DEMIX: Dual-Encoder Latent Masking Framework for Mixed Noise Reduction in Ultrasound Imaging

Anonymous authors
Paper under double-blind review

Abstract

Ultrasound imaging is widely used in noninvasive medical diagnostics due to its efficiency, portability, and avoidance of ionizing radiation. However, its utility is limited by the signal quality. Signal-dependent speckle noise, signal-independent sensor noise, and non-uniform spatial blurring caused by the transducer and modeled by the point spread function (PSF) degrade the image quality. These degradations challenge conventional image restoration methods, which assume simplified noise models, and highlight the need for specialized algorithms capable of effectively reducing the degradations while preserving fine structural details. We propose DEMIX, a novel dual-encoder denoising framework with a masked gated fusion mechanism, for denoising ultrasound images degraded by mixed noise and further degraded by PSF-induced distortions. DEMIX is inspired by diffusion models and is characterized by a forward process and a deterministic reverse process. DEMIX adaptively assesses the different noise components, disentangles them in the latent space, and suppresses these components while compensating for PSF degradations. Extensive experiments on two ultrasound datasets, along with a downstream segmentation task, demonstrate that DEMIX consistently outperforms state-of-the-art baselines, achieving superior noise suppression and preserving structural details. The code will be made available.

1 Introduction

Digital images are often degraded by noise arising from acquisition and transmission (Liu et al., 2017; Huang et al., 2017; Goyal et al., 2018). Noise can be attributed to various factors, including random thermal fluctuations of imaging sensors, inherent sensor imperfections, and interactions with the imaging environment, among others. The noisy images thus acquired have random intensity variations and degraded characteristics, which make the performance of automated downstream tasks such as segmentation and classification significantly challenging. Mathematically, image denoising can be defined as an inverse problem that aims to recover the true image from a noisy observation while preserving its structural details. Image denoising has been extensively studied in the past few decades (Buades et al., 2005; Kervrann & Boulanger, 2006; Ashouri & Eslahchi, 2022; Xie et al., 2024). Noise can be signal-dependent or signal-independent,

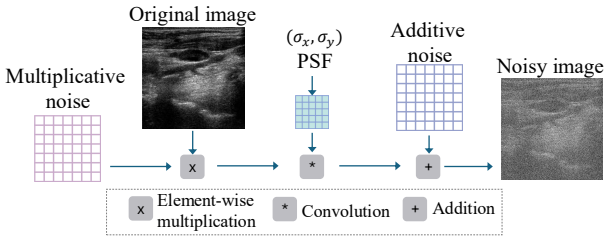

Figure 1: Noisy image generation: The original image is first multiplied with signal-dependent speckle, convolved with the PSF of the imaging system and finally added to signal-independent sensor noise.

depending on its source and the characteristics of the imaging device. The dominance of signal-independent

noise in natural illumination conditions motivates the standard additive Gaussian noise model for image denoising applications (Luisier et al., 2010). Although the simplified assumption leads to simpler analysis and algorithmic design, the assumption does not apply to many image sources. Images acquired from a wide class of imaging modalities are often degraded by various forms of noise, making the image denoising problem particularly challenging. To name a few, images acquired from ultrasound applications, fluorescence microscopy, and astronomy are affected by a mixture of different kinds of noise, which deviate significantly from the additive Gaussian noise assumption. The complex noise patterns arising in these images require specially designed denoising algorithms that go beyond simplistic models. Additionally, these images are further distorted by the point spread function (PSF) of the imaging system. The PSF characterizes the system's response to a point source and encapsulates effects such as diffraction, optical aberrations, and sensor imperfections. The final acquired image is a spatially distorted representation of the true underlying structure. Accurate PSF modeling has been a key research focus for image reconstruction and analysis, specifically from astronomical image analysis (Liaudat et al., 2023; Beltramo-Martin et al., 2020; Jia et al., 2020) to microscopy image analysis (Li et al., 2018; Stallinga et al., 2009; Shechtman et al., 2014; Zhang et al., 2007). In ultrasound images, the PSF has been widely utilized in a range of applications, including motion estimation in tissues (Saris et al., 2018), analytical methods for PSF estimation, deconvolution, and restoration algorithms (Dalitz et al., 2015; Yang & Ren, 2010; Zhao et al., 2016), PSF-informed super-resolution algorithms (Ploquin et al., 2015) and simulation frameworks with spatially varying PSFs (Varray et al., 2016; Schretter et al., 2018).

Ultrasound is a highly efficient, radiation-free, and non-invasive imaging modality for imaging soft tissues in diagnostic settings. Despite these advantages, ultrasound images are degraded by signal-dependent speckle noise, arising from the coherent interference of ultrasonic waves scattered by sub-resolution tissue structures. Speckle is an inherent characteristic of all coherent imaging systems, where the constructive and destructive interference of waves gives rise to speckle. Thus, speckle is seen in applications such as microscopy, astronomy, sonar, among several other applications, and ultrasound is used as a representative example. The speckle, combined with the additive Gaussian noise caused by thermal fluctuations in acquisition sensors and the PSF-induced distortions, severely degrades the image quality and limits the performance of automated downstream algorithms. Addressing these challenges requires a unified framework to simultaneously mitigate additive and multiplicative noise components while reducing PSF-induced distortions.

Contributions: In this work, we propose DEMIX, the first PSF-aware dual-encoder diffusion-inspired denoising framework to simultaneously reduce additive and multiplicative noise from ultrasound images. The key contributions are:

- DEMIX leverages a novel dual-encoder strategy to separately model signal-dependent speckle and signal-independent Gaussian noise, leading to superior denoising performance.
- A latent masking mechanism combines the encodings of the two encoders with a masked, attention-like gating function to disentangle complex heterogeneous noise in the latent space while preserving structural details in the reconstructed images.
- Lateral and axial PSF distortions are encoded within each encoder, allowing the model to adaptively guide the denoising model to focus on the distorted regions.
- Extensive experiments on two ultrasound image datasets, including downstream evaluations, demonstrate the strong generalizability of DEMIX in handling varying noise combinations and PSF distortions. These results highlight the robustness and generalizability of DEMIX in diverse real-world ultrasound imaging conditions.

2 Related Works

Image Denoising: Image denoising has been extensively studied for different imaging modalities with classical optimization-based methods as well as modern deep learning architectures. Spatial domain filtering algorithms have been widely studied in (Wiener, 1964; Tomasi & Manduchi, 1998; Yang et al., 1996; Bouboulis et al., 2010; Benesty et al., 2010; Yang et al., 1995) and can be broadly categorized as linear

spatial filters (Gonzalez, 2009; Benesty et al., 2010) and non-linear ones (Yang et al., 1995; Tomasi & Manduchi, 1998). Traditional image restoration algorithms include total variation regularization (Rudin et al., 1992), non-local means (NLMeans) algorithm (Buades et al., 2011) and transform domain-based algorithms (Strang, 1999; Ahmed et al., 2006; Ergen, 2012). In the past decade, a significant number of deep learning algorithms have been proposed for image denoising applications (Izadi et al., 2023; Tian et al., 2020a). Convolutional neural networks (CNNs) have been used to denoise imaging from a wide range of image sources (Tian et al., 2020b; Lefkimmiatis, 2018; Zhang et al., 2021; Jin et al., 2019b). These algorithms include supervised, semi-supervised, and unsupervised methods that have been widely adopted to design cutting-edge, efficient algorithms (Zhou et al., 2020; Kim & Ye, 2021; Laine et al., 2019; Pang et al., 2021; Cui et al., 2019). Adversarial training has also been very dominant in the design of novel image denoising algorithms (Zhong et al., 2020; Chen et al., 2020; Lin et al., 2023; Khan et al., 2021). However, all these algorithms have been proposed for images that are corrupted with a single noise form, which in most cases is additive white Gaussian noise (AWGN). However, images obtained from many sensors are corrupted with complex noise patterns arising from a mixture of noise types.

Mixed Noise Removal: Real-world images are often degraded by complex, heterogeneous noise distributions arising from different aspects of the imaging device. Image restoration algorithms for hyperspectral images corrupted with mixed noise have been widely studied (Zhuang & Ng, 2021; Zhang et al., 2019a; Zheng et al., 2019; Luo et al., 2021; Wang et al., 2020; Rasti et al., 2019; Zhuang et al., 2021). Hyperspectral images are corrupted by signal-dependent Poisson noise, signal-independent additive Gaussian noise, and impulse noise (Rasti et al., 2018). Various approaches have been proposed for restoring hyperspectral images, such as subspace learning methods (Shi et al., 2024; Chen et al., 2024; Sun et al., 2018; Wu & Li, 2024; Cao et al., 2019; Wang et al., 2023; Dong et al., 2019), Gaussian mixture models exploiting spatial and spectral information (Zhuang & Ng, 2021; Houdard et al., 2018; Jin et al., 2019a) and recent neural network-based image denoising algorithms (Zeng et al., 2024; Dong et al., 2019). Beyond hyperspectral imaging, image enhancement algorithms have been widely studied for images corrupted with additive white Gaussian noise (AWGN) and impulse noise (Huang et al., 2017; Kim et al., 2020; Jiang et al., 2014; Liu et al., 2017), images corrupted by a mixture of Poisson and Gaussian noise (Luisier et al., 2010; Le Montagner et al., 2014; Mannam et al., 2022; Zhang et al., 2019b; Khademi et al., 2021). These methods include optimization methods, statistical modeling, and deep learning frameworks.

