as scale, flip and rotation for training. Besides, one percent of this set is randomly chosen for cross-validation. The BSD68 will be used to compare PSNR and SSIM values and the ultrasound images of different organs will be used to test the performance of all evaluated methods.

A. THE BSD68

This experiment is conducted on the BSD68 corrupted by various levels of speckle noise based on equation (1) with $\sigma = 2.0$, 3.0, 4.0 and 5.0. Table 5 lists the average PSNR and SSIM values of all evaluated methods. Besides, the

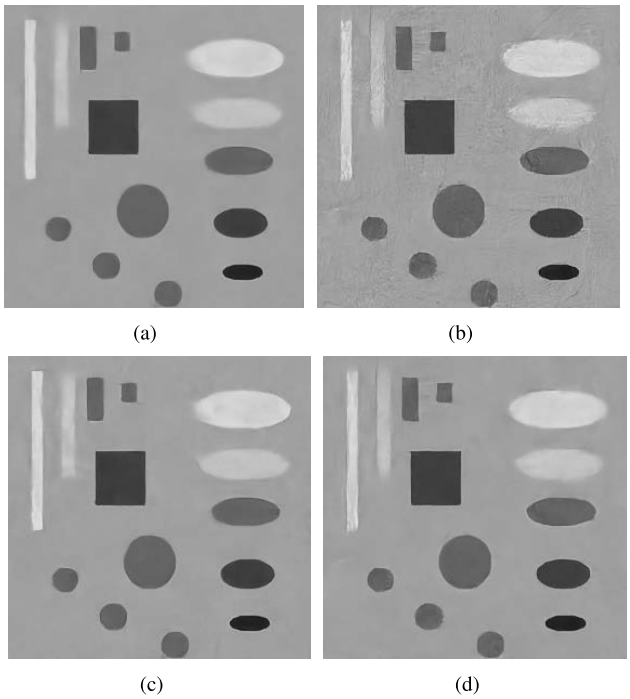


FIGURE 7. Despeckled results of the synthetic image corrupted with different speckle noise using the proposed model. (a) Model using $\sigma = 3.00$ when the real $\sigma = 2.75$, (b)Model using $\sigma = 3.00$ when the real $\sigma = 3.25$, (c) Model using $\sigma = 2.75$ when the real $\sigma = 2.75$, (d)Model using $\sigma = 3.25$ when the real $\sigma = 3.25$.

TABLE 4. The datasets used in this paper.

Datasets	Number	Size	Purpose
BSD400	400	180×180	Training
Set25600	25600	64×64	Training
Synthetic image	1	280×280	Testing
Simulated image	1	240×332	Testing
Ultrasound images	12	/	Testing
BSD68	68	480×320	Testing

parameters, depth, floating point operations (FLOPs) and average time for CNN-based methods are also listed in Table 6. The highest PSNR and SSIM values at each noise level are marked in bold. The observation from Table 5 shows that the deep learning based methods generally provide higher PSNR and SSIM values than the traditional despeckling methods except that the ID-CNN has relatively poor performance at high noise levels. However, the proposed MARU outperforms other networks due to its higher PSNR and SSIM at each noise level. The advantage of the MARU is due to its enhanced feature extraction ability resulting from the distinctive network structure and the increase in network depth and parameters. In addition, due to the adoption of encoding-decoding network structure, our method involves fewer FLOPs and the inference speed is faster than DnCNN.

B. THE SYNTHETIC IMAGE

Fig. 8 shows the despeckled results of SBF, SRAD, OBNLN, PCA-NLM, ID-CNN, DCNND, DnCNN and the proposed MARU on the synthetic image with $\sigma = 3.0$. Obviously, the SBF and SRAD methods perform poorly in speckle reduction and provide jagged or over-smoothed denoised results. By comparison, the NLM based methods deliver sufficient speckle reduction. However, the OBNLM and PCA-NLM methods generate the blurred boundary and some artifacts as shown in Fig. 8(e) and Fig. 8(f). As for deep learning based methods, the ID-CNN and DCNND remain the residual noise to different extent and generate the obvious artifacts as shown in Fig. 8(g) and Fig. 8(h). The DnCNN also produces the unwanted artifacts in Fig. 8(i). By comparison, the MARU not only removes speckle noise effectively but also preserves image details well with very little artifacts as shown in Fig. 8(j).

C. THE SIMULATED IMAGE

A more challenging and relevant image has been generated for the cyst phantom based on Field II simulation. The cyst phantom consists of a collection of point targets, five cyst regions and five highly scattering regions. The simulated image is shown in Fig. 9(a). For this simulated image, its restoration is highly difficult because the left-most point targets are heavily corrupted by noise and they are significantly smaller than other point targets. Fig. 9 shows the denoised results for all evaluated methods. In Fig. 9(b) and Fig. 9(c),

TABLE 5. The average PSNR and SSIM values of various despeckling methods on the BSD68 with different levels of speckle noise.

