

Improving image processing and noise removal using generative deep learning techniques

Firas Abdulrahman Yousif¹, Shamil Kh. Ramadhan², Ahmed A. Mostfa^{1,3}

¹Department of Computer Science, University of Al Hamdaniya, Mosul, Iraq

²Department of Medical Instrumentation Techniques, Northern Technical University, Nineveh, Iraq

³Department of Computer Center Service, Ninevah University, Mosul, Iraq

Article Info

Article history:

Received Apr 19, 2025

Revised Jul 28, 2026

Accepted Aug 12, 2026

Keywords:

Generative adversarial network

Generative learning

Image processing

Noise signal removal

Perceptual loss

U-Net

ABSTRACT

Denosing plays a crucial role in satellite and medical images. The classical methods struggle to find out the noisy values and might lead to some loss in image metadata's essitals. In this study, a generative adversarial network (GAN) is introduced along with the U-Net architecture. GAN algorithm has the capability to provide a powerful estimated fine image compared with a regular neural network approach. The Modified National Institute of Standards and Technology (MNIST) dataset is injected with synthetic random noise. The proposed work demonstrated noise reduction capabilities and significant improvement in peak signal to noise ratio (PSNR), from 11.70 dB for noise image to 19.41 dB for the estimated image. The structural similarity index (SSIM) moved from 0.61 to 0.93, showing a big restoration with stable convergence for the image details. The GAN visual analysis based on deep learning showed effective output detail recovery.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ahmed A. Mostfa

Department of Computer Science, University of Al Hamdaniya

Mosul, Iraq

Email: mostfa@uohamdaniya.edu.iq

1. INTRODUCTION

The renaissance in image processing related with deep learning approaches has considered the transformation era to enhance the quality of visual data. Noise which considers the most distortion affects in images are arises due to signal transmission errors which casue a deterioration for the quality of the images [1]. In crticial application such as satellite remote sensing, the image metadata is crucial for interpretation and management and interoperability of the data [2]. A traditional method, such as medial filters, wavelet transformation could reduce the noise based on mathematical models. These approaches could find a difficulties to reduce noise and achive the same results or might overights the outputs [3].

For these reasons, scientist looked for another method to taggle the missdifued images and denosing tasks by using deep learning approaches. Generative adversarial network (GAN) preserves the image details by subsetting the dataset and learning directly from the data with remarkable accuracy estimations [4]. Autoencoders, which are neural networks designed for unsupervised learning, compress input images into latent representations and then reconstruct them [5]. During the GAN process, the dataset passes through two steps. The generator which produces a low noise image same as the real one, and discriminator which gives an assessment and correctness between the real and fake one. The discriminator is always try to push the generator to produce a more reliable and indistinguishable free of noise image [6]. The GAN network then uses a probabilistic method to enhance the process in an incremental steps, which introduce a high quality image recovery [7]. However, a one can see GANs application in many fields, such as medical image.

Where X-ray and computed tomography (CT) scan is the target image applications [8]. While in satellite, GAN helps to reduce noise for image in difficult weather conditions and enable a stable monitor for disaster assessments. Another type of GAN application can be found in photography, where it can help in restoring degraded images and the capabilities of real time camera surveillance for improved footages in critical conditions [9].

In addition to the aftermentioned applications, one the advantages that GANs approach could have is to learn the context features by adapting the denoising method based on the structure of the image. Such as preserving textures, patterns, and edges of the estimated image [10]. Building on this trend of algorithmic optimization, a new approach used Pix2Pix GAN model to Thomson scattering diagnostic images. This approach achieved noise removal across the tested samples as well as reducing the time of image processing [11], [12].

GAN can also handle a scalable dataset and work in efficient performance in both offline and real time streaming. There is a huge interest by researchers about the way of improving the robustness and efficiency of the GAN's models, by trying to reduce the computational requirements during the network's job. Furthermore, combining a hybrid method such as auto encoders with GAN algorithm is a great approach to push the denosing levels up to the limits. In conclude, incorporating the deep learning techniques with the image process became a revolution era to enhance and visual data with massive dataset to generate a more valuable context aware estimation for the output.

2. LITERATURE REVIEW

Extracting noise is the one of the most foundational approaches in computer vision. Where a lot of deep rearch involving aiming to underlying signals from the corrupted image. For those early methods, such as wavelet transforms and fourier analysis, are important strategies for specific noise tasks, but they struggling to matain complex edge structures and inaddition to their abilities in real-time processing the dataset. Therefore, a new shifting approach goes with deep learning algorithms. Such as neural network which goes through a deep architecture to learn the hierarchical of the image to extract the noise. Some of the convolutional neural network (CNN) rely on pixel-wise loss function, like mean square error (MSE) to falsy may lead to blurry reconstruction or overly smoothing the output estimation.