Diffusion Models: Diffusion models (Ho et al., 2020; Nichol & Dhariwal, 2021; Song et al., 2020) have recently emerged as the state-of-the-art models for image generation applications such as image inpainting (Lugmayr et al., 2022; Yang et al., 2023), image super-resolution (Sahak et al., 2023; Wu et al., 2023) and inverse problems (Chung et al., 2022; Daras et al., 2024). Diffusion models comprise a forward process and a reverse process, and the model learns to generate realistic data from an isotropic Gaussian, thus being able to model complex data distributions. However, most diffusion models rely on a single noise model, which significantly restricts their applicability to real-world data that are degraded by heterogeneous noise.

Limitations of previous works: Images acquired from a wide class of imaging modalities, such as ultrasound, synthetic aperture radar (SAR), and optical coherence tomography (OCT), are corrupted by a combination of signal-dependent speckle and signal-independent Gaussian noise. However, very few studies have considered this specific combination of heterogeneous noise corruption that is observed in coherent imaging systems. Conventional denoising algorithms, which are designed exclusively for additive or multiplicative noise, struggle to enhance these images. Designing a unified generalizable framework for signal-dependent speckle and signal-independent Gaussian noise remains an underexplored and challenging research area.

3 Background: PSF for Ultrasound Images

The point spread function (PSF) of the ultrasound imaging system can be assumed to be the impulse response of the system. The PSF $p(i, j)$ is linear and space-invariant, where i and j represent the lateral and axial directions of the ultrasonic beam. The PSF $p(i, j)$ can be expressed as the convolution of the one-dimensional

functions $p_1(i)$ and $p_2(j)$, which represent the distortions along the lateral and axial directions, respectively. The lateral distortion function $p_1(i)$ accounts for the spatial response of the transducer aperture, which acts first as the transmitter and then as the receiver. The axial distortion $p_2(j)$ can be represented as a Gaussian-weighted sinusoid, also known as a Gabor function. Eq. 1 and 2 represent the lateral and axial PSFs where σ_x represents the width of the transmitted ultrasonic beam, σ_y is the pulse width of the ultrasonic wave that is transmitted through tissue, f_0 is its center frequency, and c represents the speed of sound through tissue (Guha et al., 2025).

$$p_1(i) = \exp\big(-\frac{i^2}{2\sigma_x^2}\big) \tag{1}$$

$$p_2(j) = \sin(\frac{2\pi f_0}{c}\, j)\, \exp\big(-\frac{j^2}{2\sigma_y^2}\big) \tag{2}$$

Computationally, the PSF can be modeled as a filter. For a chosen size, the lateral and axial distortions can be represented as functions of σ_x and σ_y, and the PSF is given by:

$$p(\sigma_x, \sigma_y) = p_1(\sigma_x) \,*\, p_2(\sigma_y) \tag{3}$$

A noisy ultrasound image I_{noisy} corrupted by both additive (η^a) and multiplicative (η^m) noise and further degraded by the PSF can be represented by Eq. 4. Fig. 1 illustrates Eq. 4.

$$I_{noisy} = (I + I\,\eta^m) * p_1(\sigma_x) \,*\, p_2(\sigma_y) + \eta^a \tag{4}$$

4 Method

In this section, we show the forward and reverse processes for DEMIX, the training objective, and the model architecture.

4.1 Mixed Noise Formulations

The observed image I_0 is corrupted by additive and multiplicative noise and further degraded by the PSF p of the imaging system. We consider multiple noise levels, and for a given noise level t, the noisy image is represented as I_t. The additive and multiplicative noise components are represented by η_t^a and η_t^m, where $t \in \{1, 2, ..., T\}$, and T is the maximum noise level.

The additive noise, arising from the thermal fluctuations in the sensors, is modeled as zero-mean Gaussian noise (Michailovich & Tannenbaum, 2006), and the speckle can be approximated as a zero-mean Gaussian with some variance. For a specific noise level t, additive noise (variance β_t^2) and speckle (variance α_t^2) can be represented as η_t^a and η_t^m, respectively.

$$\eta_t^m \sim \mathcal{N}(0, \alpha_t^2)\ \eta_t^a \sim \mathcal{N}(0, \beta_t^2) \tag{5}$$

Following Eq. 5, the noisy image I_t can be expressed as Eq. 6.

$$I_t = I_0(1 + \eta_t^m) * p + \eta_t^a \tag{6}$$

For $\epsilon_a, \epsilon_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the noisy image I_t can be represented by Eq. 7.

$$I_t = I_0(1 + \alpha_t \epsilon_m) * p + \beta_t \epsilon_a \tag{7}$$

Assuming both additive and multiplicative noises vary linearly, I_t can be expressed as Eq. 8 where $\alpha_t - \alpha_{t-1} = \delta$ and $\beta_t - \beta_{t-1} = \gamma\ \forall t \geq 2$ and $t \leq T$.

$$I_t = I_0 * p + (I_0 \alpha_{t-1} \epsilon_m) * p + \beta_{t-1}\epsilon_a + (I_0 \delta \epsilon_m) * p + \gamma \epsilon_a \tag{8}$$

Comparing Eq. 7 and Eq. 8,

$$I_t = I_{t-1} + (I_0 \delta \epsilon_m) * p + \gamma \epsilon_a \tag{9}$$

In terms of the initial image I_0, the noisy image I_t can be expressed as follows, where $\epsilon_a, \epsilon_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ are randomly sampled.

$$I_t = I_0 + t[(I_0 \delta \epsilon_m) * p + \gamma \epsilon_a] \tag{10}$$

Forward Process: δ and γ are positive scalars. $\epsilon_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $I_0 \epsilon_m$ represents element-wise multiplication. The system PSF p is the convolution of a zero-mean Gaussian function and a Gabor function. In the Fourier domain, this corresponds to a multiplication of the Fourier transforms of these functions and is analogous to a low-pass filter. When the frequency of the sinusoid is very high, the high-frequency components are attenuated and the resulting PSF p can be approximated as a zero-mean Gaussian. Thus, in Eq. 9 and 10, $(I_0 \delta \epsilon_m) * p = \delta (I_0 \epsilon_m) * p$ can be represented as $\delta K \epsilon'$, where K is a function of the initial image I_0 and $\epsilon' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Now, $(I_0 \delta \epsilon_m) * p + \gamma \epsilon_a$ can be expressed as $\delta K \epsilon' + \gamma \epsilon_a$, where $\epsilon', \epsilon_a \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Thus,

$$(I_0 \delta \epsilon_m) * p + \gamma \epsilon_a \sim \mathcal{N}(\mathbf{0}, (\delta^2 K^2 + \gamma^2)\mathbf{I}) \tag{11}$$

Using Eq. 11, Eq. 9 and Eq. 10 can be expressed as the following where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

$$I_t = I_{t-1} + \sqrt{\delta^2 K^2 + \gamma^2} \epsilon \tag{12}$$

$$I_t = I_0 + t\sqrt{\delta^2 K^2 + \gamma^2} \epsilon \tag{13}$$

Thus, the forward process is given by Eq. 14 and Eq. 15.

$$q(I_t | I_{t-1}) \sim \mathcal{N}(I_t; I_{t-1}, (\delta^2 K^2 + \gamma^2)\mathbf{I}) \tag{14}$$

$$q(I_t | I_0) \sim \mathcal{N}(I_t; I_0, t^2(\delta^2 K^2 + \gamma^2)\mathbf{I}) \tag{15}$$

Reverse Process: To design the reverse process, the analytical inverse of the forward process $q(I_{t-1}|I_t)$ is calculated.

$$q(I_{t-1} | I_t) \sim \mathcal{N}\big(I_{t-1}; \mu_q(I_t, I_0), \Sigma_q(I_0, t)\big) \tag{16}$$

In Eq. 16, $\mu_q(I_t, I_0) = \dfrac{(t-1)^2 I_t + I_0}{(t-1)^2 + 1}$ and $\Sigma_q(I_0, t) = \dfrac{(\delta^2 K^2 + \gamma^2)(t-1)^2}{(t-1)^2 + 1}$. Thus, the reverse process can be formulated as

$$f_\theta(I_{t-1} | I_0) \sim \mathcal{N}\big(I_t; \mu_\theta(I_t, \alpha, \beta, \psi), \Sigma_\theta(I_t, t)\big) \tag{17}$$