Methods	PSNR(dB)/SSIM			
	σ=2.0	σ=3.0	σ=4.0	σ=5.0
SBF	24.65/0.62	23.94/0.58	23.20/0.54	22.53/0.51
SRAD	27.14/0.76	26.01/0.72	25.11/0.69	24.25/0.66
OBNLM	28.07/0.77	26.55/0.72	25.38/0.67	24.39/0.63
PCA-NLM	28.26/0.78	26.79/0.73	25.62/0.69	24.39/0.63
ID-CNN	30.33/0.84	28.01/0.75	25.69/0.63	22.92/0.51
DCNND	30.69/0.86	29.10/0.82	27.87/0.77	27.03/0.74
DnCNN	30.93/0.86	29.26/0.82	27.89/0.77	26.73/0.73
MARU	**31.31/0.87**	**29.40/0.83**	**28.11/0.79**	**27.28/0.76**

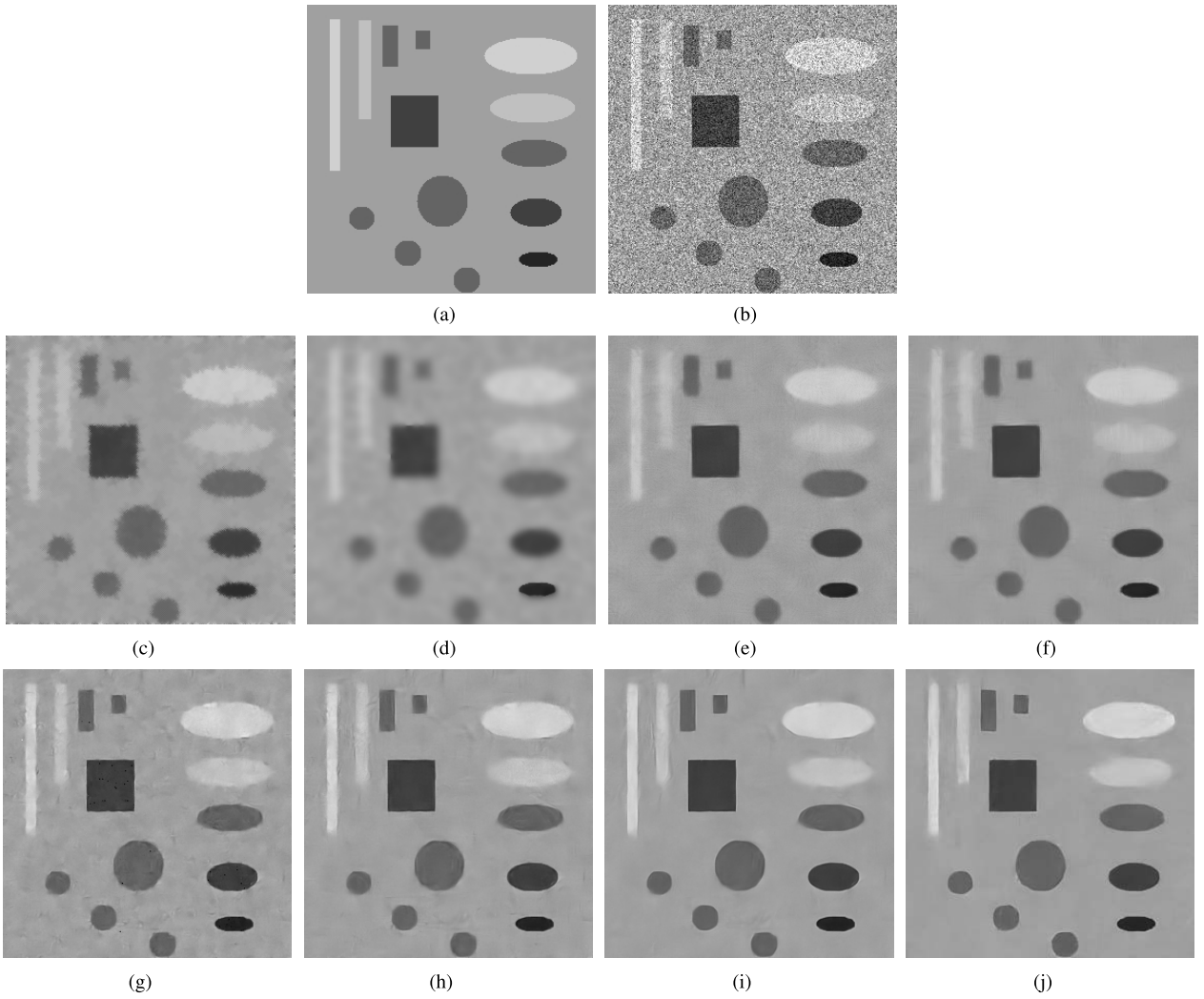


FIGURE 8. Visual comparison of despeckled results of various despeckling methods on the synthetic image with $\sigma = 3.0$. (a)The synthetic image, (b)Noisy image, (c)SBF, (d)SRAD, (e)OBNLM, (f)PCA-NLM, (g)ID-CNN, (h)DCNND, (i)DnCNN, (j)MARU.

TABLE 6. The depth, parameters, FLOPs and average time of CNN-based methods on BSD68.

