

An exploratory study on ultrasound image denoising using feature extraction and adversarial diffusion model

Yue Hu^{1,2} | Huiying Xu^{1,3} | Xinzhong Zhu^{1,3} | Xiao Huang⁴

¹School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China

²Intelligent Manufacturing College, Jinhua University of Vocational Technology, Jinhua, China

³Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China

⁴College of Education, Zhejiang Normal University, Jinhua, China

Correspondence

Huiying Xu, Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China.
Email: xhy@zjnu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 62376252; Key Project of Natural Science Foundation of Zhejiang Province, Grant/Award Number: LZ22F030003; Zhejiang Province Leading Geese Plan, Grant/Award Numbers: 2025C02025, 2025C01056

Abstract

Background: In ultrasound imaging, the generated images involve speckle noise owing to the mechanism underlying image generation. Speckle noise directly affects image analysis, necessitating its effective suppression.

Purpose: Ultrasound image denoising offers limited performance and causes structural information loss. To address these challenges and improve ultrasound image quality, we develop a new denoising method based on the diffusion model (DM).

Methods: This exploratory study proposes a DM-based denoising method, namely adversarial DM with feature extraction network (ADM-ExNet) to investigate the potential of combining diffusion models and generative adversarial Networks (GANs) for ultrasound image denoising. Specifically, we replace the reverse process of the DM with a GAN and modify the generator and discriminator as a U-Net structure. Simultaneously, a structural feature extraction network is incorporated into the model to construct a loss function, which offers enhanced detail retention. The noise levels ($\sigma = 10, 15, 20$) were simulated by adding Gaussian noise to the original ultrasound images, where σ controls the intensity of the noise. We employed three public datasets, HC18, CAMUS, and Ultrasound Nerve, which involve the ultrasound images of the fetal head circumference, heart, and nerves, respectively. Each image was adjusted to 256×256 pixels, and the training set and the validation set were divided by 9:1. The mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) were employed as primary evaluation metrics. To rigorously validate the statistical significance of performance differences, we further applied false discovery rate (FDR) correction for hypothesis testing and calculated Cohen's d effect sizes to quantify the magnitude of improvements against baselines. ADM-ExNet was compared with three traditional filtering methods and four deep learning methods with the U-Net structure.

Results: The proposed ADM-ExNet significantly enhances denoising performance across all datasets, with PSNR improvements exceeding 12 dB over noisy baselines and MSE reductions of over 90%. Notably, ADM-ExNet achieves high SSIM values (e.g., 0.941 at $\sigma = 10$ on HC18 vs. 0.369 for noisy images), demonstrating superior structural preservation without excessive smoothing. Statistical significance (FDR-adjusted $p < 0.01$) and Cohen's d effect sizes (up to $d = 3.8$ on CAMUS at $\sigma = 20$) confirm its robustness, outperforming traditional methods and deep learning competitors in both visual quality and quantitative metrics (PSNR, SSIM) across noise levels. This balance of detail retention and noise suppression highlights the exploratory potential of ADM-ExNet.

Conclusions: The proposed method improves the quality of ultrasound images with various structural features, effectively reducing noise while retaining details.

KEYWORDS

adversarial diffusion model, image denoising, ultrasound image

1 | INTRODUCTION

With the rapid development of medical big data and computer processing units, deep learning methods have been increasingly employed in medical image analysis. Ultrasound imaging has been commonly used owing to its affordability, noninvasiveness, and deep penetration ability. However, the limitations of the imaging environment and equipment performance lead to noise and artifacts, reducing image quality and causing loss of key information. In particular, speckle noise reduces the contrast of soft structural features in the image and blurs their boundary details, thereby reducing diagnostic accuracy. Extensive research has been conducted on speckle noise modeling and ultrasound image denoising. Traditional filtering methods such as image denoising, median filtering,¹ mean filtering,² and Gaussian filtering³ have been commonly used for noise reduction.

Various denoising methods have been developed using deep learning methods, particularly convolutional neural networks (CNNs) and generative adversarial networks (GANs) for noise suppression. GANs have been developed in 2014,⁴ and have resulted in transformational development. In general, a GAN employs adversarial thinking and includes a generator, which continuously generates data that more closely align with the actual labels as required, and a discriminator, which continuously distinguishes between the generator's results and the actual labels. However, original GANs struggled to generate images with specific attributes consistently and often suffered from mode collapse and training instability. To address these limitations, conditional GANs (cGANs) were developed to enforce control over generated image attributes.⁵ Its core concept entails integrating attribute information into both the generator and discriminator, where the attribute in question may be any label information. Despeckling CNNs (DsCNNs)⁶ use traditional speckle noise filtering to preprocess the ultrasound image and obtain the corresponding noise-free ultrasound image. Consequently, these data are utilized to facilitate the training of deep neural networks. To retain image details, DsCNNs use structural similarity measures as part of the loss function to guide network training. Further, researchers have employed a GAN to suppress ultrasound speckle noise.⁷ Their generator model adopted a fully convolutional U-Net architecture, which can directly transmit shallow details to the deeper layers of the network to better protect structural details.

These methods, however, have certain shortcomings in handling large-scale and multiplicative noise types.

Diffusion models (DMs)⁸ have been found to provide immense versatility across various fields,⁹ including medical image processing.¹⁰ DMs can be used for not only the conversion of different types of medical images,^{11,12} but also the reconstruction and generation of medical images,^{13–15} which can increase the medical image data available for analysis. In medical examination, DMs can be used for image segmentation^{16–18} and anomaly detection,^{19–21} which can assist doctors with image analysis. In addition, it is applied to medical image denoising. DenoOCT-DDPM²² first uses self-fusion as a preprocessing step, and then uses denoising diffusion probabilistic models (DDPMs)²³ to despeckle the optical coherence tomography retinal image via unsupervised learning. PET-DDPM²⁴ is a DDPM-based positron emission tomography denoising framework, which embeds auxiliary modes as prior information formulated by the DDPM. Compared with the denoising network based on U-Net,²⁵ this method enhances the image quality. In the context of ultrasound imaging, Stevens et al.²⁶ recently showed diffusion models' effectiveness for dehazing cardiac ultrasound images. Their approach uses a joint posterior sampling framework with separate score-based networks in the radio-frequency (RF) domain to model clean tissue and haze. It can effectively remove structured haze and retain weakly reflected tissue signals, highlighting diffusion models' potential in dealing with complex noise artifacts in medical imaging. Compared with the GAN, the DM offers improved quality and training convergence. However, the sampling requires the analysis of thousands of networks, making it costly and slow.

Two main problems persist regarding current ultrasound image denoising methods: First, the denoising effect is insufficient; second, the boundary information of structural features is easily lost after denoising. To solve these challenges, we propose an ultrasound image denoising model that combines the DM and GAN, namely, adversarial DM with feature extraction network (ADM-ExNet). The contributions of the study can be summarized as follows:

1. Considering the slow sampling speed of the DM, we replaced the reverse process of DM with a GAN structure. Simultaneously, to better retain detailed information, the U-Net structure of the GAN generator and discriminator was employed to enhance detail retention.
2. We proposed the loss function using two visual geometry group networks to extract the structural features of ultrasound images with considerably enhanced and consistent boundary details.

3. We conducted experiments on three datasets with different structural features (brain, knee, and nerve). The results show that our proposed method is highly robust and can be broadly employed in medical ultrasound imaging.

2 | MATERIALS AND METHODS

2.1 | Ultrasound image noise model

Ultrasound imaging is a broadly employed medical imaging procedure in cardiology, obstetrics, and gynecology. It can provide high-resolution images without the use of ionizing radiation and plays an important role in medical imaging applications. Ultrasound imaging equipment emits ultrasound waves. The interaction of ultrasound waves reflected and scattered when propagating through the human body results in certain granular black and white points on the ultrasound image, which constitute speckle noise.