In Eq. 17, σ, α, and ψ represent the additive noise schedule, the multiplicative noise schedule, and the PSF parameters, respectively. In the reverse process, we assume $\Sigma_\theta(I_t, t) = \Sigma_q(I_0, t)$.

$$\mu_\theta(I_t, \alpha, \beta, \psi) = \frac{(t-1)^2 I_t + I_\theta(I_t, \alpha, \beta, \psi)}{(t-1)^2 + 1} \tag{18}$$

Training Objective In the reverse process, $\mu_\theta(I_t, \alpha, \beta, \psi)$ and $\Sigma_\theta(I_t, t)$ are functions of the noisy image I_t. The first component of the loss function for DEMIX is:

$$\mathcal{L}_{\mathrm{D}} = \mathbf{E}\big[||I_\theta(I_t, \alpha, \beta, \psi) - I_0||_1\big] \tag{19}$$

The diffusion loss is complemented with the multiscale SSIM loss to enhance the quality of the reconstructed images.

$$\mathcal{L}_{\text{MS-SSIM}} = 1 - \text{MS-SSIM}(I_\theta(I_t, \alpha, \beta, \psi), I_0) \tag{20}$$

The final training objective is:

$$\mathcal{L} = \mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\text{MS-SSIM}} \tag{21}$$

4.2 Model Architecture

In this section, we present a novel dual encoder framework for denoising images that are simultaneously corrupted with additive and multiplicative noise and further degraded by the PSF of the imaging system. The model explicitly accounts for the axial and lateral distortions of the PSF, enabling adaptive restoration of noisy images while preserving critical structural details. The key components of the proposed model are:

1. The network can reduce arbitrary levels of additive and multiplicative noise along with PSF distortions along the axial and lateral directions in a single step.
2. Each encoder independently encodes the additive and multiplicative noise characteristics. Their latent representations are integrated with a masked gating mechanism to feed the bottleneck layer and the skip connections, effectively disentangling the noise components.
3. Additive and multiplicative noise schedules, σ and α, are embedded within each convolutional block, enabling the model to dynamically assess noise intensity and achieve robust denoising across images corrupted with varying noise intensities without explicit noise supervision.
4. The complete spectrum of axial and lateral PSF distortions (ψ) is encoded and integrated into the encoder layers. This enables the model to dynamically estimate the degree of structural degradation, leading to an effective restoration of clean images.

An overview of the proposed architecture is given in Fig. 2. The different components are described below.

Algorithm 1 Training DEMIX

Input: $\{I\}, \alpha, \beta, \psi, I_\theta$
Output: Trained I_θ

while not converged **do**
  $I_0 \in \{I\}$
  $\sigma_x, \sigma_y \sim \psi$
  $t \sim \text{Uniform}(1, ..., T)$
  $\epsilon_m \sim \mathcal{N}(0, I),\ \epsilon_a \sim \mathcal{N}(0, I)$
  $I_t = I_0(1 + \alpha_t \epsilon_m) * p(\sigma_x, \sigma_y) + \beta_t \epsilon_a$
  $\tilde{I} = x_\theta(I_t, \alpha, \beta, \psi)$
  Compute $\nabla_\theta \|\tilde{I} - I_0\|_1$
  Update θ
end while

Noise Encoder: The noise encoder integrates the additive and multiplicative noise schedules, $\beta \in \mathbb{R}^{1\times T}$ and $\alpha \in \mathbb{R}^{1\times T}$, into a unified embedding that quantifies noise severity without explicit supervision. For a noisy $n \times n$ sized input image, the noise schedules are first projected to a high-dimensional space, $\beta_{pos} \in \mathbb{R}^{T\times n}$ and $\alpha_{pos} \in \mathbb{R}^{T\times n}$ with sinusoidal positional embeddings and thereafter processed with a two-layer MLP with GELU activation to obtain $\beta^{'}_{pos} \in \mathbb{R}^{T\times 4n}$ and $\alpha^{'}_{pos} \in \mathbb{R}^{T\times 4n}$. The learned embeddings $\beta^{'}_{pos}$ and $\alpha^{'}_{pos}$ are processed through parallel MLP layers to obtain intermediate representations z_β and z_α, where $z_\beta = ReLU(W_\beta \beta^{'}_{pos})$ and $z_\alpha = ReLU(W_\alpha \alpha^{'}_{pos})$. These are added, and a final MLP layer is applied to fuse the noise characteristics $\alpha\beta = ReLU(W_\beta(z_\beta + z_\alpha)) \in \mathbb{R}^{T\times 4n}$.

PSF Encoder: The PSF encoder models the spatial degradations caused by the axial and lateral distortions of the ultrasound imaging system. For each direction, the minimum and maximum distortions are chosen. Next, a parameter vector $\in \mathbb{R}^{1\times m}$ is constructed by interpolating m values between the chosen minimum and maximum values of distortions and stacked to obtain $\psi \in \mathbb{R}^{2\times m}$. ψ captures the complete spectrum of PSF degradations and facilitates generalization in varying degrees of spatial degradation. ψ is first embedded with sinusoidal positional embeddings to obtain high dimensional embedding $\psi_{pos} \in \mathbb{R}^{2\times 1\times P_d}$, where $P_d = 2m$. The positional embedding is processed through a two-layer MLP with GELU activation function, yielding the final PSF embedding $\Psi = MLP(\psi_{pos}) \in \mathbb{R}^{2\times 1\times m}$. This embedding enables the model to dynamically adapt to varying degrees of PSF distortion for effective image reconstruction.

Gated Fusion Block: The Gated Fusion Block integrates the latent features of the multiplicative and additive noise encoders χ_{mul} and χ_{add}. This strategy is similar to the gating mechanism proposed in (Hu et al., 2018). To promote robust training, a masking strategy ensures that at least one of the

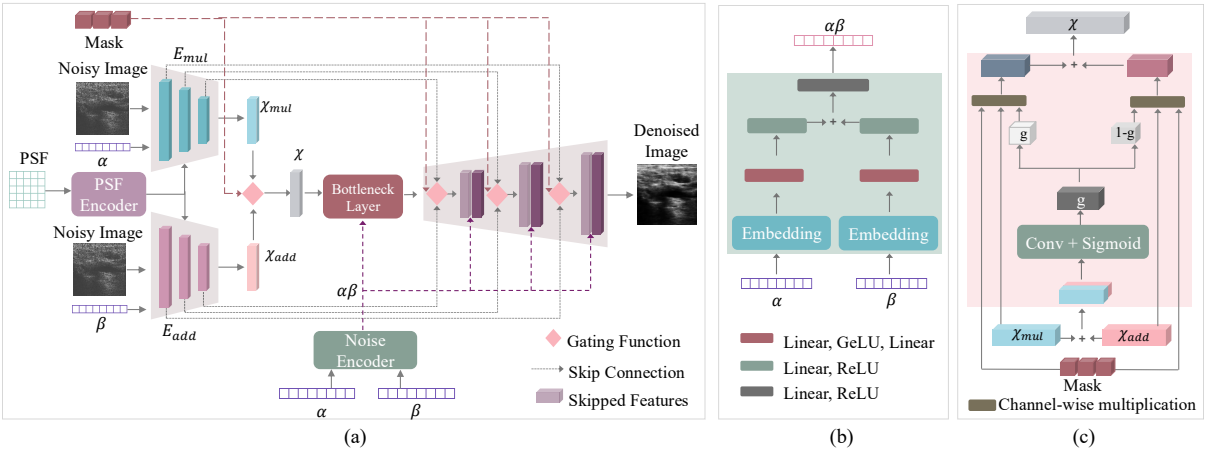

Figure 2: (a) The **overall architecture** is shown and α and β are the multiplicative and additive schedules. The dual encoder architecture processes and extracts the additive and multiplicative noise features separately and combines them through the gated fusion block. The combined features are subsequently processed by the decoder layers, and the final denoised image is obtained. (b) The **noise encoder** combines the multiplicative and additive noise schedules (α) and β, respectively) and embeds them together to create the final noise embedding vector $\alpha\beta$, which is combined with the bottleneck layer and all decoder layers of DEMIX. (c) The **gated fusion block** combines the multiplicative χ_{mul} and additive χ_{add} noise features, along with the sampled mask and fuses them to obtain a feature vector χ for further processing in the bottleneck layer and the decoder layers for efficient denoising performance.

encoders remains active, with the same masking strategy applied to both the bottleneck layer and the skip connections. Let f_1, f_2 represent multiplicative and additive features for different encoder layers, respectively. For features $f_1, f_2 \in \mathbb{R}^{B\times C\times H\times W}$, the features are fused through the learned gate $g = \beta(Conv_{1\times1}([f_1, f_2])) \in \mathbb{R}^{B\times C\times H\times W}$, where $[f_1, f_2]$ represents channelwise concatenation, $Conv_{1\times1}$ represents convolution by a 1×1 kernel and β is the Sigmoid activation function. The learned gate g determines the influence of each encoder on the final embedding. A mask $m = \{[1, 1], [1, 0], [0, 1]\}$ is randomly sampled for each image, ensuring that at least one encoder contributes to the final learned representations. The final fused representation is given by $f = m_1 g\, f_1 + m_2(1 - g)f_2$ where m_1 and m_2 are the broadcast masks and represents element-wise multiplication.