Similarly, Zou *et al.* [13] developed adversarial denoising CNNs for industrial applications, achieving high noise suppression with minimal data requirements. To address the limitations of pixel-wise optimization, researchers turned to GANs. GANs introduce an adversarial training mechanism where a generator network produces denoised images and a discriminator network evaluates their authenticity [14]. This competitive process encourages the generator to create outputs that are not only numerically accurate but also perceptually convincing. The effectiveness of this approach has been validated in numerous studies. For instance, Wang *et al.* [15] demonstrated that the residual structure and cooperative attention mechanism (RCA)-GAN architecture significantly enhanced noise suppression and detail preservation compared to traditional GAN.

However, there still a rearch limits observed, GAN could lead to unstable training performance in contrast with the transformer-based models that achieve a high performance and accuracy. A compelling framework that works on combine efficiency and structure fidelity such as U-Net is proposed. This study address GAN model along with U-Net integration. This method expecting to achieve a balance between high fifelity and outperforming against CNN denosiers. Deep learning approach such as GAN network has showed a transformation technique for image improvments [16], it has been known for its diverse application in noise removal. However, current studies may focus on specific dataset types which may limits the algorithm's capabilities [17], [18].

3. METHOD

For the following, we present the GAN method in order to provide an intense overview for the system architecture. Figure 1 shows the pipeline system process that starts from the noise input signal, going through the preprocessing and U-Net method. The hierarchy ends with the denoised output along with the evaluation values. Figure 1, as shown in is a detailed explanation of the methodology including mathematical formulations, experimental setups, and the overall workflow.

3.1. Dataset preparation and preprocessing

For preparing the dataset, we have used the fashion - Modified National Institute of Standards and Technology (MNIST) dataset. This is a suitable reference collection with about 70,000 grayscale image and serves as a good benchmark for value detail preservation. To train the model effectively, three

types of noise were synthetically corrupted to the clean images: Gaussian noise (with a standard deviation, $\sigma = 25$), Poisson noise, and salt-and-pepper noise (with a 5% density). High-resolution images with various types of noise (Gaussian, Poisson, salt-and-pepper) are included [19], and the dataset is augmented using transformations such as rotation, scaling, cropping and intensity adjustments to simulate real-world scenarios effectively [20], [21]. Mathematically, the augmentation process can be expressed as (1).

$$I' = T(I) = \alpha R(I) + \beta S(I) + \gamma C(I) \quad (1)$$

Where (I) represents the original image, (T) is the transformation function, ($R(I)$) is the rotation operation, ($S(I)$) is the scaling, and ($C(I)$) is the cropping. Parameters (α, β, γ) control the intensity of each transformation [22], [23]. Noise simulation is applied using additive or multiplicative noise models. For example, Gaussian noise is added using (2).

$$I_{\{noisy\}} = I + N(0, \sigma^2) \quad (2)$$

Where ($I_{\{noisy\}}$) is the noisy image. (I) is the clean image and ($N(0, \sigma^2)$) represents Gaussian noise with mean zero and variance (σ^2).

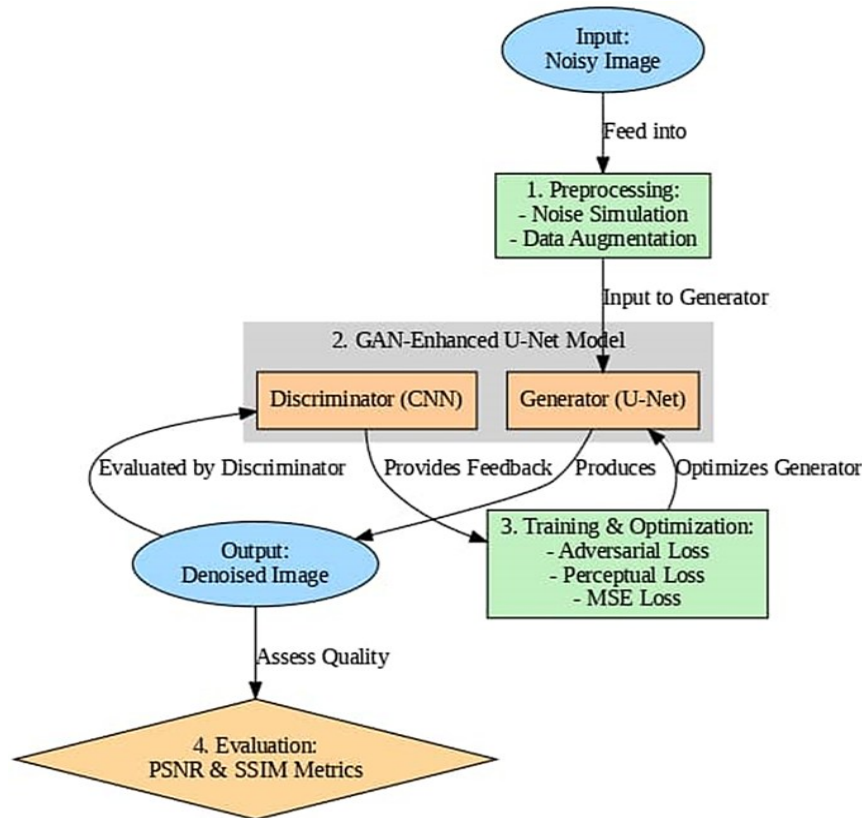


Figure 1. The proposed methodology pipeline

3.2. Proposed generative adversarial network-U-Net architecture

The study employs a U-Net-based architecture enhanced with GANs to achieve efficient denoising [24], [25] and the U-Net model consists of an encoder-decoder structure with skip connections that preserve spatial information and the encoder compresses the input image into a latent representation while the decoder reconstructs a denoised image [26]. Specifically, the architecture is broken down:

- The encoder: consists of 4 downsampling blocks. For every block, we used a 4×4 convolutional layers, and applied normalized and leaky rectified linear unit (ReLU) activation with slope 0.2 in order to capture the latent features.

