Speckle Noise Suppression in Digital Images Utilizing Deep Refinement Network

Ehab Alaa Saleh[1], **Mostafa I. Soliman**[1,2], **Mohamed Abdel-Nasser**[2,✉]

[1]Electrical Engineering Department, Aswan University, Egypt
[2]Department of Computer Science and Engineering, Egypt–Japan University of Science and Technology, Egypt
mohamed.abdelnasser@aswu.edu.eg✉

Abstract

This paper proposes a deep learning model for speckle noise suppression in digital images. The model consists of two interconnected networks: the first network focuses on the initial suppression of speckle noise. The second network refines these features, capturing more complex patterns, and preserving the texture details of the input images. The performance of the proposed model is evaluated with different backbones for the two networks: ResNet-18, ResNet-50, and SENet-154. Experimental results on two datasets, the Boss steganography, and COVIDx CXR-3, demonstrate that the proposed method yields competitive despeckling results. The proposed model with the SENet-154 encoder achieves PSNR and SNR values higher than 37 dB with the two datasets and outperforms other state-of-the-art methods (Pixel2Pixel, DiscoGAN, and BicycleGAN).

Keywords: *Speckle Noise, Image Filtering, Image Despecking, Denoising, Image Enhancement, Deep Learning.*

Received: 04 December 2023
Accepted: 31 January 2024
Online: 09 February 2024
Published: 15 July 2024

1 Introduction

The quality of digital images may degrade due to speckle noise (salt-and-pepper noise), caused by various factors during image acquisition and transmission. Speckle noise can be present in a variety of digital images from different domains, including the medical imaging domain (X-ray, ultrasound, magnetic resonance imaging (MRI)), and the remote sensing domain (e.g., satellite images). In such images, speckle noise manifests as pixels with low and high intensity distributed at random locations. Various speckle noise suppression techniques have been proposed in the literature to effectively remove speckle noise and enhance the analysis of medical and remote sensing images, such as synthetic aperture radar (SAR) images [1].

The techniques for reducing speckle noise include both traditional and deep learning-based methods. Traditional methods involve the use of filtering techniques in both spatial and spectral domains. Common filtering techniques include median filters, Wiener filters, and wavelet-based filters [12]. On the other hand, numerous deep learning-based methods have been proposed for effectively suppressing speckle noise in digital images [11, 6]. The study presented in [1] concluded that deep learning-based speckle noise suppression methods outperform traditional methods.

In recent years, various deep learning-based methods have been proposed for speckle noise suppression in digital images. For example, in [4], Karaoğlu et al. conducted a comparative analysis of five deep learning models to suppress speckle noise in ultrasound images. The evaluated models include CNN Residual Network, Dilated Convolution Autoencoder Denoising Network (D-U-NET), Generative Adversarial Denoising Network (DGAN-Net), Denoising U-Shaped Net, and BatchRenormalization U-Net. According to the study findings, the most effective results were achieved by D-U-NET and DGAN-Net, exhibiting a peak signal-to-noise ratio (PSNR) exceeding 33 dB and a structural similarity index (SSIM) surpassing 0.98.

Furthermore, Wen et al. [14] introduced a self-attention multi-scale convolutional neural network named SAMSCNN to mitigate speckle noise in SAR images. Simulation results demonstrated that SAMSCNN achieves an average PSNR of less than 28 dB and an average SSIM of less than 0.9. There is still room for improvement in the PSNR and SSIM results of existing speckle noise reduction methods.

This paper introduces a deep learning-based model for speckle noise reduction in digital images. The proposed model comprises two interconnected autoencoder networks. The first autoencoder network focuses on the initial suppression of speckle noise, extracting low-level features from the input images. The second autoencoder then refines these features, capturing more complex patterns and preserving crucial texture details. This hierarchical approach allows the model to learn both global and fine-grained features in a step-wise manner, enhancing its ability to discriminate between noise and meaningful image structures. The performance of the proposed method is assessed with different backbones for the autoencoder networks: ResNet-18, ResNet-50, and SENet-154. The method is evaluated on two datasets, including medical and real-world digital images (Boss steganography and COVIDx CXR-3). Quantitative results demonstrate that the proposed method achieves effective speckle noise re-

duction, as indicated by five different metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), PSNR, Signal-to-Noise Ratio (SNR), and Perceptual Average Mean Squared Error (PAMSE). Qualitative results are also provided to illustrate that the proposed method maintains a balance between speckle noise reduction and preserving the texture details of the digital images.