Denoising Framework: The reverse process is implemented with a dual encoder UNet backbone and a masked gated fusion block that unifies the additive χ_{add} and multiplicative χ_{mul} features for bottleneck and skip connections. To enhance robustness, one encoder is randomly masked during training, forcing the network to learn semantically meaningful features even from partial information. The noise encoder encodes both β and α in a unified representation $\alpha\beta$, which conditions the bottleneck layer and all the decoder layers. Additionally, each encoder layer is integrated with the PSF features Ψ, allowing the model to account for spatial degradations. The overall training methodology is given in Algorithm 1. Additive and multiplicative noise schedules, along with lateral and axial PSF distortions, are encoded within the framework. To enforce robustness, masks $\in \{[1, 1], [1, 0], [0, 1]\}$ are randomly sampled. For [1, 0] and [0, 1], the additive and multiplicative components are dropped, respectively. This ensures that DEMIX learns to disentangle the individual noise components and adaptively denoise different intensities of mixed noise along with varying degrees of PSF degradation, without any additional supervision.

4.3 Datasets

Ultrasound Dataset: This dataset comprises two different classes of ultrasound images acquired from human subjects. The first group contains images of the brachial plexus, lymph nodes and fetal heads (LLC, 1999). The second group consists of images of the common carotid artery (CCA) collected from 10

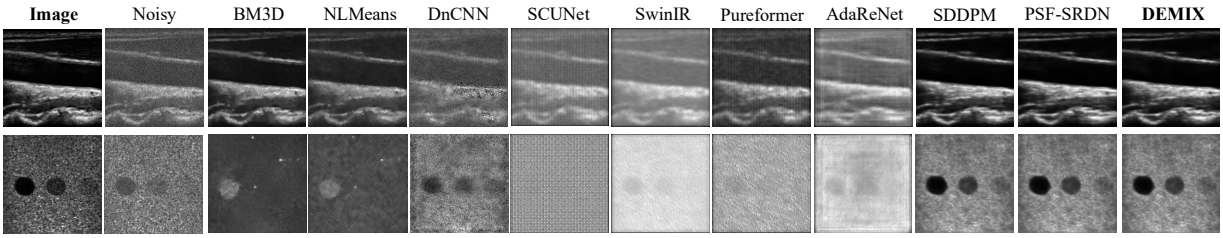

Figure 3: Qualitative comparisons of different denoising methods for $\alpha_t = 0.373, \beta_t = 0.127$ and $\sigma_x = 3, \sigma_y = 2.5$. Column 1: Ground truth images from ultrasound dataset (row 1) and the phantom dataset (row 2). Column 2: Noisy images. Columns 3 - 12: Denoised images by different algorithms. The results for the proposed method DEMIX are shown in the last column.

volunteers (Zhang & Zhang, 2022). The images of the first group are 256×320 pixels. The size of images in the second category range between 230×390 pixels and 450×600 pixels.

Phantom Dataset: This dataset comprises images of eight categories, with each category comprising 100 images. For six categories of the acquired images, each group has at least axial-lateral resolution targets or two different grayscale targets. The other two categories were acquired at different depths (40 mm and 70 mm), introducing different degrees of ultrasound attenuation (Guha et al., 2025).

Table 1: Quantitative comparisons for different denoising algorithms on different datasets. The best value is presented in bold while the second best is underlined. Higher values are expected for both PSNR and SSIM.

Methods	$\sigma_x = 2, \sigma_y = 1.5$								$\sigma_x = 4, \sigma_y = 3.5$							
	$\alpha_t = 0.373$ $\beta_t = 0.127$		$\alpha_t = 0.749$ $\beta_t = 0.251$		$\alpha_t = 1.124$ $\beta_t = 0.376$		$\alpha_t = 1.5$ $\beta_t = 0.5$		$\alpha_t = 0.373$ $\beta_t = 0.127$		$\alpha_t = 0.749$ $\beta_t = 0.251$		$\alpha_t = 1.124$ $\beta_t = 0.376$		$\alpha_t = 1.5$ $\beta_t = 0.5$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ULTRASOUND DATASET																
BM3D (Dabov et al., 2007)	26.31	0.672	19.12	0.116	15.73	0.033	14.67	0.015	26.35	0.672	19.28	0.117	15.85	0.033	14.83	0.015
NLMeans (Buades et al., 2005)	25.42	0.478	22.81	0.250	22.17	0.146	20.76	0.095	25.33	0.478	22.76	0.249	21.24	0.148	21.00	0.096
DnCNN (Zhang et al., 2017)	19.55	0.238	20.39	0.212	20.44	0.188	20.74	0.177	19.77	0.241	20.20	0.211	20.50	0.190	20.18	0.176
SCUNet (Zhang et al., 2023)	5.80	0.113	6.56	0.092	7.38	0.070	8.47	0.051	5.79	0.113	6.51	0.092	7.41	0.071	8.84	0.051
SwinIR (Liang et al., 2021)	5.35	0.154	6.63	0.109	7.88	0.071	9.48	0.044	5.42	0.154	6.63	0.109	7.76	0.071	9.51	0.044
Pureformer (Gautam et al., 2025)	5.96	0.142	7.08	0.101	8.51	0.067	11.22	0.042	5.94	0.142	7.24	0.101	9.06	0.067	11.20	0.042
AdaReNet (Liu et al., 2025)	6.98	0.111	7.74	0.080	8.69	0.058	9.92	0.040	6.91	0.111	7.77	0.080	8.96	0.057	9.91	0.040
SDDPM (Guha & Acton, 2023)	26.16	0.727	23.78	0.643	22.47	0.590	21.58	0.535	26.24	0.727	23.60	0.646	22.62	0.588	21.61	0.545
PSF-SRDN (Guha et al., 2025)	25.96	0.723	23.92	0.647	22.83	0.592	21.88	0.556	26.06	0.722	24.03	0.650	22.71	0.595	21.85	0.556
DEMIX (Ours)	**26.85**	**0.745**	**24.55**	**0.671**	**23.22**	**0.629**	**22.27**	**0.586**	**26.81**	**0.747**	**24.36**	**0.675**	**22.95**	**0.629**	**22.24**	**0.585**
PHANTOM DATASET																
BM3D (Dabov et al., 2007)	24.33	0.637	16.78	0.087	14.57	0.022	13.88	0.010	24.20	0.637	16.74	0.087	14.56	0.022	13.89	0.010
NLMeans (Buades et al., 2005)	22.91	0.442	19.89	0.245	19.67	0.148	18.18	0.096	23.34	0.472	20.02	0.242	18.80	0.138	18.17	0.086
DnCNN (Zhang et al., 2017)	19.87	0.224	19.96	0.188	19.80	0.180	19.73	0.187	19.95	0.225	20.12	0.187	19.63	0.180	19.85	0.188
SCUNet (Zhang et al., 2023)	7.08	0.158	7.62	0.142	8.39	0.120	9.49	0.095	7.08	0.158	7.63	0.142	8.39	0.121	9.47	0.095
SwinIR (Liang et al., 2021)	5.02	0.244	6.11	0.160	7.53	0.093	9.75	0.051	5.02	0.243	6.10	0.160	7.65	0.092	9.68	0.051
Pureformer (Gautam et al., 2025)	5.81	0.190	7.91	0.133	10.65	0.081	13.38	0.047	5.74	0.190	7.95	0.133	10.59	0.082	13.20	0.047
AdaReNet (Liu et al., 2025)	5.26	0.179	5.68	0.126	6.18	0.081	6.96	0.050	5.27	0.179	5.69	0.126	6.17	0.081	6.94	0.050
SDDPM (Guha & Acton, 2023)	22.96	0.652	21.06	0.600	20.06	0.584	19.31	0.577	22.94	0.652	21.08	0.600	19.98	0.584	19.31	0.577
PSF-SRDN (Guha et al., 2025)	22.14	0.636	20.79	0.603	19.79	0.583	19.06	0.574	22.14	0.636	20.76	0.603	19.74	0.583	19.09	0.569
DEMIX (Ours)	**25.10**	**0.664**	**22.33**	**0.607**	**21.29**	**0.589**	**20.76**	**0.581**	**25.00**	**0.664**	**22.41**	**0.607**	**21.38**	**0.589**	**20.79**	**0.581**