Methods	ID-CNN	DCNND	DnCNN	MARU
Depth	8	7	17	26
Parameters(M)	0.22	0.19	0.56	1.05
FLOPs(G)	34.22	28.55	85.27	31.27
Average time(ms)	20	28	43	38

the SBF method leads to the jagged boundary and performs badly in speckle noise suppression while the SRAD method produces the highly blurred image. As for the OBNLM and PCA-NLM methods, they generate better results with less remaining noise but they cannot recover the point targets effectively as shown in Fig. 9(d) and Fig. 9(e). For the deep learning based methods, the ID-CNN and DCNND remain noticeable speckle noise in the despeckled image and generates some artifacts in Fig. 9(f) and Fig. 9(g). Although the DnCNN generates better despeckled result than the ID-CNN and DCNND, it leads to the loss of some point targets as shown in Fig. 9(h). As for the proposed method, the background area has been well smoothed, and the sharpness of point targets has been preserved relatively well as shown in Fig. 9(i). The visual comparison indeed demonstrates the advantage of the MARU over other compared methods in both speckle noise reduction and detail preservation.

Since there are no noise-free images for the simulated image and real ultrasound images so that PSNR and SSIM values cannot be calculated, two widely used evaluation indexes, i.e., equivalent number of looks (ENL) and contrast-to-noise ratio (CNR) [42] are utilized to quantitatively appreciate the despeckling performance, which are defined as:

$$ENL = \frac{\mu_b^2}{\sigma_b^2} \tag{9}$$

$$CNR = \frac{|\mu_b - \mu_o|}{\sqrt{\sigma_b^2 + \sigma_o^2}} \tag{10}$$

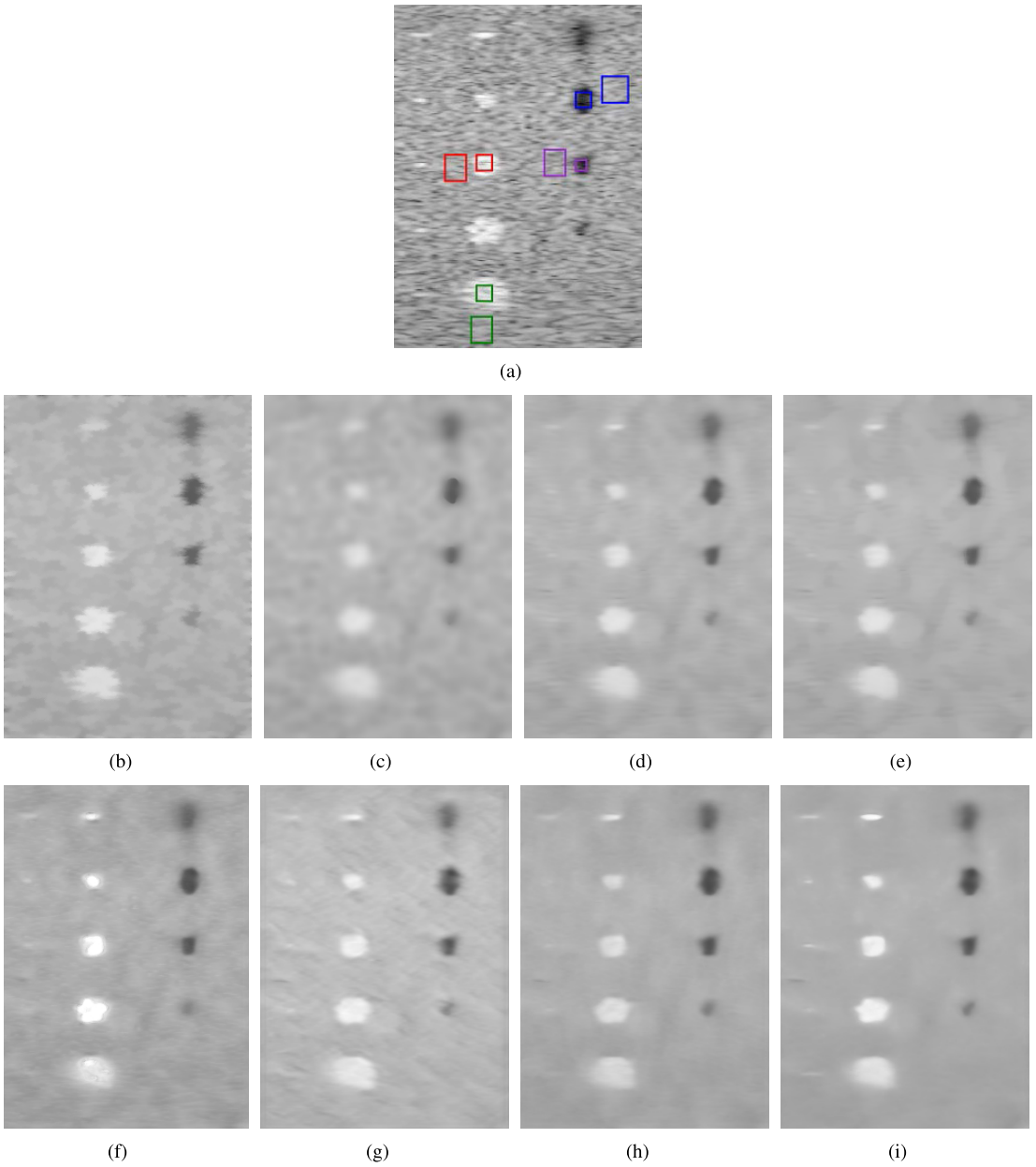


FIGURE 9. Visual comparison of despeckled results of various despeckling methods on the Field II simulated image. (a)The simulated image and four ROIs marked with different colors, (b)SBF, (c)SRAD, (d)OBNLM, (e)PCA-NLM, (f)ID-CNN, (g)DCNND, (h)DnCNN, (i)MARU.

where μ_o and μ_b denote the mean intensity of object and background regions, respectively. σ_o and σ_b denote the standard deviation of object and background regions, respectively.