The other sections of this paper include the problem formulation (Section 2), the proposed method (Section 3), the results and discussion (Section 4), and the conclusion and future work (Section 5).

2 Problem Formulation

Consider $X \in \mathbb{X}$ as the input digital image, and $Y \in \mathbb{Y}$ as the corresponding cleaned image (despeckled image). The problem of speckle noise suppression in digital images can be formally defined as a function $f : \mathbb{X} \rightarrow \mathbb{Y}$ that maps elements from the domain $\mathbb{X}$ (images having speckle noise) to elements in its co-domain $\mathbb{Y}$ (i.e., despeckled image or cleaned images). This study introduces a deep learning-based system to address this problem.

Specifically, the proposed system comprises two sequential yet interconnected autoencoder networks, denoted as $N_s(X)$ and $N_c(\hat{Y})$, where $\hat{Y}$ represents the predicted image (or cleaned image) generated by N_s. The operation of the proposed model can be expressed as follows:

$$\hat{X} = N_s(X), \tag{1}$$

$$\hat{Y} = N_c(\hat{X}), \tag{2}$$

In these expressions, X is the input image, $\hat{X}$ is the filtered image given by the first autoencoder network, N_{θ_X}, and $\hat{Y}$ is the refined despeckled image produced by the second network N_{θ_Y}.

Loss functions are employed to compare the digital image reconstructed by $N_s(X)$ and $N_c(\hat{Y})$ with the ground truth. This loss function quantifies the discrepancy between the predicted image and the ground truth image, enabling the training process to minimize this discrepancy.

3 Proposed Method

Figure 1 illustrates the main components of the proposed speckle noise reduction model, consisting of two sequential yet interconnected autoencoder networks. Such two-stage architecture enables multi-stage refinement of the denoising process. The initial denoising in the first network provides an initial despeckled image, which serves as a more informative input for the subsequent network. The second network can then focus on further refining the despeckling results based on the already improved representation, contributing to the preservation of finer details and avoiding over-smoothing.

During the training process, the predicted digital image from the first network is fed into the second network. The second network focuses on removing more challenging noise from the initially predicted image and enhancing the texture details of the images. To facilitate learning and establish relationships between the input and output domains, two loss functions are utilized for each network within the proposed model.

In this study, we adopt the concept of speckle noise reduction as an image-to-image translation task, as proposed in [5, 17]. The method proposed in [17] involves two generators and one discriminator, while the method proposed in [5] employs two generators and two discriminators. In the proposed model, we employ two networks with two loss functions to suppress speckle noise in digital images.

Indeed, the proposed method offers several advantages over single-stage speckle noise suppression methods. Firstly, the proposed method enables more efficient training of the networks. The first network can be trained on a large dataset of noisy images, while the second network can be trained on a smaller dataset of images that have already been denoised by the global network. Secondly, it facilitates superior noise removal performance, with the first network effectively removing most of the simple noise from the image, and the second network further addressing any remaining challenging noise. Lastly, the proposed method is more versatile and applicable to a broader range of image denoising tasks. For instance, the proposed two-stage approach can be employed to denoise images affected by various types of noise, including Gaussian noise, salt-and-pepper noise, and speckle noise.

3.1 Model Architecture

This subsection details the architecture of both the first network, $N_s(X)$, and the second network, $N_c(\hat{Y})$, both of which share an identical structure. The two networks comprise encoder and decoder layers. The encoder network's role is to extract features from the input digital image using convolutional filters with down-sampling. Conversely, the decoder network employs deconvolution filters to up-sample the feature maps, ultimately generating a speckle noise-free digital image. Skip connections are incorporated to preserve spatial information and anatomical structures within the images.

The decoder of each network has four convolutional layers, while the decoder of each network has four deconvolution layers. Spatial filters in each convolution and deconvolution layer have a size of 3×3, facilitating down-sampling and up-sampling of the feature maps with a stride of 2×2. Batch normalization (BN) is applied after each deconvolutional layer, and the LeakyReLU activation function with a slope of 0.2 is employed at the end of each deconvolutional layer. The $Tanh$ activation function is utilized in the last convolutional layer of the encoder, while the *sigmoid* activation function is applied in the last layer of the decoder.