The sound penetration and absorption of the echo signal are fully achieved through appropriate compensation of the ultrasound imaging system. The echo signal obtained by the final envelope detection comprises the beneficial signal and noise, and the noise can be divided into the multiplicative and additive noise.²⁷

$$f(i, j) = g(i, j)n(i, j) + w(i, j) \quad (1)$$

where $n(i, j)$ and $w(i, j)$ denote the multiplicative and additive noise, respectively. $(i, j) \in \mathbb{Z}^2$ represent the coordinates in two-dimensional (2D) space. $g(i, j)$ and $f(i, j)$ represent the original and observed signals, respectively. Denoising estimates the current real pixel value, $g(i, j)$, which is a function of the gray value $f(i, j)$ of the noise map and the local statistics of the pixel neighborhood.

The multiplicative noise, which is caused by random scattering in the resolution unit of the image, mainly forms the ultrasound noise. Compared to multiplicative noise, the degree of the additive noise is very slight and is made up of the sensor noise.²⁸ To simplify the model, the additive noise $w(i, j)$ is typically ignored. Equation (1) can be rewritten as

$$f(i, j) = g(i, j)n(i, j). \quad (2)$$

2.2 | Diffusion models

The DM represents a generative mode in deep learning and comprises the forward and reverse processes.²³

In the forward process, the incorporation of the Gaussian noise into the data occurs gradually ($x_0 \sim q(x_0)$), and the original data distribution $q(x_0)$ tends toward a

pure Gaussian distribution:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta} \cdot x_{t-1}, \beta_t \cdot \mathbf{I}),$$

$$q(x_1, x_2, \dots, x_T|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}), \quad (3)$$

where β_t is the noise variance modifier associated with the moment t , and \mathbf{I} is a unit matrix with the same dimensions as the initial state x_0 .

The reverse process through the noise estimation network at time t can be defined as:

$$p_\theta(x_1, x_2, \dots, x_T) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t), \quad (4)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)),$$

where $\mu_\theta(x_t, t)$ is the mean value of noise, $\Sigma_\theta(x_t, t)$ is the variance, and θ is the noise estimation network parameter.

Based on the existing noise state, this process obtains the state of the previous moment by learning to estimate the noise distribution, and it gradually constructs real data from the Gaussian distribution. Therefore, the optimization objective can be expressed as:

$$-\sum_{t=1}^T \mathbb{E}_{q(x_0)q(x_t|x_0)} D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t)), \quad (5)$$

which indirectly maximizes the evidence lower bound (ELBO) of $p_\theta(x_0)$, which is the likelihood, and D_{KL} denotes the Kullback–Leibler (KL) divergence.

Although DMs can generate high-quality samples and can be applied to various applications, they usually require thousands of iterations to achieve high-quality results. The inverse distribution can be assumed to be similar to a Gaussian distribution for these sampling steps. However, it becomes a non-Gaussian multimodal distribution with substantial noise. To address this problem, denoising diffusion GANs (DDGANs)²⁹ were proposed as adversarial learning schemes. The conditional distribution between $q(x_{t-1}|x_t)$ and $p_\theta(x_{t-1}|x_t)$ is matched using the cGAN, and large noise is randomly added between adjacent diffusion steps to achieve denoising with fewer steps. Its composition is summarized as follows:

$$\min_{\theta} \max_{D_{adv}} \sum_{t=1}^T \mathbb{E}_{q(x_t)} D_{adv}(q(x_{t-1}|x_t) \| p_\theta(x_{t-1}|x_t)). \quad (6)$$

The objective of the algorithm D_{adv} is to identify the discrepancy between the predicted denoised image and the sampled denoised image. The discriminator, which

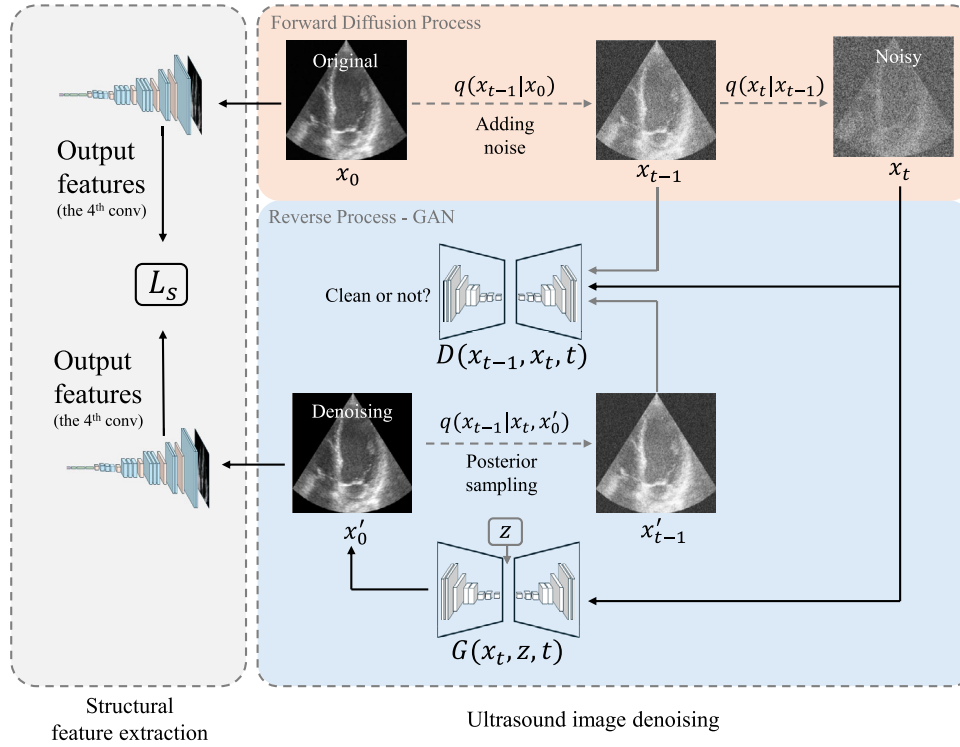


FIGURE 1 Framework of ADM-ExNet, which is mainly divided into structural feature extraction (left) and ultrasound image denoising (right) comprising the forward diffusion process (orange) and the reverse process combined with GAN (blue). GAN, generative adversarial network.

depends on time t , is denoted as $D_\phi(x_{t-1}, x_t, t)$. The loss function \mathcal{L}_{DG} can be written as

$$\begin{aligned}
 & - \sum_{t=1}^T \mathbb{E}_{q(x_0)q(x_{t-1}|x_0)q(x_t|x_{t-1})} [-\log(D_\phi(x_{t-1}, x_t, t))] \\
 & + \mathbb{E}_{p_\theta(x_{t-1}|x_t)} [-\log(1 - D_\phi(x_{t-1}, x_t, t))].
 \end{aligned} \quad (7)$$

2.3 | Proposed method

Based on DMs, this study proposes an ultrasound image denoising method combining adversarial learning and feature extraction networks, termed ADM-ExNet, with its overall framework illustrated in Figure 1. We adjusted our generator and discriminator to process the U-Net structure, similar to that²⁵ employed in NCSN++.³⁰ This U-Net structure has multiple residual network modules,³¹ also known as attTention modules.³² The generator achieves image reconstruction through progressive noise removal, while the discriminator captures local details and global semantics via multi-scale downsampling and upsampling paths, enhancing the guidance capability of adversarial training through pixel-wise distribution matching. Additionally, we added two network modules to extract structural features prior to ultrasound imaging to enhance detail retention.

We demonstrated that distinct convolutional layers can capture image features at varying semantic levels within CNNs. These features can be classified into four categories: low-level features (e.g., the target boundary), intermediate features (e.g., boundary combination), high-level features (e.g., the target part of the region), and complete object features. General features of the image, such as boundaries or textures, can be extracted in the low-level convolutional layer of the network. As the network depth increases, these features are gradually combined to form a representation of the target in the image. Simultaneously, when the boundary is extracted, the network is robust to noise. Therefore, a deep network can be employed to extract the structural feature from the ultrasound image in the low-level convolutional layer, as shown in Figure 2.