Four pairs of regions of interest (ROIs) marked with various colors as shown in Fig. 9(a) are selected to evaluate the despeckling performance of these compared methods. The ENL and CNR values are listed in Table 7. For ENL, the proposed method provides the highest values in all ROIs, which means that our method produces the best despeckled result for the background area. As for CNR, the proposed method achieves the highest values for ROI 1 and ROI 2, and slightly smaller values than the highest ones for ROI 3 and ROI 4. The ENL and CNR comparisons indicate that our method can preserve image details well while removing speckle noise effectively.

D. THE REAL ULTRASOUND IMAGES

To further verify the practicality of the proposed MARU method, it will be applied to despeckling the real ultrasound images of cyst, fetus and kidney. The despeckled results of the MARU method will be compared with those of other evaluated methods.

TABLE 7. The ENL/CNR for the evaluated methods implemented in four ROIs in the simulated image.

Methods	ROI 1 (red)	ROI 2 (green)	ROI 3 (blue)	ROI 4 (darkorchid)
simulated image	162.33/1.96	175.95/2.68	220.30/4.25	197.57/3.37
SBF	1698.11/6.49	982.61/8.21	1536.08/6.57	1590.48/5.47
SRAD	2196.72/6.35	1461.39/9.56	2112.92/7.03	2106.82/**5.98**
OBNLM	4864.95/6.99	2374.69/11.19	4384.15/6.89	3561.27/5.55
PCA-NLM	5818.46/7.62	2412.04/11.92	6599.81/**7.09**	4151.95/5.43
ID-CNN	2756.96/5.10	999.53/6.85	1742.74/7.80	1700.59/5.44
DCNND	2993.06/8.09	1232.26/7.91	2652.06/6.41	4300.19/5.43
DnCNN	8155.16/8.72	2316.44/12.93	4685.74/7.27	3673.46/5.34
MARU	**12030.70/9.14**	**3492.02/15.50**	**8472.30**/6.40	**6340.05**/5.58

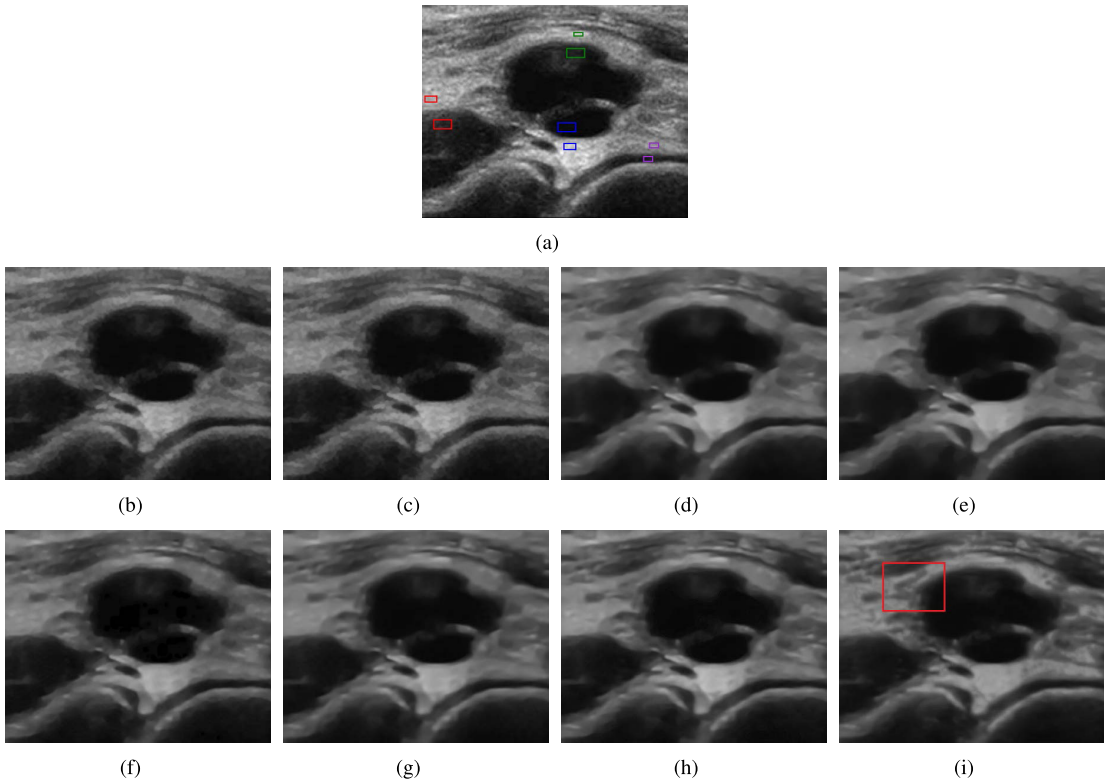


FIGURE 10. Visual comparison of despeckled results of various despeckling methods on a real cyst ultrasound image. (a)The real cyst image with four ROIs marked with different colors, (b)SBF, (c)SRAD, (d)OBNLM, (e)PCA-NLM, (f)ID-CNN, (g)DCNND, (h)DnCNN, (i)MARU.