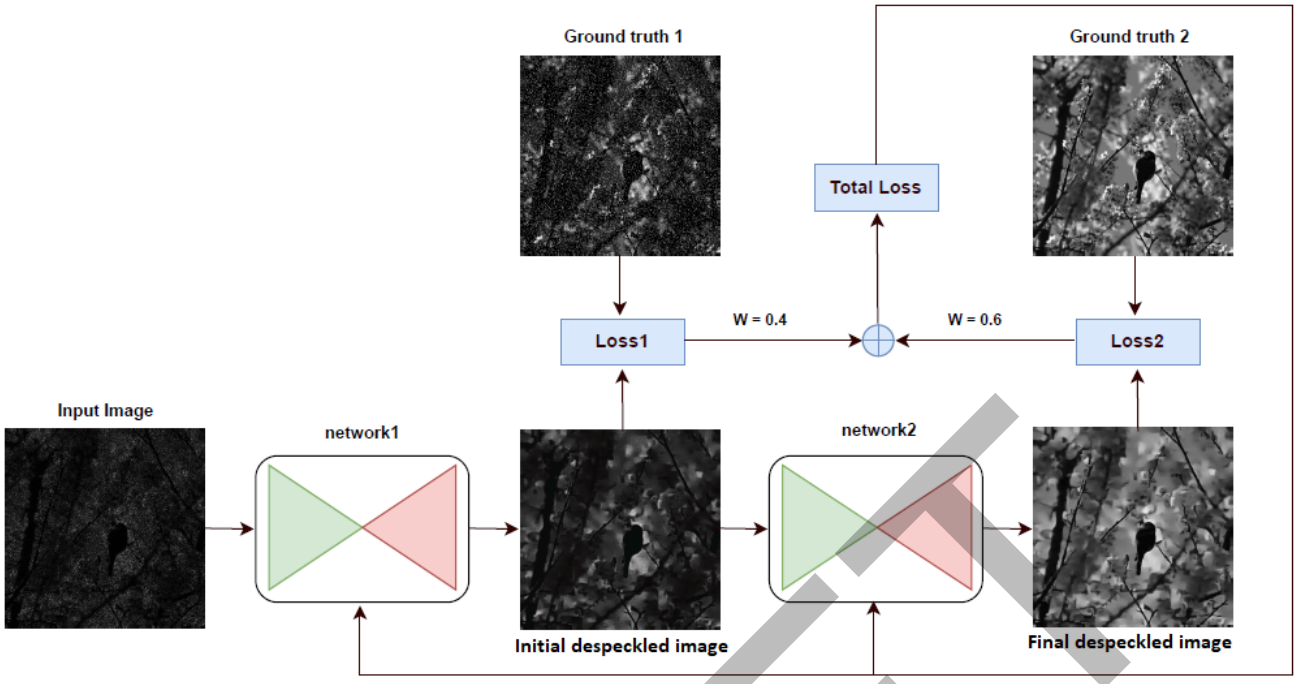

Figure 1: The main components of the proposed speckle noise suppression model.

3.2 Loss Function

In this study, we employ two loss functions to evaluate the network's performance in the training process with respect to the structural similarity between the predicted image (cleaned image) and the ground truth image. The first loss function is the mean square error (MSE) loss, which measures the average squared difference between the ground truth image $GT1_i$ (ground truth 1) and the image generated by the first network, $\hat{X}$. The MSE loss function can be defined as follows:

$$L_{MSE}(GT1, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(GT1_i - \hat{X}_i)^2, \quad (3)$$

The second loss function utilized in the training process is the $L1$ loss, also known as the absolute error. This loss calculates the mean of the absolute differences between the true values and the predicted values. The goal is to minimize the error by reducing the sum of these absolute differences between the ground truth image $GT2_i$ (ground truth 2) and the image generated by the second network, $\hat{Y}$. The $L1$ loss can be expressed as follows:

$$L_{L1}(GT2, \hat{Y}) = \sum_{i=1}^{n}|GT2_i - \hat{Y}_i|, \quad (4)$$

The overall loss function $\mathfrak{L}(GT, \hat{Y})$ of the proposed method can be expressed as follows:

$$\mathfrak{L}(GT, \hat{Y}) = \alpha L_{MSE} + (1-\alpha)L_{L1}, \quad (5)$$

where α is a hyperparameter weighting the importance of the training loss functions. In our experiments, we empirically set $\alpha = 0.4$.

3.3 Generating Ground Truth

In this study, we implement a training approach for the first network, $N_s(X)$, by generating ground truth samples (GT1). Speckle noise is introduced to the original ground truth in the training datasets. Examples of the generated ground truth images from the datasets used in our study are illustrated in Figure 2 and Figure 3. The first row displays the input noisy images, the second row shows the synthesized ground truth images (GT1), and the third row presents the actual ground truth images (i.e., GT2).