- The decoder: in the section, we apply the transposed convolution upsampling, ensuring the recovery of the fine details lost while downsampling.
- The discriminator: utilize a PatchGAN classifier structure. Instead of classifying the whole image as real/fake, it classifies $N \times N$ Local patches, forcing the generator to focus on high-frequency texture sharpness. The detailed kernel sizes, strides, and padding configurations for all layers are summarized in Table 1.

Table 1. Detailed configuration of the generator and discriminator network architectures

Network	Layer	Kernel size	Stride	Padding	Output channels	Activation/norm
Generator (encoder)	Input	-	-	-	3	-
	Conv 1	4×4	2	1	64	LeakyReLU (0.2)
	Conv 2	4×4	2	1	128	BN + LeakyReLU (0.2)
	Conv 3	4×4	2	1	256	BN + LeakyReLU (0.2)
	Conv 4	4×4	2	1	512	BN + LeakyReLU (0.2)
Generator (decoder)	DeConv 1	4×4	2	1	256	BN + ReLU
	DeConv 2	4×4	2	1	128	BN + ReLU
	DeConv 3	4×4	2	1	64	BN + ReLU
	Output	4×4	2	1	3	Tanh
Discriminator (PatchGAN)	Conv 1	4×4	2	1	64	LeakyReLU (0.2)
	Conv 2	4×4	2	1	128	BN + LeakyReLU (0.2)
	Conv 3	4×4	2	1	256	BN + LeakyReLU (0.2)
	Output	4×4	1	1	1	Sigmoid

For the GAN setup, the generator (G) learns to produce denoised images, while the discriminator (D) evaluates the quality of these images and the adversarial loss for GANs is defined as (3).

$$\{L\}_{\{GAN\}} = \{E\}_{\{x \sim p_{\{data\}}(x)\}}[\log D(x)] + \{E\}_{\{z \sim p_z(z)\}}[\log(1 - D(G(z)))] \quad (3)$$

Where (x) denotes real data, (z) represents noise input, ($p_{\{data\}}(x)$) the real data distribution, and ($p_z(z)$) the noise distribution and in addition to adversarial loss, perceptual loss is integrated to enhance high-frequency detail preservation as shown (4).

$$\{L\}_{\{perc\}} = \sum_{i=1}^N \|\phi_{i(I_{\{denoised\}})} - \phi_{i(I_{\{ground\ truth\}})}\|_2^2 \quad (4)$$

Where ϕ_i is the features extracted from the (i)-th layer of the model and (N) represents the number of layers considered [25].

3.3. Loss functions and optimization

The training process employs the MSE loss and the adversarial loss to address the limitations of using MSE alone, which results in blurry images. We identify the total objective as a weighted sum of adversarial, pixel-wise, and perceptual losses as in (5).

$$L_{total} = \lambda^1 L_{GAN} + \lambda^2 L_{MSE} + \lambda^3 L_{perc} \quad (5)$$

L_{GAN} is the adversarial loss, which ensures the generated images are realistic and difficult for the discriminator to distinguish from real ones, where L_{MSE} is the pixel-wise MSE loss, which enforces structural consistency and fidelity between the generated image and the target image, where L_{perc} is the perceptual loss, which measures the difference in high-level feature representations (from a pre-trained network) to improve visual quality and semantic similarity. The parameters λ_1 , λ_2 , and λ_3 are hyperparameters that control the relative contribution of each loss term to the overall objective. The weights were set empirically to the following values: $\lambda_1 = 1$, $\lambda_2 = 10$, $\lambda_3 = 0.1$.

This configuration is chosen in order to have a balance between the different aspects in image quality. A significantly higher weight on L_{MSE} ($\lambda^2 = 10$) ensures that the model prioritizes structural consistency and accurate pixel-level reconstruction with the ground truth and adversarial and perceptual terms ($\lambda^1 = 1$ and $\lambda^3 = 0.1$) then act as refinements. The output results of the GAN have sharper textures and also it enhance the overall realism. However, the L_{perc} tries to improve the details and edge sharpness which give a more visual interests as a result.

3.4. Training protocol and implementation details

As mentioned that λ_1 and λ_2 are hyperparameters that act as the contributions of GAN loss and MSE loss. The optimization is performed using the Adam optimizer with a learning rate of 10^{-4} . ($\beta_1 = 0.9$) and ($\beta_2 = 0.999$). The GAN training steps involve multiple looping until it reaches convergence [27]. While the model evaluation is measured using peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [28], as in (6).