To obtain the prior objective function of the structural feature, we established that the ultrasound image cannot lose details after speckle noise suppression, and its details must be as consistent as possible with those that existed before processing. Therefore, we propose two visual geometry group networks (VGG-Nets)¹⁶ for feature extraction. The extracted features are the output features of the fourth layer, and this structure is called ExNet. One of the VGG-Nets was used to extract the features before speckle noise suppression, and the other was used to extract the features after speckle noise suppression. The network structure was used to

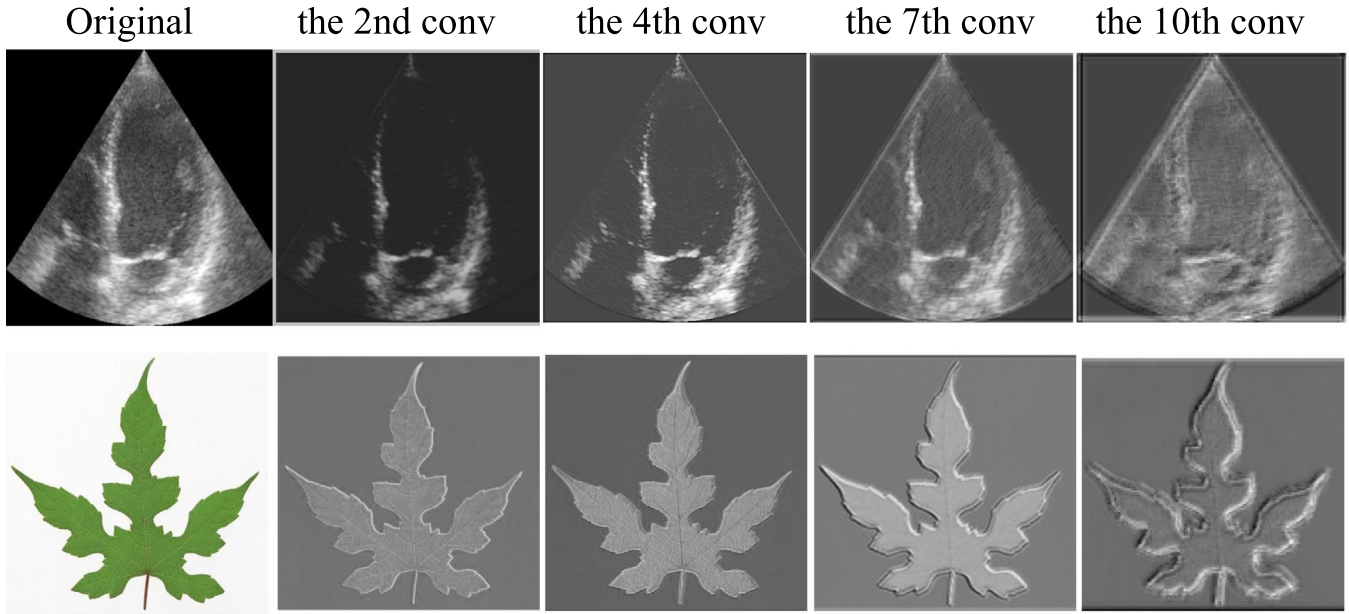


FIGURE 2 Visualization results of feature map.

obtain the feature information of the ultrasound image in the output of the fourth layer. Because the features obtained before and after speckle noise suppression must be as similar as possible, the loss function can be defined as follows:

$$\mathcal{L}_s = \|f_{vgg}(W; I_n) - f_{vgg}(W; f_d(I_n))\|_2, \quad (8)$$

where $f_{vgg}(\cdot)$ is the function representation of the network, W represents the network parameters, I_n represents the real ultrasound image containing speckle noise, and $f_d(I_n)$ represents the ultrasound image with speckle noise suppressed through the network.

The ultrasound image inputs are divided into two branches. The first is the forward diffusion process in the diffusion model. After the denoised image is directly generated by the generator, the features are extracted using ExNet. Another branch is the original image directly input into ExNet to extract the features. These two features constitute the loss function through the L2-norm. The loss function of the overall model is described as

$$\mathcal{L} = \mathcal{L}_{DG} + \omega \cdot \mathcal{L}_s. \quad (9)$$

The loss function comprises two parts, one is the generator network loss function; the other is the structural feature loss function. ω was experimentally determined as 0.001 and was used to balance the two parts of the objective function. The value was set to 0.001 based on multiple experiments. Two feature extraction networks were used as feature extractors, and their parameters were fixed.

2.4 | Implementation and evaluation

In the experiments, three datasets with different structural features were chosen. We used 90% of the data in each dataset for training and 10% for validation and adjusted image size to 256×256 pixels.

HC18: The fetal head circumference (HC) ultrasound dataset,³³ from the Automated Measurement of Fetal Head Circumference competition was used. The HC is measured during pregnancy and quantified in a specific cross-sectional area of the fetal head, designated as the standard plane. A total of 1334 2D ultrasound images of standard planes were included in the study. The initial ultrasound image had a size of 800×540 pixels.

CAMUS: The CAMUS dataset³⁴ comprises clinical examinations of 500 patients, primarily including 2D apical two-chamber and four-chamber images. Additionally, the dataset also provides time instances of the end-diastolic and end-systolic phases of the cardiac cycle and has three different image qualities (poor, good, and medium). All images were obtained using a GE Vivid E95 ultrasound scanner. In this study, only the emergency department four-cavity images of all the three image quality levels were selected. The size of the picture was not exactly the same, approximately 600×700 pixels. A total of 450 images in total with an adjusted size of 256×256 pixels were used.

Ultrasound Nerve: The dataset³⁵ comes from the Ultrasound Nerve Segmentation competition on the Kaggle site, involving a collection of nerves called brachial plexus. It contains 11 143 ultrasound images with a resolution of 580×420 pixels.

In these three datasets, the original ultrasound images are collected by means of physical correction. Clinically acceptable noise levels is involved in the ultrasound Images, which can be used as experimental data to ensure the feasibility of the model to a certain extent. Adding additive Gaussian noise to the original ultrasound image can partially simulate the noise in the ultrasound image, which can be used to control the noise level of the images in multiple sets of experiments and evaluate the robustness of the proposed method. The noise-containing image z can be expressed as

$$z = g + randn(size(g)) * \sigma / 100, \quad (10)$$

where σ is the noise variable, and $randn(size(g))$ generates data with the same dimensions as the original image g with a mean of 0 and a standard deviation of 1.

We used the mean square error (MSE), peak signal-to-noise ratio (PSNR),³⁶ and structural similarity index (SSIM)³⁷ as the evaluation indicators. A smaller MSE, value reflects a more enhanced image quality. On the contrary, the larger the PSNR value, the better the image quality. The SSIM is used to evaluate the similarity of two images. A value closer to 1, indicates an increase in the similarity of the two. They are calculated using the following equations:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \|F_d(i, j) - I(i, j)\|_2^2, \quad (11)$$

where H and W are the height and width of the ultrasound image, respectively, F_d is the image after noise removal, and I is the original image.

$$PSNR = 20 \times \log_{10} \left(\frac{255}{\sqrt{MSE}} \right), \quad (12)$$

$$SSIM(F_d, I) = \frac{(2\mu_{F_d}\mu_I + c_1)(2\sigma_{F_d}\sigma_I + c_2)}{(\mu_{F_d}^2 + \mu_I^2 + c_1)(\sigma_{F_d}^2 + \sigma_I^2 + c_2)}.$$

Here, μ_{F_d} and $\sigma_{F_d}^2$ are the mean and variance of F_d , μ_I and σ_I^2 are the mean and variance of I , and c_1 and c_2 are constants.

To rigorously evaluate the performance of ADM-ExNet against baseline methods, statistical analyses were conducted to quantify significance and effect sizes. Statistical significance across multiple comparisons was assessed using the Friedman test,³⁸ a non-parametric method suitable for comparing ranked results over three or more paired groups. Post-hoc analysis with Nemenyi correction was applied to control the false discovery rate.³⁹ Effect sizes were calculated using Cohen's d ,⁴⁰ defined as

TABLE 1 Hyperparameters for ADM-ExNet.