5 Experiments

5.1 Implementation Details

All experiments were implemented on a NVIDIA RTX 3090 GPU. The stochastic gradient descent (SGD) optimizer and a batch size of 128 were used. All models were trained on randomly cropped 64×64 patches of noisy, PSF-distorted images for 150 epochs for the ultrasound dataset and 200 epochs for the phantom dataset. The experiments were trained with an initial learning rate of 0.005, which was reduced to 10% of the value every 50 epochs. For both datasets, three randomly sampled training and validation partitions are used. Each time, 70% of the images were reserved for training and the rest 30% were used to validate the trained model. The images in the training partition were randomly augmented by rotation and Gaussian noise having $\mu = 0.025, 0.05$ and $\sigma^2 = 0, 0.001$. The augmented training set contains approximately 6000 images and 13000 images for the ultrasound and phantom datasets, respectively. The proposed mixed noise

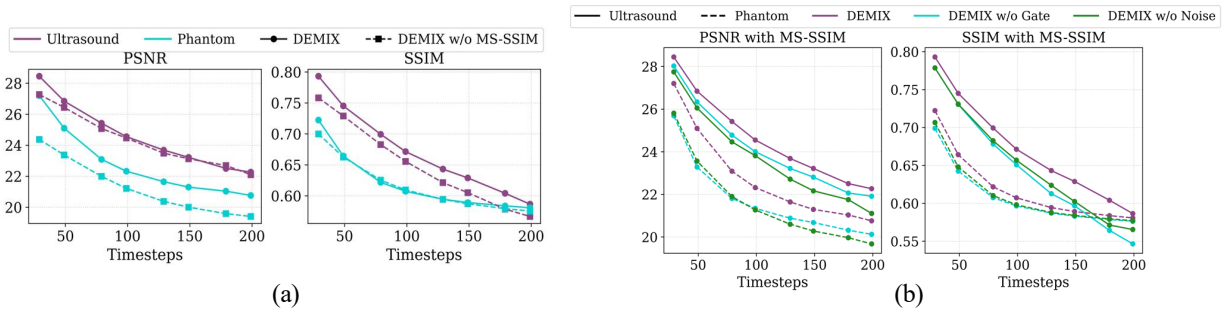

Figure 4: (a) The MS-SSIM loss (shown for $\sigma_x = 2, \sigma_y = 1.5$) improves the PSNR and SSIM scores for both datasets for the proposed method DEMIX. (b) With MS-SSIM loss, the PSNR and SSIM scores (shown for $\sigma_x = 2, \sigma_y = 1.5$) for DEMIX are consistently higher than the ablated models without the gated fusion and the noise encoder.

denoising model was trained with $\alpha = [0.005, 1.5]$, $\beta = [0.005, 0.5]$ and $T = 200$. A 3×3 PSF was considered for the ultrasound imaging system that has a center frequency $f_0 = 10\,\text{MHz}$ and the speed of sound through the tissue is $c = 1540\,\text{m/s}$. The lateral and axial distortions for the ultrasound PSF were bounded between $\sigma_x = [1, 4]$ and $\sigma_y = [0.5, 3.5]$. 50 values of ($m = 50$) were interpolated for both σ_x and σ_y to facilitate a detailed evaluation over a wide range of axial and lateral distortions.

5.2 Comparisons with the State-of-the-Art

Evaluation Metrics: We report the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) for all experiments. PSNR, reported in decibels (dB), quantifies the intensity differences between the reconstructed and ground truth images, and SSIM evaluates perceptual quality by assessing luminance, contrast, and structural consistency. Higher values of PSNR and SSIM indicate better results.

Baseline Methods: We conducted extensive experiments on the two datasets to compare the performance of DEMIX with the state-of-the-art image denoising methods, including BM3D (Dabov et al., 2007), NLMeans (Buades et al., 2005), DnCNN (Zhang et al., 2017), SwinIR (Liang et al., 2021), SCUNet (Zhang et al., 2023), Pureformer (Gautam et al., 2025), AdaReNet (Liu et al., 2025), SDDPM (Guha & Acton, 2023) and PSF-SRDN (Guha et al., 2025). BM3D (Dabov et al., 2007) and NLMeans (Buades et al., 2005) are traditional image denoising algorithms, whereas all other baselines are data-driven. BM3D (Dabov et al., 2007) denoises by grouping similar patches into 3D stacks and collaboratively filtering them in a transform domain, whereas NLMeans (Buades et al., 2005) denoises by averaging the pixel intensities having similar local neighborhoods throughout the image. DnCNN (Zhang et al., 2017) targets Gaussian denoising by learning image residuals to separate clean latent images from noisy inputs. SwinIR (Liang et al., 2021) combines Swin transformer (Liu et al., 2021) layers and residual connections for an image restoration algorithm and integrates shallow and deep feature extractions with high fidelity image reconstruction. SCUNet (Zhang et al., 2023) integrates new swin-conv blocks into a UNet backbone for a blind image denoising framework. Pureformer (Gautam et al., 2025) is an encoder-decoder architecture with transformer-based skip connections and comprises latent feature enhancer blocks that use a spatial filter bank to expand the receptive field, thereby improving image restoration. AdaReNet (Liu et al., 2025) proposes a rotation-invariant prior in a supervised learning framework, and rotation-invariant convolutional layers are used to replace all translation-invariant convolutional layers in a UNet backbone. SDDPM (Guha & Acton, 2023) and PSF-SRDN (Guha et al., 2025) are diffusion models specifically proposed to reduce signal-dependent speckle. While SDDPM is designed to reduce speckle, PSF-SRDN is additionally equipped with the PSF information that helps to reduce speckle and correct PSF-induced distortions.

Results: All methods were evaluated for different strengths of additive and multiplicative noise for varying degrees of PSF distortions. The evaluation scores for the two datasets are demonstrated in Table 1. The proposed method, DEMIX, outperforms all baseline methods for the ultrasound dataset and obtains com-

Table 2: Ablation studies of different components of DEMIX on Ultrasound and Phantom datasets. PSNR and SSIM are reported for $\sigma_x = 2, \sigma_y = 1.5$. The best scores are shown in bold letters.

Dataset	Method	Loss	$\alpha_t = 0.373$ $\beta_t = 0.127$		$\alpha_t = 0.749$ $\beta_t = 0.251$		$\alpha_t = 1.124$ $\beta_t = 0.376$		$\alpha_t = 1.5$ $\beta_t = 0.5$	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Ultrasound	DEMIX w/o Noise Encoder	$\mathcal{L}_{\text{D}}$	26.08	0.729	23.85	0.654	22.53	0.601	21.44	0.557
	DEMIX w/o Noise Encoder	$\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$	26.05	0.731	23.82	0.657	22.17	0.602	21.10	0.565
	DEMIX w/o Gated Fusion	$\mathcal{L}_{\text{D}}$	26.45	0.728	23.99	0.651	22.55	0.597	21.91	0.564
	DEMIX w/o Gated Fusion	$\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$	26.34	0.731	24.00	0.651	22.81	0.597	21.91	0.546
	DEMIX	$\mathcal{L}_{\text{D}}$	26.45	0.730	24.46	0.657	22.96	0.603	22.46	0.567
	DEMIX (proposed)	$\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$	**26.85**	**0.745**	**24.55**	**0.671**	**23.22**	**0.629**	**22.27**	**0.586**
Phantom	DEMIX w/o Noise Encoder	$\mathcal{L}_{\text{D}}$	24.17	0.650	21.96	0.600	20.94	0.586	20.25	0.578
	DEMIX w/o Noise Encoder	$\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$	23.56	0.648	21.27	0.598	20.28	0.583	19.68	0.577
	DEMIX w/o Gated Fusion	$\mathcal{L}_{\text{D}}$	23.12	0.642	21.20	0.597	20.36	0.583	19.79	0.577
	DEMIX w/o Gated Fusion	$\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$	23.28	0.643	21.34	0.597	20.68	0.583	20.13	0.576
	DEMIX	$\mathcal{L}_{\text{D}}$	23.37	0.663	21.21	**0.609**	20.00	0.587	19.40	0.575
	DEMIX (proposed)	$\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$	**25.10**	**0.664**	**22.33**	0.607	**21.29**	**0.589**	**20.76**	**0.581**