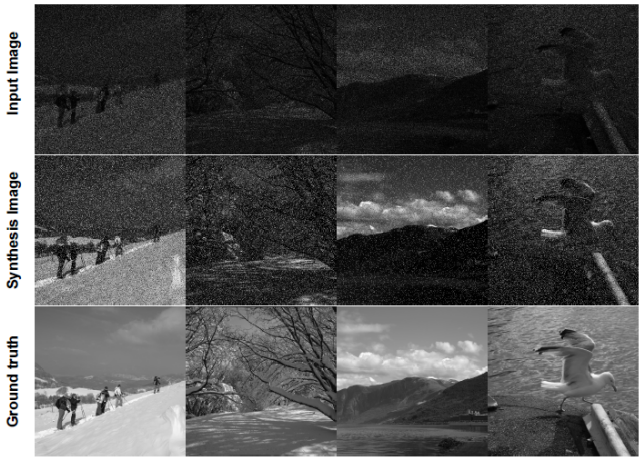

Figure 2: Examples from the Boss steganography dataset: For each image, Row 1 displays the input noisy images, Row 2 presents the synthesized ground truth images (GT1), and Row 3 shows the actual ground truth images (i.e., GT2).

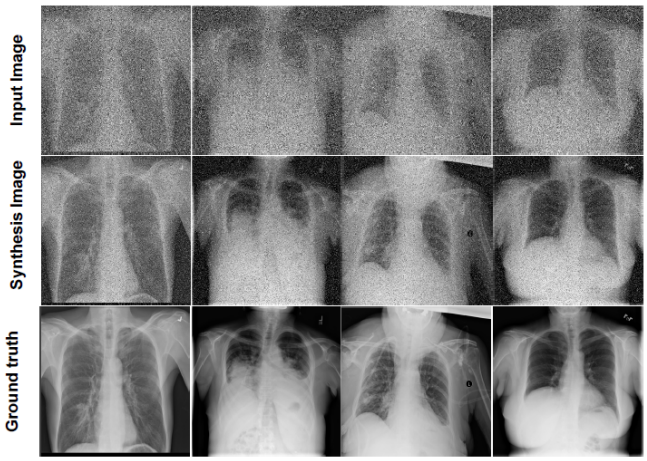

Figure 3: Examples from the COVIDx CXR-3 dataset: For each image, Row 1 displays the input noisy images, Row 2 presents the synthesized ground truth images (GT1), and Row 3 shows the actual ground truth images (i.e., GT2).

3.4 Evaluation Metrics

Five evaluation metrics are used in this study to evaluate the proposed speckle noise reduction method: Mean Absolute Error (MAE) [15], Mean Squared Error (MSE) [7], PSNR [2], Signal-to-Noise Ratio (SNR) [13], and Perceptual Average Mean Squared Error (PAMSE) [16].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|GT_{(i)} - \hat{Y}_{(i)}\right|, \tag{6}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(GT_{(i)} - \hat{Y}_{(i)}\right)^2, \tag{7}$$

$$PSNR = 20 \cdot \log_{10}\left(\frac{\max(\max(\hat{Y}))}{(PSNR_{MSE})^{0.5}}\right), \tag{8}$$

$$PSNR_{MSE} = \frac{1}{m \cdot n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|GT(i,j) - \hat{Y}(i,j)\right\|^2, \tag{9}$$

$$SNR = 100 \cdot \log_{10}\left(\frac{\alpha^2}{\alpha c^2}\right), \tag{10}$$

$$PAMSE(X,\hat{Y}) = \frac{1}{n}\left\|h \otimes (X,\hat{Y})\right\|^2, \tag{11}$$

In these expressions, n stands for the total number of images. In (6), (7), (8), and (9), GT represents the ground truth image, and $\hat{Y}$ is the corresponding despeckled image produced by the proposed model. In equation (10), α^2 represents the variance of the input image, and αc^2 stands for the variance of the predicted image. In (11), $\otimes$ stands for the convolution operator.

4 Experiments and Results

This section provides a description of the experiments conducted to evaluate the proposed model, including details about the datasets used and the evaluation metrics employed in the experiments.

4.1 Datasets

A series of experiments were conducted to assess the performance of the proposed model on two datasets, including both real-world and medical images. The two datasets are the Boss steganography dataset and the COVIDx CXR-3 dataset.

The Boss steganography dataset [1] comprises 3000 images. The dataset was randomly partitioned: 80% for training and 20% for testing. All images were resized from 512×512 to 320×320.