Resolution	256
Diffusion steps	4
EMA	0.999
Batch size	128
learning rate of generator	2e-4
learning rate of discriminator	1e-4

$$\frac{(\mu_1 - \mu_2)}{\sigma_{pooled}}, \quad (13)$$

where μ_1 and μ_2 represent the mean values of compared methods, and σ_{pooled} denotes the pooled standard deviation. Thresholds of 0.2, 0.5, and 0.8 were adopted to interpret small, medium, and large effects, respectively. And Table 1 lists the hyperparameters for ADM-ExNet.

To further confirm the effectiveness of our method, generalization ability evaluation will be conducted on **the Point-of-Care Ultrasound (POCUS) dataset**.⁴¹ The POCUS dataset was collected in emergency department settings for rapid assessment of COVID-19 pneumonia. This dataset comprises lung ultrasound scans from 267 RT-PCR-confirmed patients, acquired under clinically urgent conditions. Notably, the images contain significant noise artifacts (e.g., motion blur, sub-optimal probe contact) due to real-world emergency acquisition protocols. A total of 2189 keyframe images were included, with original resolutions of 800×600 pixels. All images were resized to 256×256 pixels for experimental consistency.

For generalization ability evaluation where ground truth is unavailable, three no-reference metrics were adopted: Natural Image Quality Evaluator (NIQE),⁴² Contrast-to-Noise Ratio (CNR),⁴³ and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE).⁴⁴ The NIQE quantifies naturalness degradation using spatial domain statistics, where lower values indicate better perceptual quality. The CNR is defined as:

$$CNR = \frac{\mu_{ROI} - \mu_{background}}{\sigma_{background}}, \quad (14)$$

and higher CNR demonstrates improved tissue contrast preservation. The BRISQUE predicts quality distortions via natural scene statistics, with lower scores reflecting fewer artifacts.

3 | RESULTS

3.1 | Quantitative analysis

We compared ADM-ExNet with seven other methods, namely, three traditional methods, and four deep learning

TABLE 2 Comparison of results on the HC18, CAMUS, and Ultrasound Nerve datasets.

Dataset	Model	$\sigma = 10$			$\sigma = 15$			$\sigma = 20$		
		MSE ↓	SSIM ↑	PSNR ↑	MSE ↓	SSIM ↑	PSNR ↑	MSE ↓	SSIM ↑	PSNR ↑
HC18	Noise	196.541	0.369	25.358	384.317	0.225	21.379	594.648	0.194	19.795
	NLM	42.328‡	0.438†	30.576‡	45.137‡	0.427†	29.375‡	52.794‡	0.457†	30.231‡
	Bitonic	33.347‡	0.883‡	34.977‡	37.619‡	0.836‡	32.751‡	43.903‡	0.813‡	31.837‡
	ASWMean	11.674‡	0.928‡	36.976‡	15.473‡	0.902‡	34.867‡	25.390‡	0.880‡	33.617‡
	RED	22.128‡	0.925‡	36.486‡	27.438‡	0.889‡	34.219‡	37.307‡	0.865‡	32.918‡
	RDN10	11.206‡	0.931‡	37.658‡	15.024‡	0.911‡	35.089‡	24.903‡	0.890‡	34.152‡
	RED-SENet	11.268‡	0.929‡	37.470‡	15.127‡	0.907‡	35.132‡	24.931‡	0.889‡	34.027‡
	GAN-RW	<u>11.187‡</u>	<u>0.934‡</u>	<u>37.755‡</u>	<u>14.937‡</u>	<u>0.912‡</u>	<u>35.157‡</u>	<u>24.597‡</u>	<u>0.893‡</u>	<u>34.224‡</u>
	ADM-ExNet	11.037‡###	0.941‡###	37.816‡###	14.792‡###	0.916‡###	35.671‡###	24.373‡###	0.902‡###	34.367‡#
CAMUS	Noise	217.107	0.499	18.113	441.218	0.368	15.394	627.891	0.324	12.348
	NLM	59.174‡	0.522†	24.813‡	71.970‡	0.497†	22.176‡	84.246‡	0.387†	20.097‡
	Bitonic	47.860‡	0.761‡	30.594‡	64.347‡	0.718‡	29.037‡	73.549‡	0.697‡	27.681‡
	ASWMean	25.539‡	0.924‡	35.061‡	40.134‡	0.908‡	34.092‡	50.346‡	0.792‡	31.982‡
	RED	30.641‡	0.939‡	33.415‡	50.671‡	0.902‡	32.540‡	61.546‡	0.873‡	30.637‡
	RDN10	25.012‡	<u>0.943‡</u>	36.306‡	39.910‡	0.912‡	34.255‡	50.171‡	0.889‡	32.090‡
	RED-SENet	25.098‡	0.942‡	36.165‡	39.795‡	0.911‡	34.244‡	50.087‡	0.891‡	32.394‡
	GAN-RW	<u>24.834‡</u>	0.945‡	<u>36.383‡</u>	<u>39.825‡</u>	<u>0.920‡</u>	<u>34.276‡</u>	<u>49.698‡</u>	<u>0.904‡</u>	<u>32.421‡</u>
	ADM-ExNet	24.813‡#	0.942‡#	36.962‡###	39.873‡#	0.931‡#	34.567‡###	49.348‡#	0.912‡#	32.519‡#
Ultrasound Nerve	Noise	210.482	0.453	24.902	445.222	0.267	20.375	665.259	0.257	18.679
	NLM	39.984‡	0.689†	29.573‡	46.339‡	0.593‡	26.681‡	62.076‡	0.438†	27.572‡
	Bitonic	27.312‡	0.807‡	32.167‡	35.907‡	0.671‡	30.381‡	43.044‡	0.753‡	30.867‡
	ASWMean	16.988‡	0.901‡	35.880‡	25.792‡	0.816‡	32.816‡	32.112‡	0.867‡	32.013‡
	RED	23.576‡	0.893‡	35.608‡	31.109‡	0.784‡	31.972‡	37.070‡	0.862‡	31.915‡
	RDN10	16.630‡	0.913‡	35.967‡	25.498‡	0.884‡	33.944‡	31.732‡	0.870‡	32.983‡
	RED-SENet	16.673‡	0.905‡	35.889‡	25.509‡	0.879‡	33.953‡	31.683‡	0.871‡	33.007‡
	GAN-RW	16.435‡	<u>0.914‡</u>	36.001‡	<u>25.412‡</u>	<u>0.888‡</u>	<u>34.067‡</u>	31.598‡	<u>0.873‡</u>	<u>33.160‡</u>
	ADM-ExNet	<u>16.567‡</u>	0.916‡#	<u>35.994‡</u>	25.401‡#	0.890‡#	34.119‡#	<u>31.616‡</u>	0.881‡#	33.167‡#

Note: Bold values indicate best performance; underlined values indicate second-best. Significance markers: † ($p < 0.05$ vs. Noise); ‡ ($p < 0.01$ vs. Noise); # ($p < 0.05$ vs. GAN-RW); ## ($p < 0.01$ vs. GAN-RW).

Abbreviations: MSE, mean square error; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

methods. Nonlocal means (NLM)⁴⁵ is a classical image denoising method. Bitonic⁴⁶ involves linear filtering in a morphological framework, and ASWMean⁴⁷ is an adaptive switching weight mean filter designed for the denoising of images affected by salt and pepper noise. The RED⁴⁸ approach to image restoration employs a convolutional encoder-decoder network with symmetrical skip connections, to facilitate information transfer between layers. RDN10⁴⁹ is a recently proposed simplified version of the RDN model for super-resolution and denoising of images. RED-SENet⁵⁰ represents a channel-adaptive denoising model based on a residual codec and employs a squeeze-and-excitation network to achieve its denoising capabilities. Further, GAN-RW⁵¹ is a GAN comprising residual dense connectivity and a weighted joint loss. The denoising network is based on the U-Net architectural

framework, comprising four encoder and four decoder modules.