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (6)$$

Where MAX_I is the maximum possible pixel value and MSE error as shown in (7).

$$SSIM = \frac{\{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)\}}{\{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)\}} \quad (7)$$

Where (μ_x, μ_y) are the means, (σ_x, σ_y) the variances, and (σ_{xy}) the covariance of images (x) and (y). The output results are inspected using histogram equalization as a post-processing techniques as in [29]. In Table 2, we expose the difference aspect methods used in our case study in compare with the traditional methods.

Table 2. Comparative analysis of proposed and traditional image denoising methods

Aspect	Proposed method	Traditional methods
Architecture	U-Net with GAN	CNN-based denoisers
Loss function	GAN + Perceptual + MSE	MSE only
Optimization algorithm	Adam ($10 - 4 * 10^{-4}$)	SGD
Noise types handled	Gaussian, poisson, salt-and-pepper	Gaussian only
Evaluation metrics	PSNR and SSIM	PSNR

The deep learning method shows outperforming conventional approach with high quality denosing the output image in compared with the other algorithms. The model performed a 50 epochs along with a batch size equal 128, and the model is measured using (6) and (7).

- i) Generator (U-Net architecture): for the generator part, it is based on U-Net build model, in which it has a strong encoder and decoder structure. Where the first term contains a sequence of down sampling blocks in order to capture the contextual information. While the second term (decoder) works on transpose convolution for up sampling. The layers between the two methods are interconnected using skip connections. These connections are important enabling the reconstruction of the estimated image.
- ii) Discriminator architecture: the discriminator works as a binary classifier, where the image goes through a multiple setpes of convolutional layers to determine if the image is real or fake denoised output.
- iii) Loss functions and optimization: to achieve a balance between pixel-level accuracy and perceptual realism, the generator is optimized using a hybrid loss function.
 - Loss-adversarial (LGAN): works on the competitive training process as defined in (3).
 - Loss-perceptual (L Perc): enhances high-frequency preservation by looking for the feature maps from a pre-trained visual geometry group (VGG) network, as defined in (4).
 - Loss- mean square error (LMSE): ensures structural stability and works on minimizing the pixel-wise differences between the denoised and original clean images.

The total loss function for the generator is a weighted combination of these components as in (5). Where λ_1 , λ_2 and λ_3 are hyperparameters that balance the influence of each loss term. Optimization is performed using the Adam optimizer with a learning rate of 10^{-4} and beta parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$.

4. RESULTS AND DISCUSSION

In this paper, a GAN is used to denoise and improve the quality of the image. The results are analyzed as follows, focusing on the technical aspects of each form and its practical results: Figure 2 shows the comparison the original (clean) input images and the corrupted images. The first row observes the ground truth images from the fashion-MNIST dataset, which express fine details, patterns, and object boundaries. In sharp contrast, the bottom row shows the effect of the added noise, which was generated using a normal distribution Gaussian noise. The outcome simulated degradation significantly blurs boundaries and obscures key visual features, hindering feature extraction and object recognition. This figure establishes the necessary baseline to visually measure the effectiveness of the proposed denoising model.

Figure 3 represents the loss curve of the model during training and is a basic indicator used to evaluate the performance of the GAN. The curve shows a gradual decrease in the loss value with increasing number of epochs. Which reflects the improved performance of the generator in accurately reconstructing the original images, and the improved performance of the discriminator in distinguishing between real images and produced images.