The COVIDx CXR-3 dataset [9] is a large-scale benchmark dataset for CXR (Chest X-Ray) images. The dataset comprises 30,386 CXR images obtained from a multinational cohort of 17,026 patients from at least 51 countries. It consists of 29,986 images for the training set and 400 images for the testing set. In our experiments, all images in COVIDx CXR-3 were resized from 1024×1024 to 320×320.

4.2 Parameter Settings

During training, we employed the Adam optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and an initial learning rate of 0.0001. The optimal combination of parameters was achieved with a batch size of 2 and 100 epochs.

All experiments were conducted on a system with a 64-bit Core i7-6700, 3.40GHz CPU, 16GB of memory, and an NVIDIA GTX 1080 GPU. We used the Ubuntu 16.04 operating system and the PyTorch deep learning framework [8]. The training process of the proposed method required approximately 20 minutes per epoch with a batch size of 2, and the online prediction time averaged around 0.0198 seconds.

The implementation of the proposed method will be available at https://github.com/egnaser/Speckle-Noise-Suppression-MENDEL.

4.3 Results and Discussion

First, we investigate the impact of different loss functions on the performance of the proposed speckle noise reduction model. Three loss combinations are studied: MSE, L1, and MSE+L1. In the first loss combination, MSE is used as a loss function for both Network 1 and Network 2 of the proposed model. In the second loss combination, L1 is used as a loss function for both networks. In the third loss combination, MSE is used for Network 1, and L1 is used for Network 2. In these experiments, the SENet-154 encoder was employed in both networks. As presented in Table 1, the first loss

[1] https://rb.gy/hcpaoc

Table 1: Analyzing the performance of the proposed model with different loss functions.

Loss	MAE ↓	MSE ↓	SNR ↑	PSNR ↑	PAMSE ↑
MSE	0.046551	0.002167	33.19 ± 3.7	36.05 ± 1.5	0.129648
L1	0.0497795	00.002478	33.32 ± 4.6	35.79 ± 1.7	0.130156
MSE + L1	0.0404104	0.001633	33.97 ± 3.1	36.57 ± 1.1	0.132695

combination, MSE, outperforms the second loss combination in terms of the five evaluation metrics. The best performance is achieved with the MSE+L1 loss combination, yielding a PSNR of 36.57 ± 1.1.

Second, we explore the performance of the proposed model with different encoder networks: ResNet-18, ResNet-50, SENet-154. We also compare the proposed model's performance with three state-of-the-art deep learning architectures: Pixel2Pixel [3], DiscoGAN [5], and BicycleGAN [18]. These three methods were trained to suppress speckle noise using the same datasets used to train the proposed model (the Boss steganography and COVIDx CXR-3 datasets).

Table 2 presents the performance of the proposed model with the ResNet-18 encoder. It is essential to note that the same ResNet-18 encoder is used in all models presented in this table. With the Boss steganography dataset, the proposed model obtains SNR and PSNR values higher than 36 dB. Pixel2Pixel and BicycleGAN models yield lower results with SNR values less than 27 dB. DiscoGAN produces a MAE of 0.132, significantly higher than that of the proposed model. With the COVIDx CXR-3 dataset, the proposed model achieves a PSNR of 37.87 ± 0.1, surpassing the other three models.

Table 3 presents the performance of the proposed model with the ResNet-50 encoder, also used in other models. With both datasets, the proposed model achieves MAE values less than 0.08 and PSNR values higher than 37 dB. The MAE of DiscoGAN and BicycleGAN models is higher than that of other models. Pixel2Pixel obtains the lowest SNR values with both datasets.

As shown in Table 4, the proposed model with the SENet-154 encoder achieves PSNR and SNR values higher than 37 dB with the two datasets. The proposed model has an MSE error of 0.000965, significantly lower than those of Pixel2Pixel, DiscoGAN, and BicycleGAN models.

Tables 2, 3, 4 indicate that the SENet-154 encoder yields the best results with the Boss steganography and COVIDx CXR-3 datasets. Moreover, the proposed model with the SENet-154 encoder outperforms the other compared methods (Pixel2Pixel [3], DiscoGAN [5], and BicycleGAN [18]) using the same settings.