Fig. 10 shows the despeckled results for the various methods on the real cyst ultrasound image [44] whose noise level is chosen to be 2.75 for our model based on the noise estimation result. Clearly, the results of the SBF and SRAD methods remain much noise in the denoised images. As for such NLM based methods as the OBNLM and the PCA-NLM, they smooth the structure information somewhat excessively as shown in Fig. 10(d) and Fig. 10(e). The ID-CNN, DCNND and DnCNN methods lead to the loss of some image details as shown in Fig. 10(f)-10(h). By comparison, the MARU can smooth out speckle noise effectively while better preserving the image details such as the fine details marked with the red box in Fig. 10(i).

To quantitatively analyze the despeckling performance of various methods, we will select four ROIs as shown in Fig. 10(a) to calculate their ENL and CNR. Table 8 lists ENL and CNR values for each evaluated method. Consistent with the visual effects in Fig. 10, the ENL values of NLM based methods are generally higher while our method follows closely. As for CNR, the proposed method achieves the highest values in all regions. the comprehensive consideration of ENL and CNR demonstrates that our method performs very well in speckle noise suppression and detail preservation.

Besides, another clinical fetus ultrasound image [45] is used to further visually compare the despeckling performance of all evaluated methods. For this image, the chosen noise

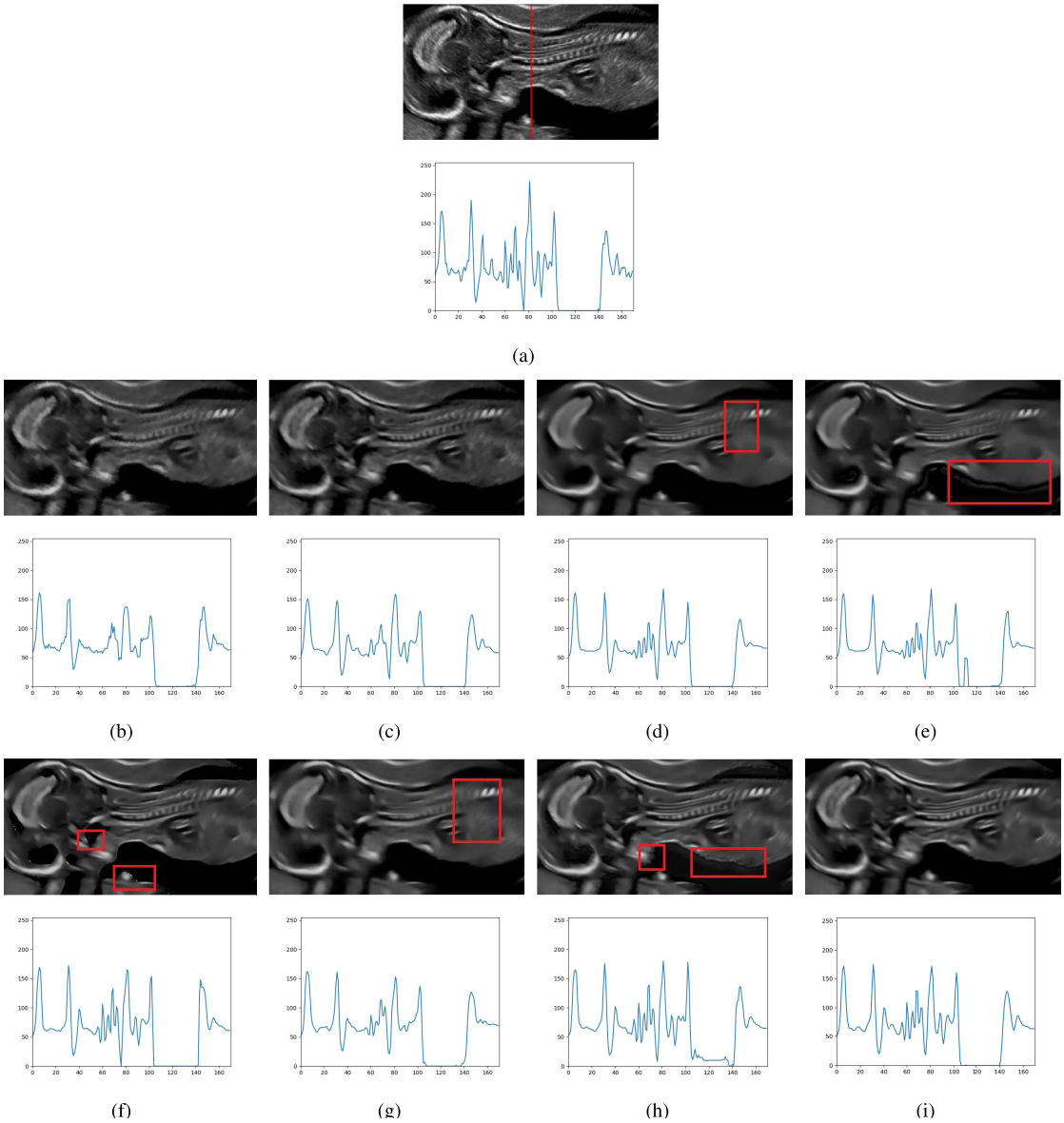


FIGURE 11. Visual comparison of despeckled results and profiles of pixel intensities along the highlighted line for the various despeckling methods implemented on a real fetus ultrasound image. (a) The fetus ultrasound image, (b) SBF, (c) SRAD, (d) OBNLM, (e) PCA-NLM, (f) ID-CNN, (g) DCNND, (h) DnCNN, (i) MARU.