As shown in Table 2, which summarizes averaged experimental results across three ultrasound image datasets, deep learning-based denoising methods significantly outperform traditional approaches. Table 3 provides the Cohen's d effect sizes quantifying the magnitude of performance improvements (PSNR) against the noise baseline, with thresholds for small ($d \geq 0.2$), medium ($d \geq 0.5$), and large ($d \geq 0.8$) effects. Among these, the proposed ADM-ExNet achieves superior performance across most metrics and datasets, with statistical significance (FDR-adjusted $p < 0.01$) in critical comparisons. Specifically, ADM-ExNet demonstrates consistent superiority, attaining the highest PSNR and SSIM values on HC18 (37.816 dB, 0.941 at $\sigma = 10$) and Ultrasound Nerve (35.994 dB PSNR, 0.916 SSIM at

TABLE 3 Cohen's d effect sizes for all models on PSNR (vs. Noise Image).

Dataset	Model	$\sigma = 10$			$\sigma = 15$			$\sigma = 20$		
		Δ PSNR	Cohen's d	Effect size	Δ PSNR	Cohen's d	Effect Size	Δ PSNR	Cohen's d	Effect Size
HC18	NLM	5.2	1.0	Large	8.0	1.6	Large	10.4	2.1	Large
	Bitonic	9.6	1.9	Large	11.4	2.3	Large	12.0	2.4	Large
	ASWMean	11.6	2.3	Large	13.5	2.7	Large	13.8	2.8	Large
	RED	11.1	2.2	Large	12.8	2.6	Large	13.1	2.6	Large
	RDN10	12.3	2.5	Large	13.7	2.7	Large	14.4	2.9	Large
	RED-SENet	12.1	2.4	Large	13.8	2.8	Large	14.4	2.9	Large
	GAN-RW	12.4	2.5	Large	13.8	2.8	Large	14.4	2.9	Large
	ADM-ExNet	12.5	2.5	Large	14.3	2.9	Large	14.6	2.9	Large
CAMUS	NLM	6.7	1.3	Large	8.0	1.6	Large	9.5	1.9	Large
	Bitonic	12.5	2.5	Large	14.0	2.8	Large	15.5	3.1	Large
	ASWMean	16.9	3.4	Large	16.9	3.4	Large	17.2	3.4	Large
	RED	15.3	3.1	Large	15.3	3.1	Large	15.8	3.2	Large
	RDN10	18.2	3.6	Large	18.2	3.6	Large	18.5	3.7	Large
	RED-SENet	18.1	3.6	Large	18.1	3.6	Large	18.4	3.7	Large
	GAN-RW	18.3	3.7	Large	18.3	3.7	Large	18.6	3.7	Large
	ADM-ExNet	18.8	3.8	Large	18.8	3.8	Large	18.9	3.8	Large
Ultrasound nerve	NLM	4.7	0.9	Large	6.3	1.3	Large	7.9	1.6	Large
	Bitonic	7.3	1.5	Large	9.0	1.8	Large	11.1	2.2	Large
	ASWMean	11.0	2.2	Large	12.0	2.4	Large	12.2	2.4	Large
	RED	10.7	2.1	Large	11.6	2.3	Large	12.1	2.4	Large
	RDN10	11.2	2.2	Large	12.5	2.5	Large	13.2	2.6	Large
	RED-SENet	11.0	2.2	Large	12.4	2.5	Large	13.2	2.6	Large
	GAN-RW	11.1	2.2	Large	12.5	2.5	Large	13.3	2.7	Large
	ADM-ExNet	11.1	2.2	Large	12.5	2.5	Large	13.4	2.7	Large

Note: Small ($d \geq 0.2$), medium ($d \geq 0.5$), large ($d \geq 0.8$). Abbreviation: PSNR, peak signal-to-noise ratio.

$\sigma = 10$), while maintaining competitiveness on CAMUS (36.962 dB PSNR at $\sigma = 10$, 0.931 SSIM at $\sigma = 15$). Its robustness against high noise is evident at $\sigma = 20$, where it surpasses baselines by substantial margins (e.g., HC18: +14.6 dB PSNR over noise; CAMUS: Δ PSNR = 18.9 dB), supported by Cohen's d values up to $d = 3.8$ (CAMUS, $\sigma = 20$), far exceeding the "Large" threshold. Notably, ADM-ExNet achieves the largest effect sizes in all datasets (e.g., $d = 2.9$ for HC18 and $d = 2.7$ for Ultrasound Nerve at $\sigma = 20$), reinforcing its practical impact. Combined with superior SSIM scores, these results highlight ADM-ExNet's ability to balance noise reduction with structural preservation, solidifying its versatility for clinical ultrasound denoising across diverse noise levels and anatomical complexities.

3.2 | Visual quality improvement

Figure 3 shows the noise removal results of different types of ultrasound images (enlarged images, which display more details, are provided). All methods can reduce

the influence of noise to a certain extent. Although NLM has been proposed in 2013, it can obtain effective regional smoothing owing to the similarity of nonlocal regions but relatively vague feature boundaries. Bitonic and ASWMean preserve more boundary information than NLM. The four deep learning methods remove noise, with no noticeable differences. However, the loss of boundary details appears to be more substantial for GAN-RW. In terms of visual effects, ASWMean, RDN10, and ADM-ExNet all provide detailed feature boundaries.

3.3 | Ablation studies

Ablation experiments were conducted to evaluate the addition of the feature extraction network (ExNet) and the use of the generator and discriminator with the U-Net structure. Inspired by DDGANs,²⁹ we proposed our time-dependent G and D convolutional networks with residual network blocks. D determines whether X_{t-1} is true or false under X_t and t conditions. Consequently, a minibatch standard deviation is added to each