Figure 4 illustrates the despeckling results of the first network in the proposed model (i.e., network1) and the final despeckled images provided by the second network (i.e., network2). All examples are randomly picked up from the Boss steganography dataset. Despite the input images exhibiting dense speckle noise, the despeckling results of the first network in the proposed model (column 3) closely approximate the ground truth images. Notably, the proposed model produces final despeckling images (column 4) that preserve the discontinuities and small details of the objects, closely aligning with the ground truth images. This demonstrates that the two-stage architecture is more adaptable to different types and levels of noise. The first network can specialize in handling specific characteristics of noise, while the second network can adapt to the specific noise characteristics that may persist after the initial denoising stage.

For a qualitative assessment of the Boss steganography dataset, Figure 5 showcases the predicted images generated by the proposed model using the SENet-154 encoder alongside three state-of-the-art methods (Pixel2Pixel [3], DiscoGAN [5], and BicycleGAN [18]). It is evident that our proposed model excels in producing more accurate despeckled images compared to the other models.

5 Conclusion

This study has introduced a deep learning-based model designed for speckle noise reduction in digital images, utilizing two interconnected autoencoder networks. The first network focuses on suppressing speckle noise, while the subsequent network refines despeckling results, preserving essential texture details in the input images. Evaluations were conducted on two diverse datasets, encompassing medical and real-world digital images (Boss steganography and COVIDx CXR-3). Various combinations of loss functions and backbone encoders (ResNet-18, ResNet-50, and SENet-154) were explored for the two networks. The results demonstrate that SENet-154, combined with the MSE+L1 loss, yields the most effective despeckling outcomes. The proposed model consistently achieves PSNR and SNR values exceeding 37 dB with the two datasets, outperforming other state-of-the-art deep learning architectures trained for despeckling digital images (Pixel2Pixel, DiscoGAN, and BicycleGAN) with an MSE lower than 0.001. Qualitative assessments highlight the proposed method's ability to maintain the balance between speckle noise reduction and preserving crucial texture details in digital images.

Future work will involve exploring additional refinements for the proposed speckle noise suppression method, assessing its scalability to larger datasets, and considering potential adaptations for real-time image despeckling.Additionally, the proposed method will incorporate the pointwise convolutions with parallel

Table 2: Quantitative results of the proposed model with the ResNet-18 encoder and the three state-of-art methods (Pixel2Pixel [3], DiscoGAN [5], and BicycleGAN [18]).

Dataset	Methods	MAE ↓	MSE ↓	PSNR ↑	SNR ↑	PAMSE ↑
Boss steganography	Pixel2Pixel [3]	0.0963327	0.00928	36.31 ± 1.3	26.3 ± 9.0	0.14183
	DiscoGAN [5]	0.1327026	0.01761	36.31 ± 1.3	35.66 ± 1.8	0.14183
	BicycleGAN [18]	0.1458423	0.02127	34.62 ± 2.6	26.95 ± 8.5	0.13523
	Proposed	0.086133	0.007387	36.83 ± 0.9	36.31 ± 1.3	0.14386
COVIDx CXR-3	Pixel2Pixel [3]	0.0766159	0.00965	36.83 ± 0.9	27.6 ± 8.0	0.14386
	DiscoGAN [5]	0.0992673	0.009854	36.83 ± 0.9	36.7 ± 1.0	0.14386
	BicycleGAN [18]	0.0993478	0.00987	37.74 ± 0.2	28.38 ± 7.4	0.14742
	Proposed	0.083366	0.00587	37.87 ± 0.1	36.83 ± 0.9	0.14792

Table 3: Quantitative results of the proposed model with the ResNet-50 encoder and the three state-of-art methods (Pixel2Pixel [3], DiscoGAN [5], and BicycleGAN [18]).

Dataset	Methods	MAE ↓	MSE ↓	PSNR ↑	SNR ↑	PAMSE ↑
Boss steganography	Pixel2Pixel [3]	0.093359	0.008716	37.09 ± 0.7	26.04 ± 9.2	0.14488
	DiscoGAN [5]	0.101079	0.010217	36.83 ± 0.9	36.7 ± 1.0	0.14386
	BicycleGAN [18]	0.111085	0.012340	37.09 ± 0.7	37.09 ± 0.7	0.14488
	Proposed	0.073212	0.005360	37.87 ± 0.1	37.22 ± 0.6	0.14792
COVIDx CXR-3	Pixel2Pixel [3]	0.084409	0.007125	37.22 ± 0.6	28.51 ± 7.3	0.14539
	DiscoGAN [5]	0.092412	0.00854	36.96 ± 0.8	36.83 ± 0.9	0.14437
	BicycleGAN [18]	0.0978263	0.00957	37.09 ± 0.7	37.22 ± 0.6	0.14488
	Proposed	0.0621128	0.003858	37.35 ± 0.5	37.48 ± 0.4	0.14589

Table 4: Quantitative results of the proposed model with the SENet-154 encoder and the three state-of-art methods (Pixel2Pixel [3], DiscoGAN [5], and BicycleGAN [18]).