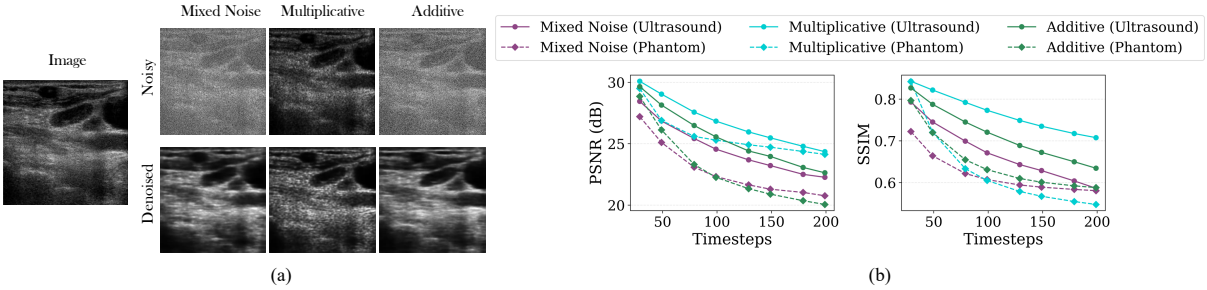

Figure 5: (a) $\sigma_x = 3, \sigma_y = 2.5$, $\alpha_t = 0.598$, $\beta_t = 0.201$ (b) For PSF parameters $\sigma_x = 2, \sigma_y = 1.5$, we show how the PSNR and SSIM scores vary for the different noise combinations for the two datasets. All other PSF distorted images follow similar trends.

petitive scores for the phantom dataset. DEMIX achieves the best SSIM scores for both datasets for all noise levels, outperforming most state-of-the-art methods by a significant margin. SDDPM and PSF-SRDN are specifically proposed for multiplicative noise and perform better than other baselines. Figure 3 shows the original image, the noisy image, and the denoised images obtained from the state-of-the-art denoising algorithms. It can be observed that the images denoised by DEMIX recover the fine structural details present in the original image. Moreover, it can be verified from Table 1 that the PSNR and SSIM scores for the images denoised by DEMIX are significantly higher than all comparative methods even for low SNR and high noise scenarios. Figure 5 shows the generalizability of DEMIX. Figure 5 (a) shows the denoised images reconstructed from different noisy inputs. Despite being trained on both additive and multiplicative noise, the model performs equally well when either component is absent. This indicates the strong generalizability of DEMIX.

Downstream Task: We evaluated the performance of the denoised images for the downstream segmentation task. We used the thyroid nodule segmentation dataset, TN3K (Gong et al., 2021; 2022), which contains 3493 thyroid nodule images along with nodule mask labels. We have used UNet (Ronneberger et al., 2015), AttenUNet (Oktay et al., 2018), Segformer (Xie et al., 2021), and TransUNet (Chen et al., 2021) to evaluate the segmentation performance of the TN3K dataset for the original and denoised images, where the train and test partitions contained 2879 and 614 images, respectively. Table 3 shows the Jaccard Similarity (JS) and the Dice Scores (%). To evaluate the segmentation models with the denoised images, both the training and test partitions were denoised with the model pre-training on the ultrasound data set. It can be seen that, for the majority of cases, the denoised images perform better, which demonstrates the superiority of the proposed denoising method.

Table 3: Downstream evaluation: Segmentation scores for the original dataset and the dataset denoised by DEMIX. We report the Jaccard Similarity (JS) and the Dice Score (%) for an overlap of 0.5.

Method	Original		Denoised	
	JS	Dice (%)	JS	Dice(%)
UNet (Ronneberger et al., 2015)	0.675	80.54	**0.696**	**82.00**
AttnUNet (Oktay et al., 2018)	0.681	80.89	**0.691**	**81.60**
Segformer (Xie et al., 2021)	0.337	50.27	**0.593**	**74.01**
TransUNet (Chen et al., 2021)	**0.637**	77.72	0.634	**77.84**

5.3 Ablation Studies

We study the effect of different components of the proposed architecture. Table 2 shows the PSNR and SSIM scores for different variations of DEMIX for $\sigma_x = 2, \sigma_y = 1.5$. All other combinations of PSF distortions demonstrate similar trends. It can be seen that the proposed architecture and the overall objective function ($\mathcal{L}_{\text{D}} + \mathcal{L}_{\text{MS-SSIM}}$) for DEMIX, including the noise encoder and the gating mechanism, yield the highest evaluation scores among all other variations. Fig. 4 (a) shows the improvement in the evaluation scores for DEMIX with the inclusion of $\mathcal{L}_{\text{MS-SSIM}}$, while Fig. 4 (b) shows the effectiveness of the noise encoder and the gated fusion block in the final evaluation scores. The multiscale SSIM loss consistently improves the evaluation scores.

6 Discussions

6.1 Generalizability of the proposed model

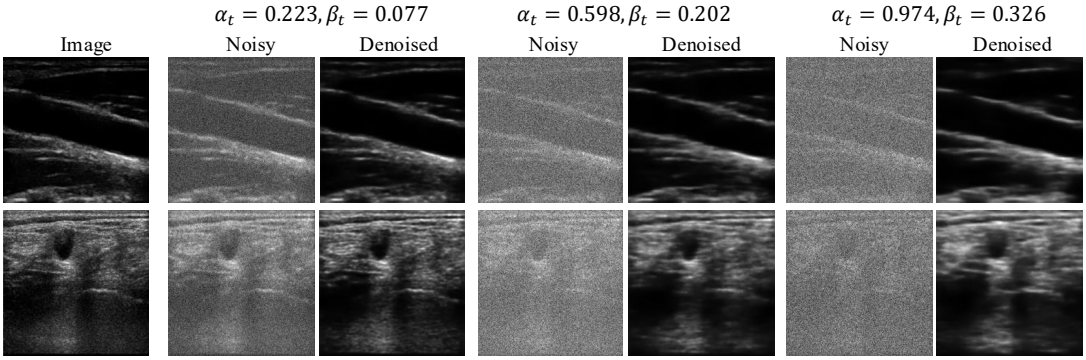

Figure 6: Each row shows the original image and the noisy and denoised images for different combinations of additive and multiplicative noise.

The proposed model is generalizable and can disentangle the different noise components. This is demonstrated by Fig. 5. DEMIX can adaptively suppress the individual noise components and does not overfit to the joint distribution of the additive and multiplicative noise components. This makes DEMIX well-suited for real-world denoising, where the PSF distortion parameters or the exact noise combinations are not explicitly known. In real-world image denoising scenarios, either of the noise components can be absent, and this might lead to undesired performances from the denoising model. The efficiency of DEMIX is further demonstrated by Fig. 6. For different levels of additive and multiplicative noise components, the proposed method recovers the high-frequency details from the noisy, corrupted images without over-smoothing the edges. Additionally, DEMIX does not require that the noise level or the PSF distortion be known beforehand. Since the PSF formulation of the ultrasound images remains the same for all acquisition methods, the model is trained to dynamically adapt to different levels of noise and distortions. This makes the model suitable for denoising noisy acquisitions in real-world settings.

6.2 Efficient training

DEMIX is essentially an image denoising framework proposed in an inverse problem setting. Instead of learning the distribution of images, DEMIX learns the local noise characteristics and reverses the degradation process. This significantly simplifies the training process. Instead of using the full resolution images, the model can be trained on randomly cropped, small patches that capture the representative noise patterns. As a result, the training complexity is significantly reduced. The final model can be generalized to all image sizes. After training, DEMIX can be used for denoising images of arbitrary sizes, without any additional fine-tuning or retraining. Essentially, DEMIX learns the noise characteristics from image patches, which is seamlessly transferred to full-sized images during inference. Patch-efficient training, combined with the formulation of inverse problems, makes DEMIX highly efficient and scalable.

7 Conclusions

This work presents DEMIX, a novel, dual-encoder denoising framework inspired by the diffusion models for mixed noise reduction along with non-uniform, spatially varying PSF-induced distortions. DEMIX demonstrates superior qualitative and quantitative performance for both datasets compared to state-of-the-art image denoising algorithms. The masked gated fusion mechanism makes the model capable of denoising additive, multiplicative, or any combination of additive or multiplicative noise components. This method holds promise for denoising a wide class of images from various applications such as microscopy, astronomy, and sonar, where images are degraded by signal-dependent and signal-independent noise components and additionally degraded by the system PSF.

References

Nasir Ahmed, T_ Natarajan, and Kamisetty R Rao. Discrete cosine transform. *IEEE transactions on Computers*, 100(1):90–93, 2006.

Fatemeh Ashouri and MR Eslahchi. A new pde learning model for image denoising. *Neural Computing and Applications*, 34(11):8551–8574, 2022.

Olivier Beltramo-Martin, Sam Ragland, Romain Fétick, Carlos Correia, Trent Dupuy, Giulian Fiorentino, Thierry Fusco, Laurent Jolissaint, Sebastian Kamann, Antonino Marasco, et al. Review of psf reconstruction methods and application to post-processing. In *Adaptive Optics Systems VII*, volume 11448, pp. 22–36. SPIE, 2020.