TABLE 8. The ENL and CNR for the evaluated methods implemented in four ROIs in the cyst ultrasound image.

Methods	ROI 1 (red)	ROI 2 (green)	ROI 3 (blue)	ROI 4 (darkorchid)
Noisy image	33.88/11.23	39.76/13.32	68.33/18.59	36.99/11.16
SBF	51.36/15.78	59.71/24.53	172.83/23.65	74.31/17.77
SRAD	37.78/17.12	47.45/25.47	46.38/28.04	41.20/17.58
OBNLM	67.02/22.28	160.92/18.56	**327.42**/36.94	123.81/24.87
PCA-NLM	**84.56**/25.61	**213.93**/18.69	298.61/51.88	126.05/28.79
ID-CNN	48.73/19.80	91.20/25.70	3.09/23.63	82.01/24.88
DCNND	56.54/22.12	157.52/20.27	113.89/40.93	57.52/23.93
DnCNN	59.79/25.26	99.52/23.49	61.47/31.56	79.44/26.58
MARU	83.85/**27.09**	140.66/**30.81**	205.50/**93.24**	**190.35/40.20**

level is 2.50 based on the noise estimation result. To make a clear comparison, the profiles of pixel intensities along the highlighted line plotted in the original image are also shown in Fig. 11(a). As shown in Fig. 11(b) and Fig. 11(c), the SBF

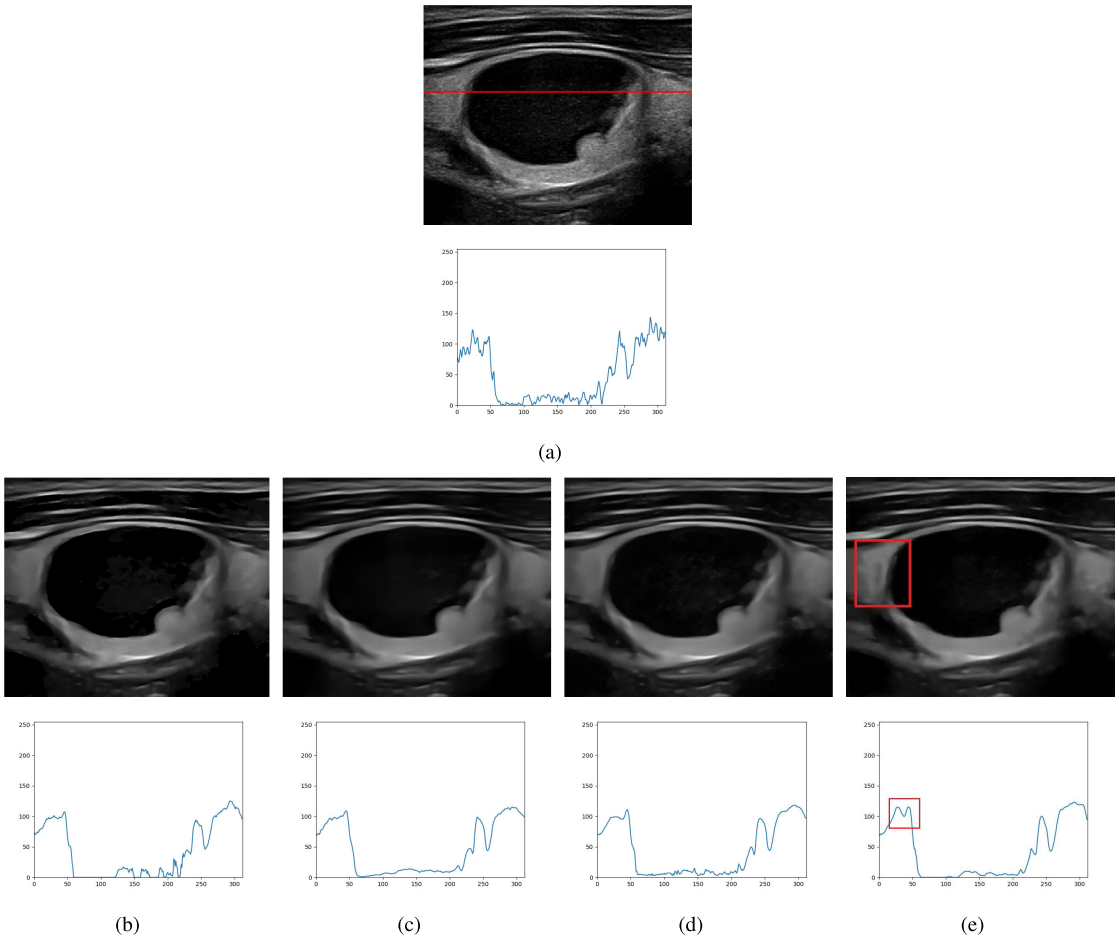


FIGURE 12. Visual comparison of despeckled results and profiles of pixel intensities along the highlighted line for the various despeckling methods implemented on a real liver ultrasound image. (a)The kidney ultrasound image, (b)ID-CNN, (c)DCNND, (d)DnCNN, (e)MARU.