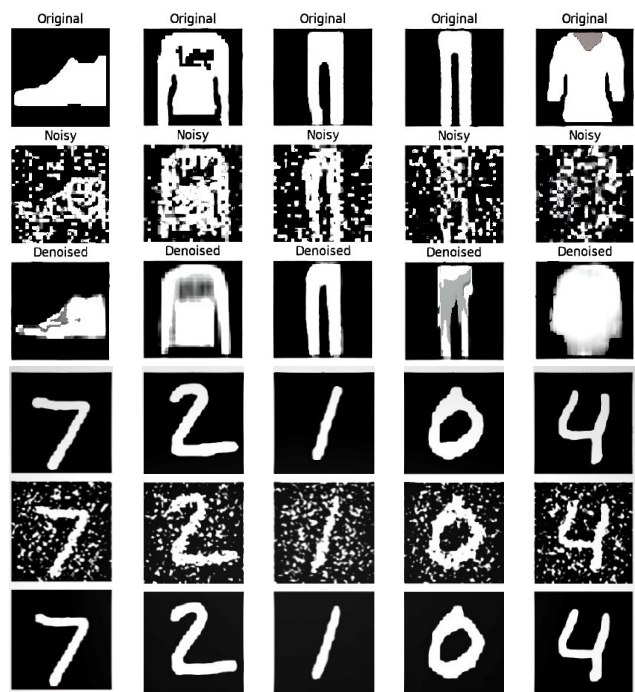


Figure 2. Comparison of original and noisy images

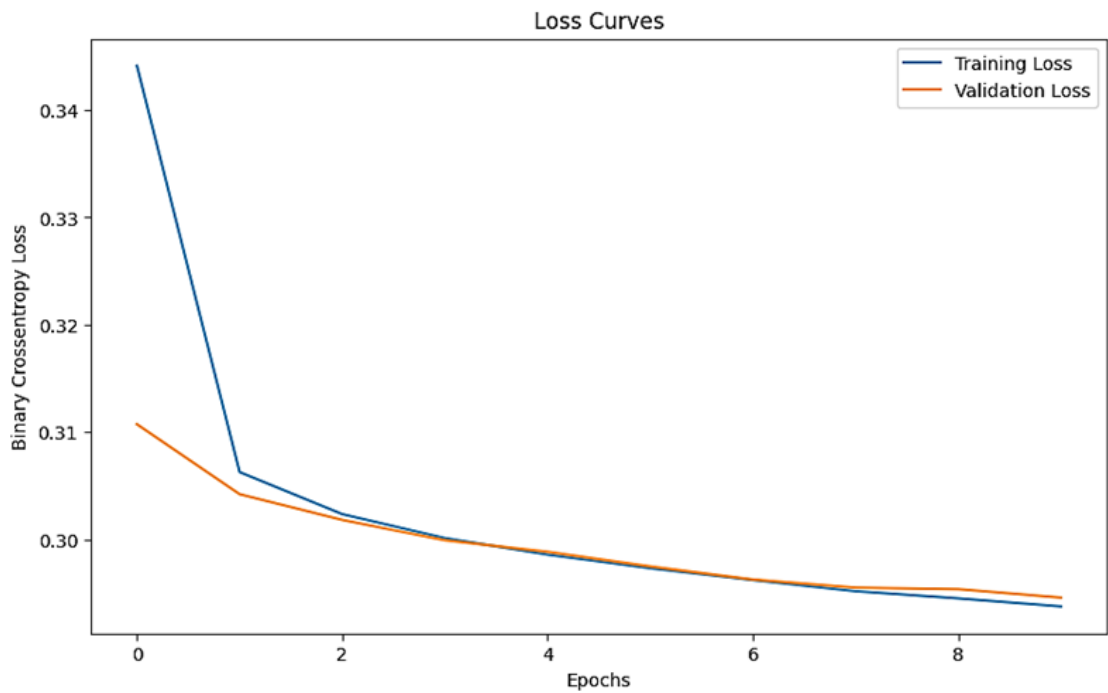


Figure 3. Training loss curve for the GAN model

Initially, the loss is high due to the large divergence between the resulting images and the original images and for training continues, the generator learns to improve the quality of the resulting images so that they become very close to the original images. The discriminator evaluates the produced images more accurately and the stabilization of the curve in the last loops indicates that the GAN has reached an equilibrium state. Where the generator has learned to effectively represent the basic patterns of noisy images and the loss curve also reflects the model's generalization ability. As the divergence between the loss of the training set and the validation set remains small, indicating that the model does not exhibit an overfitting problem.

Figure 4 is the most important for model evaluation as it shows a bar chart gives an empirical comparison of image quality, which is measured through the PSNR in two different conditions. The orange bar which represents noisy images implies low quality of the baseline, as the PSNR is about 9.5 dB. In contrast, the green bar that displays denoised images achieves significantly better quality as the PSNR will go above 17.5 dB. The given considerable enhancement of PSNR justifies the effectiveness of the developed image denoising algorithm. PSNR can be obtained based on the MSE and is a common quantitative measure; larger numbers indicate a high-quality image reconstruction.