Jacob Benesty, Jingdong Chen, and Yiteng Huang. Study of the widely linear wiener filter for noise reduction. In *2010 IEEE international conference on acoustics, speech and signal processing*, pp. 205–208. IEEE, 2010.

Pantelis Bouboulis, Konstantinos Slavakis, and Sergios Theodoridis. Adaptive kernel-based image denoising employing semi-parametric regularization. *IEEE Transactions on Image Processing*, 19(6):1465–1479, 2010.

A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pp. 60–65 vol. 2, 2005. doi: 10.1109/CVPR.2005.38.

Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Non-local means denoising. *Image processing on line*, 1:208–212, 2011.

Chunhong Cao, Jie Yu, Chengyao Zhou, Kai Hu, Fen Xiao, and Xieping Gao. Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(3):973–988, 2019. doi: 10.1109/JSTARS.2019.2896031.

Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

Songkui Chen, Daming Shi, Muhammad Sadiq, and Xiaochun Cheng. Image denoising with generative adversarial networks and its application to cell image enhancement. *IEEE Access*, 8:82819–82831, 2020. doi: 10.1109/ACCESS.2020.2988284.

Yong Chen, Jinshan Zeng, Wei He, Xi-Le Zhao, Tai-Xiang Jiang, and Qing Huang. Fast large-scale hyperspectral image denoising via noniterative low-rank subspace representation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024. doi: 10.1109/TGRS.2024.3458395.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Jianan Cui, Kuang Gong, Ning Guo, Chenxi Wu, Xiaxia Meng, Kyungsang Kim, Kun Zheng, Zhifang Wu, Liping Fu, Baixuan Xu, et al. Pet image denoising using unsupervised deep learning. *European journal of nuclear medicine and molecular imaging*, 46(13):2780–2789, 2019.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

Christoph Dalitz, Regina Pohle-Fröhlich, and Thorsten Michalk. Point spread functions and deconvolution of ultrasonic images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 62:531–544, 2015. ISSN 08853010. doi: 10.1109/TUFFC.2014.006717.

Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio. A survey on diffusion models for inverse problems. *arXiv preprint arXiv:2410.00083*, 2024.

Weisheng Dong, Huan Wang, Fangfang Wu, Guangming Shi, and Xin Li. Deep spatial–spectral representation learning for hyperspectral image denoising. *IEEE Transactions on Computational Imaging*, 5(4):635–648, 2019.

Burhan Ergen. *Signal and image denoising using wavelet transform*. InTech London, UK, 2012.

Arnim Gautam, Aditi Pawar, Aishwarya Joshi, Satya Narayan Tazi, Sachin Chaudhary, Praful Hambarde, Akshay Dudhane, Santosh Vipparthi, and Subrahamanyam Murala. Pureformer: Transformer-based image denoising. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 1441–1449, 2025.

Haifan Gong, Guanqi Chen, Ranran Wang, Xiang Xie, Mingzhi Mao, Yizhou Yu, Fei Chen, and Guanbin Li. Multi-task learning for thyroid nodule segmentation with thyroid region prior. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 257–261, 2021. doi: 10.1109/ISBI48211.2021.9434087.

Haifan Gong, Jiaxin Chen, Guanqi Chen, Haofeng Li, Fei Chen, and Guanbin Li. Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. *Computers in Biology and Medicine*, 106389:1–12, 2022.

Rafael C Gonzalez. *Digital image processing*. Pearson education india, 2009.

Bhawna Goyal, Sunil Agrawal, and BS Sohi. Noise issues prevailing in various types of medical images. *Biomedical & Pharmacology Journal*, 11(3):1227, 2018.

Soumee Guha and Scott T Acton. Sddpm: Speckle denoising diffusion probabilistic models. *arXiv preprint arXiv:2311.10868*, 2023.

Soumee Guha, Yi Huang, John A. Hossack, and Scott T. Acton. Psf-srdn: Point spread function-aware speckle reducing diffusion network. In *2025 IEEE International Conference on Image Processing (ICIP)*, pp. 1618–1623, 2025. doi: 10.1109/ICIP55913.2025.11084687.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Antoine Houdard, Charles Bouveyron, and Julie Delon. High-dimensional mixture models for unsupervised image denoising (hdmi). *SIAM Journal on Imaging Sciences*, 11(4):2815–2846, 2018.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

Tao Huang, Weisheng Dong, Xuemei Xie, Guangming Shi, and Xiang Bai. Mixed noise removal via laplacian scale mixture modeling and nonlocal low-rank approximation. *IEEE Transactions on Image Processing*, 26(7):3171–3186, 2017.

Saeed Izadi, Darren Sutton, and Ghassan Hamarneh. Image denoising in the deep learning era. *Artificial Intelligence Review*, 56(7):5929–5974, 2023.

Peng Jia, Xuebo Wu, Huang Yi, Bojun Cai, and Dongmei Cai. Psf–net: a nonparametric point-spread function model for ground-based optical telescopes. *The Astronomical Journal*, 159(4):183, 2020.

Jielin Jiang, Lei Zhang, and Jian Yang. Mixed noise removal by weighted encoding with sparse nonlocal regularization. *IEEE transactions on image processing*, 23(6):2651–2662, 2014.

Qiwen Jin, Yong Ma, Erting Pan, Fan Fan, Jun Huang, Hao Li, Chenhong Sui, and Xiaoguang Mei. Hyperspectral unmixing with gaussian mixture model and spatial group sparsity. *Remote Sensing*, 11(20):2434, 2019a.

Zhi Jin, Muhammad Zafar Iqbal, Dmytro Bobkov, Wenbin Zou, Xia Li, and Eckehard Steinbach. A flexible deep cnn framework for image restoration. *IEEE Transactions on Multimedia*, 22(4):1055–1068, 2019b.

C. Kervrann and J. Boulanger. Optimal spatial adaptation for patch-based image denoising. *IEEE Transactions on Image Processing*, 15(10):2866–2878, 2006. doi: 10.1109/TIP.2006.877529.

Wesley Khademi, Sonia Rao, Clare Minnerath, Guy Hagen, and Jonathan Ventura. Self-supervised poisson-gaussian denoising. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2131–2139, 2021.

Aamir Khan, Weidong Jin, Amir Haider, MuhibUr Rahman, and Desheng Wang. Adversarial gaussian denoiser for multiple-level image denoising. *Sensors*, 21(9):2998, 2021.

Dai-Gyoung Kim, Mukhtar Hussain, Muhammad Adnan, Muhammad Asif Farooq, Zahid Hussain Shamsi, and Asif Mushtaq. Mixed noise removal using adaptive median based non-local rank minimization. *IEEE Access*, 9:6438–6452, 2020.

Kwanyoung Kim and Jong Chul Ye. Noise2score: tweedie's approach to self-supervised image denoising without clean images. *Advances in Neural Information Processing Systems*, 34:864–874, 2021.

Samuli Laine, Tero Karras, Jaakko Lehtinen, and Timo Aila. High-quality self-supervised deep image denoising. *Advances in neural information processing systems*, 32, 2019.

Yoann Le Montagner, Elsa D Angelini, and Jean-Christophe Olivo-Marin. An unbiased risk estimator for image denoising in the presence of mixed poisson–gaussian noise. *IEEE Transactions on Image processing*, 23(3):1255–1268, 2014.

Stamatios Lefkimmiatis. Universal denoising networks: a novel cnn architecture for image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3204–3213, 2018.

Yiming Li, Markus Mund, Philipp Hoess, Joran Deschamps, Ulf Matti, Bianca Nijmeijer, Vilma Jimenez Sabinina, Jan Ellenberg, Ingmar Schoen, and Jonas Ries. Real-time 3d single-molecule localization using experimental point spread functions. *Nature methods*, 15(5):367–369, 2018.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844, 2021.

Tobías I Liaudat, Jean-Luc Starck, and Martin Kilbinger. Point spread function modelling for astronomical telescopes: a review focused on weak gravitational lensing studies. *Frontiers in Astronomy and Space Sciences*, 10:1158213, 2023.

Xin Lin, Chao Ren, Xiao Liu, Jie Huang, and Yinjie Lei. Unsupervised image denoising in real-world scenarios via self-collaboration parallel generative adversarial branches. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12642–12652, 2023.

Hanze Liu, Jiahong Fu, Qi Xie, and Deyu Meng. Rotation-equivariant self-supervised method in image denoising. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12720–12730, 2025.

Licheng Liu, CL Philip Chen, Xinge You, Yuan Yan Tang, Yushu Zhang, and Shutao Li. Mixed noise removal via robust constrained sparse representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2177–2189, 2017.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows, 2021. URL `https://arxiv.org/abs/2103.14030`.

MultiMedia LLC. MS Windows NT kernel description, 1999. URL `http://splab.cz/en/download/databaze/ultrasound`.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11461–11471, 2022.