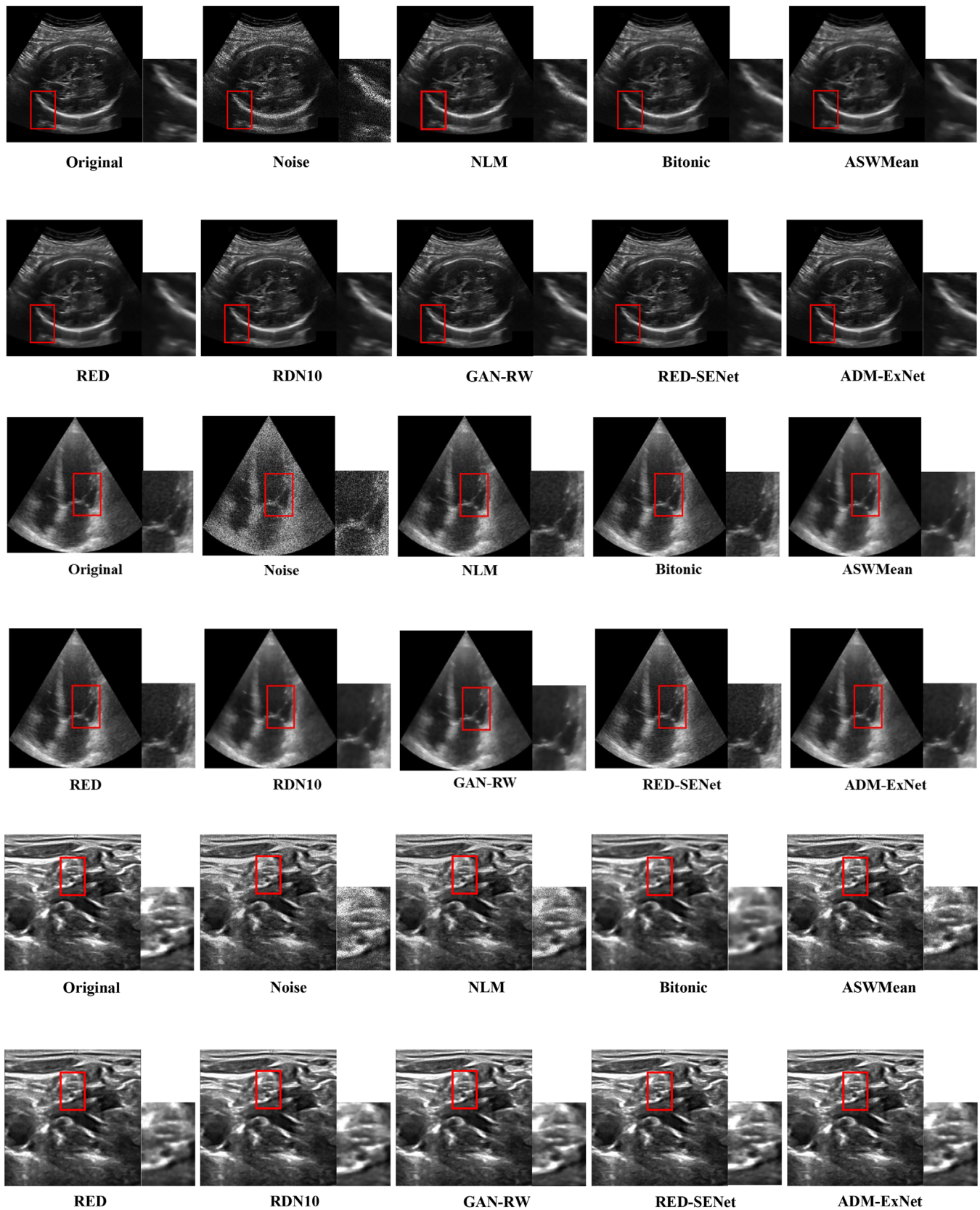


FIGURE 3 Results obtained using various methods on the HC18, CAMUS and Ultrasound Nerve datasets. To better show noise reduction and boundary information retention effects, specific areas (enclosed with red rectangles) were enlarged and placed on the right of the images

TABLE 4 Network structure for G and D , with the number of channels in each of the remaining network blocks provided on the right.

1×1 conv2d, 128
Residual Block, 128
Residual Block down, 256
Residual Block down, 512
Residual Block down, 512
Residual Block down, 512
minibatch standard layer
Global Sum Pooling
FC layer \rightarrow scalar

TABLE 5 Results of ablation studies obtained on the CAMUS dataset.

	MSE \downarrow	PSNR \uparrow	SSIM \uparrow
Noise	217.107	18.113	0.499
ADM-ExNet	24.813\ddagger	36.962\ddagger	0.942\ddagger
w/o ExNet	27.244 \dagger	36.672 \ddagger	0.935 \ddagger
Non-UNet G w/o ExNet	27.516 \dagger	34.937 \dagger	0.924 \dagger
Non-UNet D w/o ExNet	27.602 \dagger	34.721 \dagger	0.917 \dagger
Non-UNet G&D w/o ExNet	28.769 \dagger	33.855 \dagger	0.908 \dagger
Non-UNet G w/ ExNet	26.982 \ddagger	35.092 \ddagger	0.937 \ddagger
Non-UNet D w/ ExNet	26.543 \ddagger	36.421 \ddagger	0.939 \ddagger
Non-UNet G&D w/ ExNet	27.060 \dagger	34.232 \dagger	0.911 \dagger

Note: The values in bold indicate the highest level of statistical significance. $\dagger = p < 0.05$ versus Noise; $\ddagger = p < 0.01$ versus Noise.

residual network block. The specific structure is presented in Table 4. As shown in Table 5, we obtained the best indicator values by adding the ExNet and modifying the network structure.

Modifying either the generator or the discriminator to possess a U-Net structure increases the evaluation indicators (compared with the case of the unmodified network structure). The U-shaped network architecture offers upsampling and downsampling, as well as the operation of the feature concatenate. This can better process image-related information, especially when the discriminator has a U-Net structure. Conventional discriminators can output only global binary logic (true or false). The discriminator of the U-Net structure can distinguish detail information from different levels. When the generator and discriminator are constructed using the U-Net structure but ExNet is not added, the three evaluation indicators slightly decline, indicating the necessity of the feature extraction network.

Figure 4 shows the visualization results of our ablation experiments. The included subimages show the post-experimental magnification effect for the images obtained from the same region of the same original image. The combination of the GAN and DM provides improved noise reduction and boundary information

retention. Although the denoised image without ExNet is relatively smoother, it indicates the loss of a few structural features.

3.4 | Generalization ability evaluation

To address the generalization capability of ADM-ExNet beyond synthetic noise conditions, additional evaluation was conducted on the POCUS Lung COVID-19 dataset. These clinical lung images were acquired without pre-processing and exhibit significant noise artifacts due to real-world emergency acquisition protocols. These raw images were directly fed into the pre-trained ADM-ExNet without retraining. Using the no-reference metrics defined in 2.4, Table 6 showed significant improvements: NIQE decreased by 31.6% (from 5.73 ± 0.82 to 3.92 ± 0.71 , $p < 0.01$), CNR increased by 60.5% (from 1.85 ± 0.34 to 2.97 ± 0.41 , $p < 0.01$), and BRISQUE decreased by 30.5% (from 38.21 ± 6.17 to 26.54 ± 5.23 , $p < 0.01$). As shown in Figure 5, the qualitative results indicate that in lung ultrasound, effective speckle suppression can be achieved around the pleural line, while the key A-line texture is retained, without introducing artificial texture. These findings confirm ADM-ExNet's capability to generalize to real-world noise distributions while preserving diagnostically critical features.

4 | DISCUSSION

Medical ultrasound images contain a large amount of image information. Low image contrast increases the difficulty of distinguishing the lesion site from the background thereby increasing the risk of inaccurate diagnosis. We combined and optimized the DM and GAN to process three types of ultrasound images with noise and enhanced noise reduction. As shown in Table 2, our proposed method, ADM-ExNet, achieves optimal or second-best performance on the HC18 and Ultrasound Nerve datasets across all noise levels ($\sigma = 10, 15, 20$). For example, on the HC18 dataset at $\sigma = 10$, ADM-ExNet achieves a PSNR of 37.816 dB and an SSIM of 0.941, outperforming traditional methods (e.g., NLM, Bitonic) and other deep learning approaches (e.g., GAN-RW, RED-SENet). However, on the CAMUS dataset, it slightly trails RED-SENet in SSIM at $\sigma = 10$ (0.942 vs. 0.945) under low noise conditions, suggesting potential refinement for subtle cardiac structural details. Statistical significance and Cohen's d effect sizes (e.g., $d = 3.8$ at $\sigma = 20$ on CAMUS) confirm its overall superiority, with performance gaps narrowing as noise increases (e.g., CAMUS $\sigma = 20$: ADM-ExNet PSNR = 32.519 dB vs. GAN-RW = 32.421 dB). This underscores ADM-ExNet's robustness in high-noise scenarios while highlighting targeted opportunities for optimization in low-noise cardiac imaging.