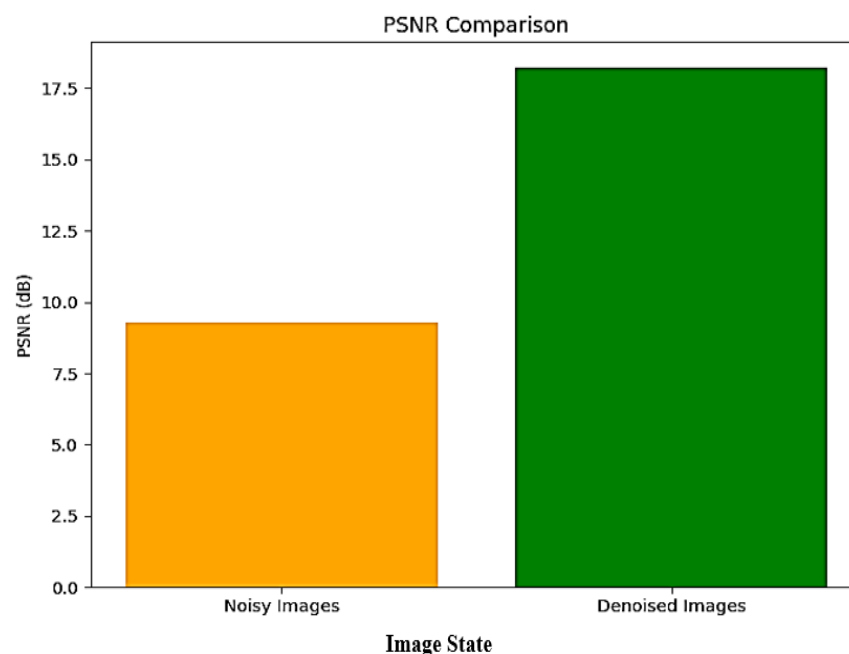


Figure 4. Comparison of original, noisy and denoised images and measured in PSNR

Moreover, the resulting images do not contain any additional noise, which is evidence that the generator did not introduce any random elements that might affect the quality of the images. This balance is a direct result of GAN training where the generator and discriminator work together to continually improve image quality. The use of a GAN here was a strategic decision to remove noise, as the generator learns the visual patterns of clean images through the feedback, it receives from the discriminator and the generator attempts to produce images that the discriminator cannot differentiate from the original images. On the other hand, the discriminator tries to improve its ability to distinguish between original and generated images and this competitive process allows incremental improvements of the model. Ultimately leading to restored images with high resolution and what distinguishes these results is that the model not only removed noise, but also improved the images comprehensively, making them very close to the original images in terms of quality and accuracy. The model showed high efficiency in recovering small details in images such as fine patterns on clothes, which is a remarkable achievement considering the complexity of the added noise and these results demonstrate the great potential offered by GAN in image enhancement applications and confirm its ability to deal with noise-related problems and restore original images with high accuracy and quality.

- i) PSNR for the noisy images was calculated compared to the original images and the value was 11.70 dB. This value is relatively low, indicating that the noisy images contain a large amount of noise that negatively affects their quality. In contrast, the PSNR value for the purified images was 19.41 dB,

reflecting a significant improvement in the quality of the restored images after applying the model. The high PSNR indicates that the model was able to effectively remove noise while reducing the error between the restored images and the original images.

- ii) SSIM was used to measure the extent of structural similarity between images and in the noisy images, the SSIM value was 0.61, reflecting a significant degradation in the textual structure of the images due to noise and the denoising images achieved SSIM of 0.93, indicating that the model restored the structural and detail very accurately and making the restored images very close to the original images in terms of quality and text and these calculations show that the GAN-based model was able to effectively reduce noise in the images, showing significant improvements in both PSNR and SSIM and this indicates that the restored images not only removed noise but also preserved the fine details and visual patterns that characterize the original images. Higher SSIM reflects the stability of the textural structure, while higher PSNR indicates reduced numerical differences between the original and restored images and the numerical results support the visual analysis of images and confirm that the model is an effective and powerful solution for image enhancement in applications that require high resolution such as medical imaging and visual data analysis.

In Table 3, GAN and U-Net have an excellent result among different algorithms used for image enhancement. With a PSNR equal to 19.41 dB and SSIM equal to 0.93. Using a T-test statistical analysis, the results with p-value < 0.01 showed a significant performance in compared with CNN denoisers. This corresponds to a substantial improvement over the standard CNN base model (+1.59 dB) and the traditional median filter (+4.16 dB). Regarding computational efficiency, the inference time of model (0.042 s/frame) is slightly higher than that of the resource-efficient CNN (0.015 s), yet it remains within an acceptable range for real-time applications on modern GPUs. This model used 3.5 million parameters, reflecting the complication of the U-Net generator to capture texture information. This constitutes a justifiable trade-off: a slightly higher computational overhead translates into a significant improvement in perceived image quality.

Table 3. Comparative performance analysis on the fashion-MNIST dataset

Method	PSNR (dB)	SSIM	Inference time (s/image)	Model parameters (M)
Noisy image (input)	11.70	0.61	–	–
Median filter (traditional)	15.25	0.75	0.002	–
CNN denoiser (e.g., DnCNN)	17.8	0.86	0.015	0.6
GAN-U-Net (Ours)	19.41	0.93	0.042	3.5

4.1. Error analysis and limitations

The analysis of failure cases shows certain limitations. The GAN parameter sometimes struggles with high-frequency textures in low-contrast images, such as dark 'Coat' or 'Shirt' classes, where the generator may fail to capture subtle fabric details as noise, leading to diminished texture details. On the other hand, under rare circumstances ($< 5\%$), the model detects slight “hallucination” artifacts. In which plausible but non-existent textures are generated. Future work may discuss the limitation by introducing self-attention to identify the difference between noise and semantic features.

4.2. Potential application scenarios

The image processing demonstrated by the architecture that forms gated attention unit (GAU) are of an amazing nature, a fact that can be attributed to their selective attention processes. GAUs adopt salience weighting and denoising methods in the process of denoising tasks of visual data on a large scale. These features have therefore been significantly used in improving medical imaging. The model architecture demonstrates adaptability suitable for deployment in the following:

- Medical imaging: the model can be adapted to remove Rician noise from magnetic resonance imaging (MRI) scans, where preserving tumor boundaries (edges) is critical for diagnosis.
- Satellite imaging: in remote sensing, the GAN-U-Net can work on reducing the sensor data noise, contributing to achieving feature analysis.
- Real-time surveillance: with a low inference time of 0.05s per frame, the model is able to enhance low-light video feeds, improving object recognition in noisy environments.