and SRAD methods perform poorly in speckle reduction. The two NLM based methods both over-smooth the images to some extent. Meanwhile, the OBNLM method leads to the loss of such details as spine marked with the red box in Fig. 11(d) and the PCA-NLM method introduces very obvious artifacts marked with the red box in Fig. 11(e). As for deep learning based methods, the ID-CNN and the DnCNN generate some artifacts around the boundary marked with the red boxes in 11(f) and Fig. 11(h). The DCNND damages such details as the spine marked with a red box in 11(g). The denoised result and the profile of pixel intensities in Fig. 11(i) show that the proposed method can suppress speckle noise more sufficiently than the DCNND and it preserves such details as the spine better than the DnCNN without introducing artifacts.

We further compare the deep learning based methods on the real kidney ultrasound images [46]. Likewise, the visual comparison of despeckled results and the profiles of pixel intensities along the highlighted lines in the original images will be made. For this artery image, the chosen noise level is 3.00 based on the noise estimation result. It is easy to see from Fig. 12 that the proposed MARU method can produce smoother background regions and preserve image structure information better than the compared deep learning based methods, especially in the red box area.

V. CONCLUSION

In this paper, a novel convolutional neural network is proposed for real-time ultrasound image despeckling. The introduction of residual network and the mixed-attention mechanism boosts the denoising performance of this network effectively. Besides, the speckle noise estimation algorithm automates the denoising process and ensures its application to real ultrasound image despeckling. Experiments on the BSD68, the synthetic image, the simulated and clinical ultrasound images quantitatively demonstrate the advantage of the proposed method over other compared methods in terms of PSNR, SSIM, ENL and CNR. Visual comparison shows that the proposed method outperforms the compared despeckling methods in terms of speckle noise reduction and detail preservation. Our future work will be focused on the estimation of speckle noise levels in the real ultrasound images using deep

learning and extension of the proposed method to denoising other medical images such as computed tomography (CT), magnetic resonance (MR) and positron emission computed tomography (PET) images.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments that improved the quality of this article greatly. They would also like to thank the Medical Ultrasound Laboratory, Huazhong University of Science and Technology, for providing hardware support.

REFERENCES

[1] Q. Zhang and B. Li, "Formation principle and model of ultrasonic speckle noise," *Electron. Technol. Softw. Eng*, vol. 14, pp. 118–120, Jul. 2014.

[2] H. Yu, M. Ding, X. Zhang, and J. Wu, "PCANet based nonlocal means method for speckle noise removal in ultrasound images," *PLoS ONE*, vol. 13, no. 10, Oct. 2018, Art. no. e0205390.

[3] G. Slabaugh, G. Unal, T. Fang, and M. Wels, "Ultrasound-specific segmentation via decorrelation and statistical region-based active contours," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jun. 2006, pp. 45–53.

[4] Z. Tao, H. D. Tagare, and J. D. Beaty, "Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1483–1491, Nov. 2006.

[5] C. R. Mittermayr, S. G. Nikolov, H. Hutter, and M. Grasserbauer, "Wavelet denoising of Gaussian peaks: A comparative study," *Chemometric Intell. Lab. Syst.*, vol. 34, no. 2, pp. 187–202, Sep. 1996.

[6] D. Boto-Giralda, F. J. Díaz-Pernas, D. González-Ortega, J. F. Díez-Higuera, M. Antón-Rodríguez, M. Martínez-Zarzuela, and I. Torre-Díez, "Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 25, no. 7, pp. 530–545, Aug. 2010.

[7] M. Srivastava, C. L. Anderson, and J. H. Freed, "A new wavelet denoising method for selecting decomposition levels and noise thresholds," *IEEE Access*, vol. 4, pp. 3862–3877, Jul. 2016.

[8] Z. Gan, F. Zou, N. Zeng, B. Xiong, L. Liao, H. Li, X. Luo, and M. Du, "Wavelet denoising algorithm based on NDOA compressed sensing for fluorescence image of microarray," *IEEE Access*, vol. 7, pp. 13338–13346, Jan. 2019.

[9] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.

[10] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive restoration of images with speckle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 373–383, Mar. 1987.

[11] P. C. Tay, C. D. Garson, S. T. Acton, and J. A. Hossack, "Ultrasound despeckling for contrast enhancement," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1847–1860, Jul. 2010.

[12] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1260–1270, Nov. 2002.

[13] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 60–65.

[14] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2221–2229, Oct. 2009.

[15] J. Yang, J. Fan, D. Ai, X. Wang, Y. Zheng, S. Tang, and Y. Wang, "Local statistics and non-local mean filter for speckle noise reduction in medical ultrasound image," *Neurocomputing*, vol. 195, pp. 88–95, Jun. 2016.