Florian Luisier, Thierry Blu, and Michael Unser. Image denoising in mixed poisson–gaussian noise. *IEEE Transactions on image processing*, 20(3):696–708, 2010.

Yi-Si Luo, Xi-Le Zhao, Tai-Xiang Jiang, Yu-Bang Zheng, and Yi Chang. Hyperspectral mixed noise removal via spatial-spectral constrained unsupervised deep image prior. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:9435–9449, 2021.

Varun Mannam, Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Vignesh Sundaresan, Siyuan Zhang, Cody Smith, Paul W Bohn, and Scott S Howard. Real-time image denoising of mixed poisson–gaussian noise in fluorescence microscopy images using imagej. *Optica*, 9(4):335–345, 2022.

O.V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 53(1):64–78, 2006. doi: 10.1109/TUFFC.2006.1588392.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *CoRR*, abs/1804.03999, 2018. URL `http://arxiv.org/abs/1804.03999`.

Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2043–2052, 2021.

Marie Ploquin, Adrian Basarab, and Denis Kouamé. Resolution enhancement in medical ultrasound imaging. *Journal of Medical Imaging*, 2:017001, 3 2015. ISSN 2329-4302. doi: 10.1117/1.jmi.2.1.017001.

Behnood Rasti, Paul Scheunders, Pedram Ghamisi, Giorgio Licciardi, and Jocelyn Chanussot. Noise reduction in hyperspectral imagery: Overview and application. *Remote Sensing*, 10(3):482, 2018.

Behnood Rasti, Pedram Ghamisi, and Jon Atli Benediktsson. Hyperspectral mixed gaussian and sparse noise reduction. *IEEE Geoscience and Remote Sensing Letters*, 17(3):474–478, 2019.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL `http://arxiv.org/abs/1505.04597`.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Hshmat Sahak, Daniel Watson, Chitwan Saharia, and David Fleet. Denoising diffusion probabilistic models for robust image super-resolution in the wild. *arXiv preprint arXiv:2302.07864*, 2023.

Anne E.C.M. Saris, Stein Fekkes, Maartje M. Nillesen, Hendrik H.G. Hansen, and Chris L. de Korte. A psf-shape-based beamforming strategy for robust 2d motion estimation in ultrafast data. *Applied Sciences (Switzerland)*, 8, 3 2018. ISSN 20763417. doi: 10.3390/app8030429.

Colas Schretter, Shaun Bundervoet, David Blinder, Ann Dooms, Jan D'Hooge, and Peter Schelkens. Ultrasound imaging from sparse rf samples using system point spread functions. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65:316–326, 3 2018. ISSN 08853010. doi: 10.1109/TUFFC.2017.2772916.

Yoav Shechtman, Steffen J. Sahl, Adam S. Backer, and W. E. Moerner. Optimal point spread function design for 3d imaging. *Physical Review Letters*, 113, 9 2014. ISSN 10797114. doi: 10.1103/PhysRevLett.113.133902.

Kexin Shi, Jiangjun Peng, Jing Gao, Yisi Luo, and Shuang Xu. Hyperspectral image denoising via double subspace deep prior. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3457792.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Sjoerd Stallinga, Bernd Rieger, E Betzig, G H Patterson, R Sougrat, O W Lindwasser, S Olenych, J S Bonifacino, M W Davidson, J Lippincott-Schwartz, H F Hess, M Heilemann, S van de Linde, M Schttpelz, R Kasper, B Seefeldt, A Mukherjee, P Tinnefeld, and M Sauer. Imaging intracellular fluorescent proteins at nanometer resolution. Technical report, 2009.

Gilbert Strang. The discrete cosine transform. *SIAM review*, 41(1):135–147, 1999.

Le Sun, Byeungwoo Jeon, Bushra Naz Soomro, Yuhui Zheng, Zebin Wu, and Liang Xiao. Fast superpixel based subspace low rank learning method for hyperspectral denoising. *IEEE Access*, 6:12031–12043, 2018. doi: 10.1109/ACCESS.2018.2808474.

Chunwei Tian, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin. Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275, 2020a.

Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 121:461–473, 2020b.

Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pp. 839–846. IEEE, 1998.

Francois Varray, Olivier Bernard, Sonia Assou, Christian Cachard, and Didier Vray. Hybrid strategy to simulate 3-d nonlinear radio-frequency ultrasound using a variant spatial psf. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63:1390–1398, 9 2016. ISSN 08853010. doi: 10.1109/TUFFC.2016.2579666.

Hailin Wang, Jiangjun Peng, Xiangyong Cao, Jianjun Wang, Qibin Zhao, and Deyu Meng. Hyperspectral image denoising via nonlocal spectral sparse subspace representation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:5189–5203, 2023.

Minghua Wang, Qiang Wang, Jocelyn Chanussot, and Dan Li. Hyperspectral image mixed noise removal based on multidirectional low-rank modeling and spatial–spectral total variation. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):488–507, 2020.

Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*. The MIT press, 1964.

Yue Wu and Weisheng Li. Robust and adaptive subspace learning for fast hyperspectral image denoising. *Applied Intelligence*, 54(22):11400–11411, 2024.

Zhanxiong Wu, Xuanheng Chen, Sangma Xie, Jian Shen, and Yu Zeng. Super-resolution of brain mri images based on denoising diffusion probabilistic model. *Biomedical Signal Processing and Control*, 85:104901, 2023.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

Xinheng Xie, Yue Wu, Hao Ni, and Cuiyu He. Node-imgnet: A pde-informed effective and robust model for image denoising. *Pattern Recognition*, 148:110176, 2024.

Guang-Zhong Yang, Peter Burger, David N Firmin, and SR Underwood. Structure adaptive anisotropic image filtering. *Image and Vision Computing*, 14(2):135–145, 1996.

Lihong Yang and Jianyue Ren. Remote sensing image restoration using estimated point spread function. In *2010 International Conference on Information, Networking and Automation (ICINA)*, volume 1, pp. V1–48. IEEE, 2010.

Ruikang Yang, Lin Yin, Moncef Gabbouj, Jaakko Astola, and Yrjö Neuvo. Optimal weighted median filtering under structural constraints. *IEEE transactions on signal processing*, 43(3):591–604, 1995.

Shiyuan Yang, Xiaodong Chen, and Jing Liao. Uni-paint: A unified framework for multimodal image inpainting with pretrained diffusion model. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3190–3199, 2023.

Haijin Zeng, Jiezhang Cao, Kai Zhang, Yongyong Chen, Hiep Luong, and Wilfried Philips. Unmixing diffusion for self-supervised hyperspectral image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27820–27830, 2024.

Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. Technical report, 2007.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7), 2017. ISSN 1941-0042. doi: 10.1109/tip.2017.2662206. URL `http://dx.doi.org/10.1109/TIP.2017.2662206`.

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021.

Kai Zhang, Yawei Li, Jingyun Liang, Jiezhang Cao, Yulun Zhang, Hao Tang, Deng-Ping Fan, Radu Timofte, and Luc Van Gool. Practical blind image denoising via swin-conv-unet and data synthesis. *Machine Intelligence Research*, 20(6):822–836, 2023.

Lun Zhang and Junhua Zhang. Ultrasound image denoising using generative adversarial networks with residual dense connectivity and weighted joint loss. *PeerJ Computer Science*, 8:e873, 2022.

Qiang Zhang, Qiangqiang Yuan, Jie Li, Xinxin Liu, Huanfeng Shen, and Liangpei Zhang. Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10):7317–7329, 2019a.

Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, Cody Smith, and Scott Howard. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11710–11718, 2019b.

Ningning Zhao, Qi Wei, Adrian Basarab, Denis Kouamé, and Jean-Yves Tourneret. Blind deconvolution of medical ultrasound images using a parametric model for the point spread function. In *2016 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4. IEEE, 2016.

Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Tai-Xiang Jiang, Tian-Hui Ma, and Teng-Yu Ji. Mixed noise removal in hyperspectral image via low-fibered-rank regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1):734–749, 2019.

Yue Zhong, Lizhuang Liu, Dan Zhao, and Hongyang Li. A generative adversarial network for image denoising. *Multimedia Tools and Applications*, 79(23):16517–16529, 2020.

Long Zhou, Joshua D Schaefferkoetter, Ivan WK Tham, Gang Huang, and Jianhua Yan. Supervised learning with cyclegan for low-dose fdg pet image denoising. *Medical image analysis*, 65:101770, 2020.

Lina Zhuang and Michael K Ng. Fasthymix: Fast and parameter-free hyperspectral image mixed noise removal. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8):4702–4716, 2021.

Lina Zhuang, Michael K Ng, and Xiyou Fu. Hyperspectral image mixed noise removal using subspace representation and deep cnn image prior. *Remote Sensing*, 13(20):4098, 2021.