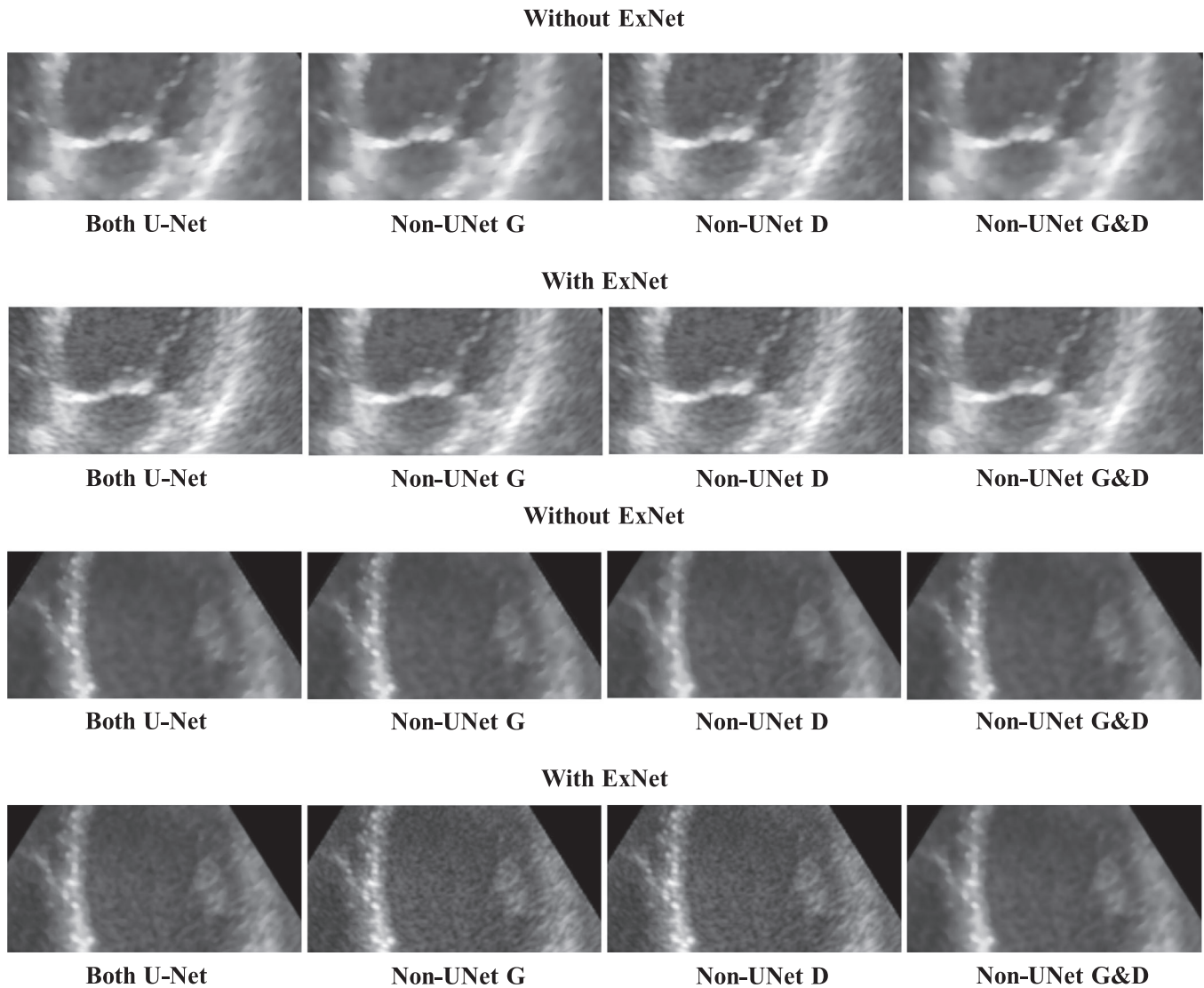


FIGURE 4 Visualization results of ablation experiments in terms of noise reduction and boundary information retention.

TABLE 6 Results of generalization ability evaluation obtained on the POCUS dataset.

Metric	Original images	ADM-ExNet output	Improvement
NIQE ↓	5.73 ± 0.82	3.92 ± 0.71	31.6% ($p < 0.01$)
CNR ↑	1.85 ± 0.34	2.97 ± 0.41	60.5% ($p < 0.01$)
BRISQUE ↓	38.21 ± 6.17	26.54 ± 5.23	30.5% ($p < 0.01$)

Note: NIQE, CNR and BRISQUE are presented as mean \pm std.

Abbreviations: BRISQUE, blind/referenceless image spatial quality evaluator; CNR, contrast to-noise ratio; NIQE, natural image quality evaluator.

Our study has two main limitations. First, we focused on static, rather than dynamic, ultrasound images, limiting the applicability of our method to dynamic imaging scenarios, such as echocardiography. Dynamic imaging provides critical information about heart function, particularly for assessing left ventricular torsion—a sensitive indicator of myocardial fiber contraction that is more informative than ejection fraction. Although our

method shows promise for static images, extending it to dynamic ultrasound videos remains a challenge, as current diffusion models are primarily designed for static data. Second, we assumed that speckle noise is uniformly distributed in ultrasound images, which may not fully capture the complex interactions between the ultrasound probe's sound field and the received signals. In future work, we plan to integrate linear sound field calculations into our model to better simulate real-world noise conditions and further enhance denoising performance.

It is important to explicitly acknowledge the exploratory nature of this study. Although our approach demonstrates technical advancements in noise reduction, its clinical applicability remains constrained by the use of synthetic noise and a moderate sample size. The simulated noise conditions, though useful for controlled experimentation, may not fully replicate the variability and complexity of clinical environments. To translate these methodological innovations into clinical

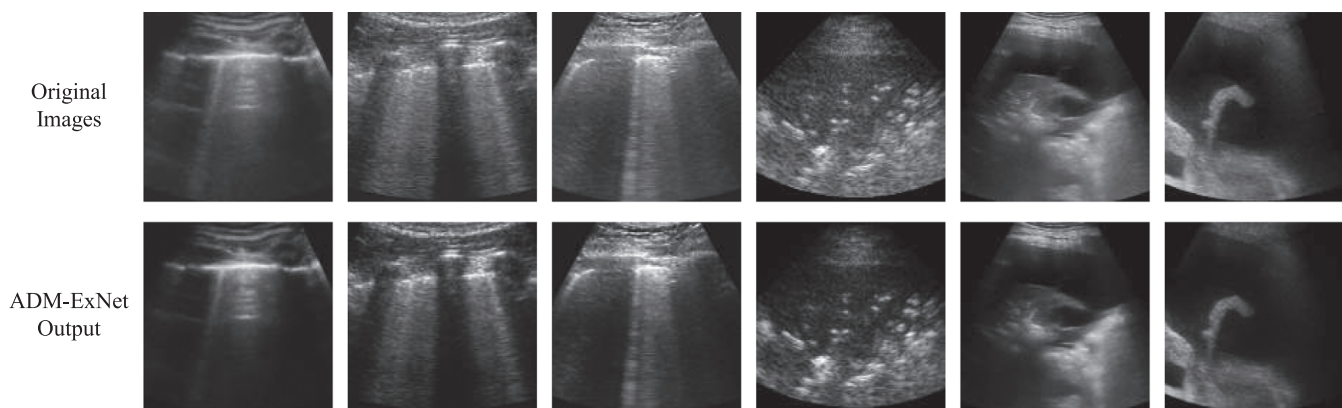


FIGURE 5 Visualization results of generalization ability evaluation.

practice, future confirmatory studies must prioritize real-world validation. This includes utilizing diverse clinical datasets with naturally occurring noise patterns, incorporating evaluations by radiologists to assess the diagnostic relevance of denoised images, and adopting pre-registered statistical plans to ensure methodological rigor and reproducibility.

5 | CONCLUSIONS

In this study, we developed an ultrasound image denoising method based on the DM and GAN. First, the DM and GAN were combined to improve the denoising effect, and the generator and discriminator were modified to possess the U-Net structure. Subsequently, a network module for feature extraction was added to avoid information loss after denoising. The superiority of our method in ultrasound image denoising was experimentally demonstrated vis-à-vis state-of-the-art deep learning-based denoising methods. The image denoising capability was considerably improved to retain more detailed feature information.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (62376252); Key Project of Natural Science Foundation of Zhejiang Province (LZ22F030003); Zhejiang Province Leading Geese Plan (2025C02025, 2025C01056).