[16] C. A. N. Santos, D. L. N. Martins, and N. D. A. Mascarenhas, "Ultrasound image despeckling using stochastic distance-based BM3D," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2632–2643, Jun. 2017.

[17] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. Adv. Neural Inf. Process. Syst.* Vancouver, BC, Canada: Hyatt Regency, Dec. 2009, pp. 1096–1104.

[18] Y. Bengio, "How auto-encoders could provide credit assignment in deep networks via target propagation," 2014, *arXiv:1407.7906*. [Online]. Available: http://arxiv.org/abs/1407.7906

[19] H. Nasir Khan, A. R. Shahid, B. Raza, A. H. Dar, and H. Alquhayz, "Multi-view feature fusion based four views model for mammogram classification using convolutional neural network," *IEEE Access*, vol. 7, pp. 165724–165733, Nov. 2019.

[20] S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion methods for CNN-based automatic modulation classification," *IEEE Access*, vol. 7, pp. 66496–66504, May 2019.

[21] Y. Zhao, G. Li, W. Xie, W. Jia, H. Min, and X. Liu, "GUN: Gradual upsampling network for single image super-resolution," *IEEE Access*, vol. 6, pp. 39363–39374, Jul. 2018.

[22] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," *IEEE Access*, vol. 7, pp. 31959–31970, Mar. 2019.

[23] J. Hu, H. Wang, S. Gao, M. Bao, T. Liu, Y. Wang, and J. Zhang, "S-UNet: A bridge-style U-net framework with a saliency mechanism for retinal vessel segmentation," *IEEE Access*, vol. 7, pp. 174167–174177, Sep. 2019.

[24] L. Qi, H. Zhang, W. Tan, S. Qi, L. Xu, Y. Yao, and W. Qian, "Cascaded conditional generative adversarial networks with multi-scale attention fusion for automated bi-ventricle segmentation in cardiac MRI," *IEEE Access*, vol. 7, pp. 172305–172320, Nov. 2019.

[25] X. Nie, M. Duan, H. Ding, B. Hu, and E. K. Wong, "Attention mask R-CNN for ship detection and segmentation from remote sensing images," *IEEE Access*, vol. 8, pp. 9325–9334, Jan. 2020.

[26] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 3929–3938.

[27] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[28] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "SAR image despeckling through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Fort Worth, TX, USA, Jul. 2017, pp. 5438–5441.

[29] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, Dec. 2017.

[30] F. Shi, N. Cai, Y. Gu, D. Hu, Y. Ma, Y. Chen, and X. Chen, "DeSpecNet: A CNN-based method for speckle reduction in retinal optical coherence tomography images," *Phys. Med. Biol.*, vol. 64, no. 17, Sep. 2019, Art. no. 175010.

[31] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7794–7803.

[32] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," 2019, *arXiv:1904.11492*. [Online]. Available: http://arxiv.org/abs/1904.11492

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, Oct. 2015, pp. 234–241.

[34] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. 14th Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 630–645.

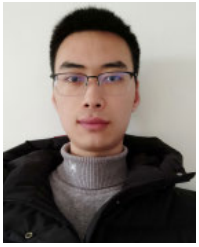
[37] X. Zhang, Y. Zou, and W. Shi, "Dilated convolution neural network with LeakyReLU for environmental sound classification," in *Proc. 22nd Int. Conf. Digit. Signal Process. (DSP)*, London, U.K., Aug. 2017, pp. 1–5.

[38] S. Shah, P. Ghosh, L. S. Davis, and T. Goldstein, "Stacked U-nets: A no-frills approach to natural image segmentation," 2018, *arXiv:1804.10343*. [Online]. Available: http://arxiv.org/abs/1804.10343

[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826.

[40] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016, *arXiv:1603.07285*. [Online]. Available: http://arxiv.org/abs/1603.07285

[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[42] M. Szkulmowski, I. Gorczynska, D. Szlag, M. Sylwestrzak, A. Kowalczyk, and M. Wojtkowski, "Efficient reduction of speckle noise in optical coherence tomography," *Opt. Express*, vol. 20, no. 2, pp. 1337–1359, 2012.
[43] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 205–229, Apr. 2009.
[44] Accessed: May 5, 2019. [Online]. Available: https://www.philips.com.cn/healthcare/product/HC795204V/-
[45] Accessed: May 5, 2019. [Online]. Available: https://www.philips.com.cn/healthcare/product/HC795200W/-
[46] Accessed: May 5, 2019. [Online]. Available: https://www.philips.com.cn/healthcare/product/HC795200V/-



YANCHENG LAN received the B.S. degree in biomedical engineering from the Huazhong University of Science and Technology, China, in 2018, where he is currently pursuing the master's degree in biomedical engineering. His research interests include ultrasound image denoising and deep learning.

XUMING ZHANG received the B.S. and M.S. degrees in material science and engineering from the Wuhan University of Technology, China, in 1998 and 2001, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, China, in 2005. From 2006 to 2008, he worked in the field of mechanical and electrical engineering with the Huazhong University of Science and Technology (HUST), as a Postdoctoral Researcher. From 2008 to 2009, he held a postdoctoral research position with the University of California at Davis, USA. Since 2009, he has been an Associate Professor with the Department of Biomedical Engineering, HUST. His research interests include image de-noising and image registration.