The multimodal large language models (MLLMs) [30], showed an increasing convergence between language understanding and visual content generation. This trend, although not directly focused on denoising, underscores the capability of advanced models to comprehend and process complex visual information, a fundamental ability also for effective denoising tasks. The accelerator architectures for deep neural networks discussed [31] provide the necessary hardware foundation to support these complex models. The outcomes

observed are similar to the results of Li *et al.* [4], who also showed the importance of GAN-based models in denoising tasks. The model presented successfully combines the GAN method with the consistency of the encoder-decoder architecture.

The methodology of this research focuses on the use of GANs, and Li *et al.* [32] presents an interesting approach for lightweight GANs for text-guided image manipulation. Despite the different application, the principle of building efficient and lightweight GAN models resonates directly with your goal of creating robust and scalable solutions. While recent state-of-the-art models, particularly those based on the transformer architecture (e.g., SwinIR and Restormer), demonstrate marginal performance superiority in certain metrics, our proposed GAN-U-Net architecture achieves an optimal trade-off. It provides high denoising efficacy (evidenced by the PSNR jump from 11.70 dB to 19.41 dB) combined with superior computational efficiency and significantly lower latency, making it a more viable solution for resource-constrained and real-time deployment scenarios. Research results that showed significant improvement in PSNR and SSIM using a U-Net-based GAN, are evidence of the power of this approach. Recent literature is heavily leaning towards diffusion models, as demonstrated [33] with the diffusion probabilistic models (DPM)-solver for fast sampling.

However, powerful generative techniques come with challenges, Westerlund [34] discusses the emergence of deepfake technology, Santosh *et al.* [35] highlights the “cat and mouse game” between diffusion models and detection methods, while Bianchi *et al.* [36] illustrates how easily accessible text-to-image generation can amplify demographic stereotypes. Oppenlaender [37] suggest the potential on the creativity of text-to-image generation for integrating advanced denoising techniques into more intuitive and interactive content creation tools. Schulhoff *et al.* [38] explores jailbreaking large language models and mitigating their vulnerabilities, although your goal is constructive denoising.

Peebles and Xie [39] on scalable diffusion models with Transformers, and Tian *et al.* [40], which focuses on self-supervised discrimination in diffusion transformers, these models, which often outperform GANs in generation quality, could offer future avenues for improving the denoising results you have achieved, the ability of your model to restore fine details (as shown in Figure 3) could be compared to what these more modern models can achieve. These ethical considerations and security challenges are important for understanding the technological landscape in which you operate, many subsequent studies focus on text-to-image generation using transformers, such as Thakur *et al.* [41] on scaling autoregressive models, Khmag [42] with Muse, Ding *et al.* [43] with CogView, and Yang *et al.* [44] with CogView2. While these differ from the denoising task, advancements in Transformer architectures and their ability to model complex relationships in data could be inspirational for improving the encoder-decoder components of your U-Net architecture, or even for developing hybrid models.

Finally, there is the issue of the efficiency of these powerful models. The research [45], [46] presents methods for accelerating diffusion models through various distillation techniques, these techniques could be crucial for making advanced models (whether improved GANs or future diffusion models) practical for the real-time applications you target, such as medical imaging or satellite data analysis, where processing speed can be a critical factor. Additionally, Chen *et al.* [47] demonstrates the cutting-edge capabilities of large commercial models in generating high-quality images from text, setting a benchmark that academic research can aspire to or attempt to achieve similar efficiency with open-source techniques.

5. CONCLUSION

In this study, a generative deep learning is proposed with GAN and U-Net to enhance image quality with a perceptual loss and adversarial learning. The proposed method removes noise while preserving the image details, outperforming traditional approaches in both quantitative and qualitative metrics. The methodology showcases adaptability across various noise types and imaging scenarios, emphasizing its utility in domains such as medical imaging, environmental monitoring and real-time video analysis. This study used the fashion-MNIST dataset as a benchmark for structural detail recovery. The GAN-U-Net architecture is designed to be domain-agnostic where integration of perceptual loss makes it highly suitable for more complex real-world scenarios. Such as, in medical imaging, this framework can be adapted to remove Rician noise from MRI scans without blurring tumor boundaries also in satellite remote sensing applications. It can effectively mitigate cloud cover and sensor noise, preserving crucial topographical details. A future work may focus on external validation using the BSD68 and Set12 datasets, as well as reducing the computational overhead to enable deployment on edge devices for real-time surveillance applications. A one should focus on reducing computational overhead and exploring hybrid generative models to further optimize performance for real-time.

FUNDING INFORMATION

The authors declare that the research described in their article was not funded by any organization or institution.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Firas Abdulrahman	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Yousif														
Shamil Kh. Ramadhan		✓				✓		✓	✓	✓	✓	✓		
Ahmed A. Mostfa	✓		✓				✓			✓	✓		✓	✓

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [AAM], upon reasonable request.