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

REFERENCES

- Erkan U, Enginoğlu S, Thanh DNH, Hieu LM. Adaptive frequency median filter for the salt and pepper denoising problem. *IET Image Process.* 2020;14:1291-1302.
- Feng L, Wang J. Research on image denoising algorithm based on improved wavelet threshold and non-local mean filtering. *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP).* 2021:493-497.
- Jia M, Dong M. Analysis and comparison of Gaussian noise denoising algorithms. *J Phys: Conf Ser.* 2021;1846:012069.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. In: *Neural Information Processing Systems.* 2014:2672-2680.
- Mirza M, Osindero S. Conditional Generative Adversarial Nets. *ArXiv abs/1411.1784.* 2014.
- Mishra D, Tyagi S, Chaudhury S, Sarkar M, Soan AS. Despeckling CNN with Ensembles of Classical Outputs. *2018 24th International Conference on Pattern Recognition (ICPR).* 2018:3802-3807.
- Dietrichson F, Smistad E, Ostvik A, Løvstakken L. Ultrasound Speckle Reduction Using Generative Adversarial Networks. *2018 IEEE International Ultrasonics Symposium (IUS).* 2018:1-4.
- Sohl-Dickstein J, Weiss EA, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics. In: *Proceedings of the 32nd International Conference on Machine Learning - Volume 37 (ICML'15).* 2015:2256-2265. JMLR.org.
- Croitoru F-A, Hondru V, Ionescu RT, Shah M. Diffusion Models in Vision: A Survey. *IEEE Trans Pattern Anal Mach Intell.* 2022;45:10850-10869.
- Kazerouni A, Aghdam EK, Heidari M, et al. Diffusion models in medical imaging: a comprehensive survey. *Med Image Anal.* 2022;88:102846.
- Lyu Q, Wang G. Conversion between CT and MRI images using diffusion and score-matching models. *arXiv.* 2022;abs/2209.12104.
- Meng X, Gu Y, Pan Y, et al. A novel unified conditional score-based generative framework for multi-modal medical image completion. *arXiv.* 2022;abs/2207.03430.
- Song Y, Shen L, Xing L, Ermon S. Solving inverse problems in medical imaging with score-based generative models. In: *Proceedings of the International Conference on Learning Representations (ICLR);* 2022.
- Xie Y, Li Q. Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* 2022:655-664.
- Kim B, Ye JC. Diffusion deformable model for 4D temporal medical image generation. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2022. 25th International Conference;* September 18-22, 2022; Singapore. Part I. Berlin, Heidelberg: Springer-Verlag; 2022:539-548, doi: [10.1007/978-3-031-16431-6_51](https://doi.org/10.1007/978-3-031-16431-6_51)
- Fernandez V, Pinaya WHL, Borges P, et al. Can segmentation models be trained with fully synthetically generated data? In: Zhao C, Svoboda D, Wolterink JM, Escobar M, eds. *Simulation and Synthesis in Medical Imaging. 7th International Workshop,*

- SASHIMI 2022, Held in Conjunction with MICCAI 2022. vol. 13570. Lecture Notes in Computer Science. Springer; 2022:79-90, doi: [10.1007/978-3-031-16980-9_8](https://doi.org/10.1007/978-3-031-16980-9_8)
17. Kim B, Oh Y, Ye JC. Diffusion adversarial representation learning for self-supervised vessel segmentation. In: *Proceedings of the International Conference on Learning Representations (ICLR) 2023*; 2022.
 18. Wolleb J, Sandkühler R, Bieder F, Valmaggia P, Cattin PC. Diffusion models for implicit image segmentation ensembles. In: *International Conference on Medical Imaging with Deep Learning*. 2021:1336-1348.
 19. Wyatt J, Leach A, Schmon SM, Willcocks CG. AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Workshops (CVPRW)*; 2022:649-655.
 20. Sanchez P, Kascenas A, Liu X, O'Neil AQ, Tsaftaris SA. What is healthy? Generative counterfactual diffusion for lesion localization. In: *Deep Generative Models: Second MICCAI Workshop, DGM4MICCAI 2022, Held in Conjunction with MICCAI 2022*; September 22, 2022; Singapore, Berlin, Heidelberg: Springer-Verlag; 2022:34-44, doi: [10.1007/978-3-031-18576-2_4](https://doi.org/10.1007/978-3-031-18576-2_4)
 21. Wolleb J, Sandkühler R, Bieder F, Cattin PC. The Swiss army knife for image-to-image translation: multi-task diffusion models. *arXiv*. 2022;abs/2204.02641.
 22. Hu D, Tao YK, Oguz I. Unsupervised denoising of retinal OCT with diffusion probabilistic model. In: *Medical Imaging*. 2022:4.
 23. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS)*; 2020. Article 5746840-6851.
 24. Gong K, Johnson KA, Fakhri GE, Li Q, Pan T. PET image denoising based on denoising diffusion probabilistic model. *Eur J Nucl Med Mol Imag*. 2022;51:358-368.
 25. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Vol 9351. Lecture Notes in Computer Science*. Springer; 2015:234-241, doi: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)
 26. Stevens TSW, Meral FC, Yu J, et al. Dehazing Ultrasound Using Diffusion Models. *IEEE Trans Med Imaging*. 2024;43:3546-3558.
 27. Bhuiyan MIH, Swamy MNS, Ahmad M. Spatially adaptive thresholding in wavelet domain for despeckling of ultrasound images. *IET Image Proc*. 2009;3:147-162.
 28. Amirmazlaghani M, Amindavar H. Wavelet domain Bayesian processor for speckle removal in medical ultrasound images. *IET Image Proc*. 2012;6:580-588.
 29. Xiao Z, Kreis K, Vahdat A. Tackling the generative learning trilemma with denoising diffusion GANs. In: *Proceedings of the International Conference on Learning Representations (ICLR) 2022*; 2021.
 30. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. In: *Proceedings of the International Conference on Learning Representations*; 2021.
 31. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015:770-778.
 32. Vaswani A, Shazeer NM, Parmar N, et al. Attention is all you need. In: *Neural Information Processing Systems*. 2017;6000-6010.
 33. van den Heuvel TLA, de Bruijn D, de Korte CL, Ginneken Bv. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS ONE* 2018;13(8):e0200412, doi: [10.1371/journal.pone.0200412](https://doi.org/10.1371/journal.pone.0200412)
 34. Leclerc S, Smistad E, Pedrosa J, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging*. 2019;38:2198-2210.
 35. Montoya A, Hasnin, kaggle446, shirzad, Will Cukierski, and yffud. Ultrasound nerve segmentation. Kaggle. <https://www.kaggle.com/competitions/ultrasound-nerve-segmentation>. Published 2016. Accessed September 18, 2025.
 36. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
 37. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
 38. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc*. 1937;32(200):675-701.
 39. Nemenyi PB. *Distribution - Free Multiple Comparisons*. Ph.D. thesis. Princeton University; 1963.
 40. Cohen J. *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Lawrence Erlbaum Associates; 1988.
 41. Born J, Brändle G, Cossio M, et al. POCUS: point-of-care ultrasound dataset. *Sci Data*. 2022;9:655.
 42. Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. *IEEE Trans Image Process*. 2012;21:4695-4708.
 43. Zhou Z, Wu W, Wu S, et al. Contrast-to-noise ratio evaluation in focused ultrasound imaging. *Ultrasonics* 2014;54:116-122.
 44. Mittal A, Soundararajan R, Bovik AC. Making a "completely blind" image quality analyzer. *IEEE Signal Process Lett*. 2013;20:209-212.
 45. Kumar BKS. Image denoising based on non-local means filter and its method noise thresholding. *Signal Image Video Process*. 2013;7:1211-1227.
 46. Treece G. The Bitonic Filter: Linear Filtering in an Edge-Preserving Morphological Framework. *IEEE Trans Image Process*. 2016;25:5199-5211.
 47. Thanh DNH, Hien NN, Kalavathi P, Prasath VBS. Adaptive switching weight mean filter for salt and pepper image denoising. *Procedia Comput Sci*. 2020;171:292-301.
 48. Couturier R, Perrot G, Salomon M. Image denoising using a deep encoder-decoder network with skip connections. In: *International Conference on Neural Information Processing*. 2018;554-565.
 49. Zhang Y, Tian Y, Kong Y, Zhong B, Fu YR. Residual dense network for image restoration. *IEEE Trans Pattern Anal Mach Intell*. 2020;43:2480-2495.
 50. Xianhua Z, Yancheng L, Ge G, Xueting Z. Channel adaptive ultrasound image denoising method based on residual encoder-decoder networks. *J Electron Inf Technol*. 2022;44:2547-2558.
 51. Zhang J, Zhang J. Ultrasound image denoising using generative adversarial networks with residual dense connectivity and weighted joint loss. *Peer J Comput Sci*. 2022;8:e873.

How to cite this article: Hu Y, Xu H, Zhu X, Huang X. An exploratory study on ultrasound image denoising using feature extraction and adversarial diffusion model. *Med Phys*. 2025;52:e70023. <https://doi.org/10.1002/mp.70023>