Dataset	Methods	MAE ↓	MSE ↓	PSNR ↑	SNR ↑	PAMSE ↑
Boss steganography	Pixel2Pixel [3]	0.063576	0.004042	37.48 ± 0.4	37.74 ± 0.2	0.14640
	DiscoGAN [5]	0.094398	0.008911	37.61 ± 0.3	37.87 ± 0.1	0.146914
	BicycleGAN [18]	0.1763065	0.031084	37.48 ± 0.4	37.74 ± 0.2	0.146406
	Proposed	0.04041	0.001633	37.61 ± 0.3	37.87 ± 0.1	0.146914
COVIDx CXR-3	Pixel2Pixel [3]	0.0507937	0.00258	37.74 ± 0.2	37.87 ± 0.1	0.147426
	DiscoGAN [5]	0.08655	0.007491	37.74 ± 0.2	37.87 ± 0.1	0.147421
	BicycleGAN [18]	0.124442	0.015486	37.74 ± 0.2	37.74 ± 0.2	0.147422
	Proposed	0.0310644	0.000965	37.74 ± 0.2	37.87 ± 0.1	0.147428

branching approach suggested in [10] to further enhance the results of speckle noise suppression.

References

[1] Baraha, S., Sahoo, A. K., and Modalavalasa, S. A systematic review on recent developments in nonlocal and variational methods for sar image despeckling. *Signal Processing 196* (2022), 108521.

[2] Huynh-Thu, Q., and Ghanbari, M. Scope of validity of psnr in image/video quality assessment. *Electronics letters 44*, 13 (2008), 800–801.

[3] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134.

[4] Karaoğlu, O., Bilge, H. Ş., and Uluer, İ. Removal of speckle noises from ultrasound images using five different deep learning networks. *Engineering Science and Technology, an International Journal 29* (2022), 101030.

[5] Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J. Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning* (2017), PMLR, pp. 1857–1865.

[6] Li, X., Wang, Y., Zhao, Y., and Wei, Y. Fast speckle noise suppression algorithm in breast ultrasound image using three-dimensional deep learning. *Frontiers in Physiology 13* (2022), 698.

[7] Marmolin, H. Subjective mse measures. *IEEE transactions on systems, man, and cybernetics 16*, 3 (1986), 486–489.

[8] Paszke, A., Gross, S., Chintala, S., and Chanan, G. Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration, 2017.

[9] Pavlova, M., Tuinstra, T., Aboutalebi, H., Zhao, A., Gunraj, H., and Wong, A. Covidx cxr-3: A large-scale, open-source bench-

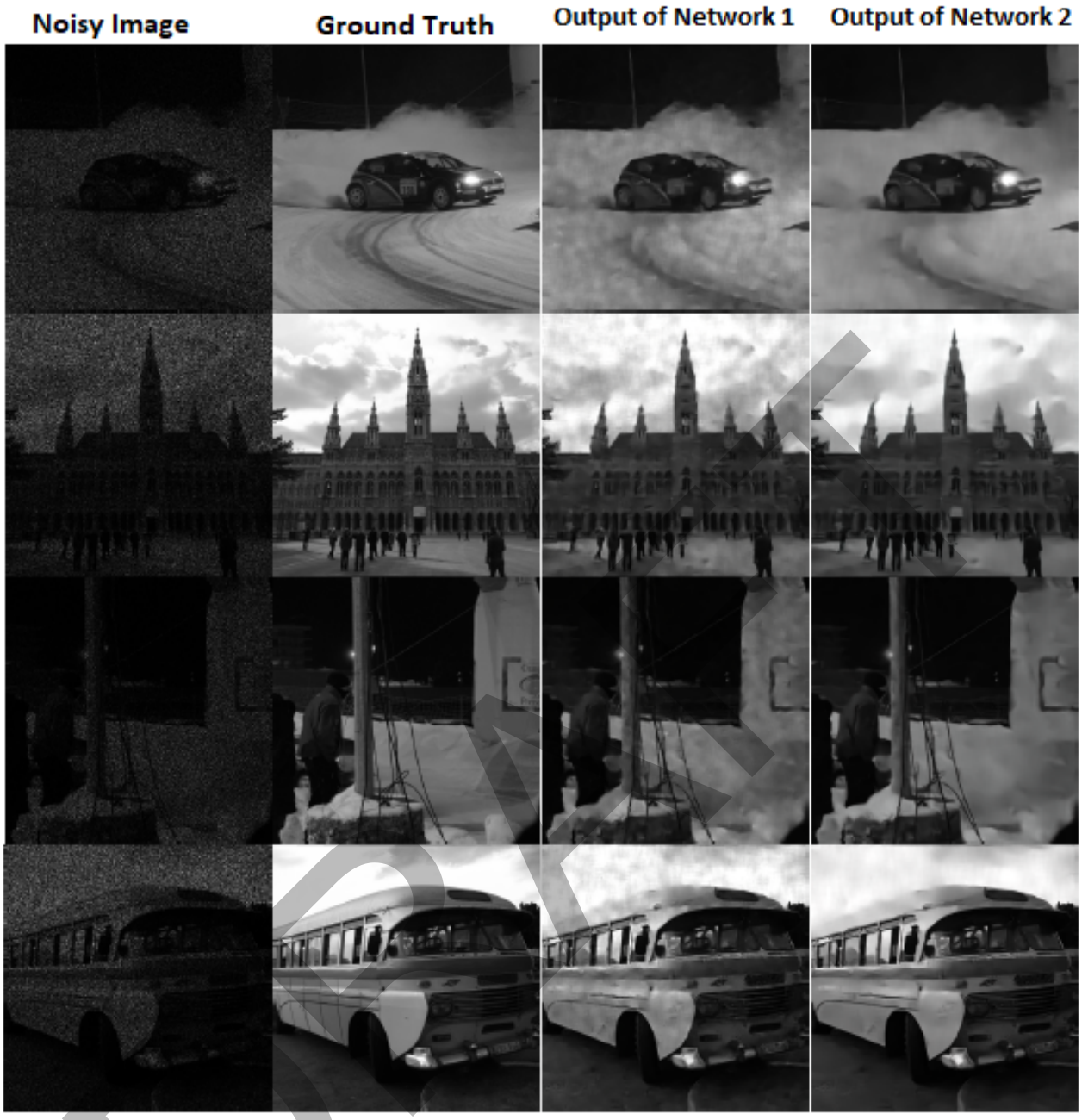

Figure 4: Examples of the despeckling results from the proposed model. The first column presents the input noisy images, the second column shows the ground truth, the third column displays the despeckling results of the first network of the proposed model (i.e., network1), and the fourth column exhibits the final despeckling results provided by the second network of the proposed model (i.e., network2).

mark dataset of chest x-ray images for computer-aided covid-19 diagnostics. *arXiv preprint arXiv:2206.03671* (2022).

[10] Schuler, J. P. S., Romani, S., Abdel-Nasser, M., Rashwan, H., and Puig, D. Grouped pointwise convolutions reduce parameters in convolutional neural networks. *Mendel 28*, 1 (2022), 23–31.

[11] Shukla, A. K., Dwivedi, S. K., Chandra, G., and Shree, R. Deep learning-based suppression of speckle-noise in synthetic aperture radar (sar) images: A comprehensive review. In *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2* (2023), Springer, pp. 693–705.

[12] Sudha, S., Suresh, G., and Sukanesh, R. Speckle noise reduction in ultrasound images by wavelet thresholding based on weighted variance. *International journal of computer theory and engineering 1*, 1 (2009), 7.

[13] Tandra, R., and Sahai, A. Snr walls for signal detection. *IEEE Journal of selected topics in Signal Processing 2*, 1 (2008), 4–17.

Figure 5: Examples from the test results with the Boss steganography dataset using the Pixel2Pixel model, DiscoGAN model, BicycleGAN model, and Our model. For each image, we display (column 1) the input image, (column 2) the ground truth, (column 3) the output for the Pixel2Pixel model, (column 4) the output for the DiscoGAN model, (column 5) the output for the BicycleGAN model, and (column 6) the final output with our model.

[14] Wen, Z., He, Y., Yao, S., Yang, W., and Zhang, L. A self-attention multi-scale convolutional neural network method for sar image despeckling. *International Journal of Remote Sensing 44*, 3 (2023), 902–923.

[15] Willmott, C. J., and Matsuura, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research 30*, 1 (2005), 79–82.

[16] Xue, W., Mou, X., Zhang, L., and Feng, X. Perceptual fidelity aware mean squared error. In *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 705–712.

[17] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint* (2017).

[18] Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. Toward multimodal image-to-image translation. *Advances in neural information processing systems 30* (2017).