REFERENCES

[1] M. Yessenova *et al.*, “Application of informative textural law’s masks methods for processing space images,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4557–4566, 2023, doi: 10.11591/ijece.v13i4.pp4557-4566.

[2] J. Wang *et al.*, “Towards a unified identifier of satellite remote sensing images,” *Remote Sensing*, vol. 17, no. 3, Jan. 2025, doi: 10.3390/rs17030465.

[3] Q. Zhou, M. Wen, Y. Wang, M. Ding, and X. Zhang, “Collaborative masking based speckle disentanglement for self-supervised optical coherence tomography image despeckling,” *Optics and Lasers in Engineering*, vol. 183, 2024, doi: 10.1016/j.optlaseng.2024.108499.

[4] Y. Li *et al.*, “NTIRE 2023 challenge on image denoising: methods and results,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2023, pp. 1905–1921, doi: 10.1109/CVPRW59228.2023.00188.

[5] Y. Shen *et al.*, “Noise-aware fully webly supervised object detection,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11323–11332, doi: 10.1109/CVPR42600.2020.01134.

[6] H. Liang, A. Yu, M. Shao, and Y. Tian, “Multi-feature guided low-light image enhancement,” *Applied Sciences*, vol. 11, no. 11, May 2021, doi: 10.3390/app11115055.

[7] S. Shariflou, A. Agar, K. Rose, C. Bowd, and S. M. Golzan, “Objective quantification of spontaneous retinal venous pulsations using a novel tablet-based ophthalmoscope,” *Translational Vision Science and Technology*, vol. 9, no. 4, pp. 1–9, 2020, doi: 10.1167/tvst.9.4.19.

[8] S. Ghose, N. Singh, and P. Singh, “Image denoising using deep learning: convolutional neural network,” in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2020, pp. 511–517, doi: 10.1109/Confluence47617.2020.9057895.

[9] B. E. Soyly, M. S. Guzel, G. E. Bostanci, F. Ekinici, T. Asuroglu, and K. Acici, “Deep-learning-based approaches for semantic segmentation of natural scene images: a review,” *Electronics*, vol. 12, no. 12, 2023, doi: 10.3390/electronics12122730.

[10] Q. Zhang, S. Wang, X. Wang, Z. Sun, S. Kwong, and J. Jiang, “A multi-task collaborative network for light field salient object detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1849–1861, 2021, doi: 10.1109/TCSVT.2020.3013119.

[11] S. K. Ramadhan, F. A. Yousif, and F. S. Hamid, “Enhancing user experience with hand gesture volume control: overcoming challenges of camera distance,” *Cybernetics and Physics*, vol. 14, no. 4, pp. 375–382, Dec. 2025, doi: 10.35470/2226-4116-2025-14-4-375-382.

[12] R. Correa, G. Farias, E. Fabregas, S. D.-Canto, I. Pastor, and J. Vega, “Deep learning models to reduce stray light in TJ-II Thomson scattering diagnostic,” *Sensors*, vol. 24, no. 9, 2024, doi: 10.3390/s24092764.

[13] L. Zou, Y. Li, and F. Xu, “An adversarial denoising convolutional neural network for fault diagnosis of rotating machinery under noisy environment and limited sample size case,” *Neurocomputing*, vol. 407, pp. 105–120, 2020, doi: 10.1016/j.neucom.2020.04.074.

- [14] C. Chen, N. Hou, Y. Hu, S. Shirol, and E. S. Chng, "Noise-robust speech recognition with 10 minutes unparalleled in-domain data," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 4298–4302, doi: 10.1109/ICASSP43922.2022.9747755.
- [15] Y. Wang, S. Luo, L. Ma, and M. Huang, "RCA-GAN: an improved image denoising algorithm based on generative adversarial networks," *Electronics*, vol. 12, no. 22, 2023, doi: 10.3390/electronics12224595.
- [16] P. Nagpal, S. A. Bhinge, and A. Shitole, "A comparative analysis of ResNet architectures," in *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, Dec. 2022, pp. 1–8, doi: 10.1109/SMARTGENCON56628.2022.10083966.
- [17] S. Shanmugasundaram and N. Palaniappan, "Detection accuracy improvement on one-stage object detection using AP-loss-based ranking module and ResNet-152 backbone," *International Journal of Image and Graphics*, vol. 24, no. 3, May 2024, doi: 10.1142/S021946782450030X.
- [18] D. Bhattacharjee, T. Zhang, S. Süssstrunk, and M. Salzmann, "MuIT: an end-to-end multitask learning transformer," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 12021–12031, doi: 10.1109/CVPR52688.2022.01172.
- [19] R. M. Samant, M. R. Bachute, S. Gite, and K. Kotecha, "Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions," *IEEE Access*, vol. 10, pp. 17078–17097, 2022, doi: 10.1109/ACCESS.2022.3149798.
- [20] D. Yuan, X. Li, Z. He, Q. Liu, and S. Lu, "Visual object tracking with adaptive structural convolutional network," *Knowledge-Based Systems*, vol. 194, 2020, doi: 10.1016/j.knsys.2020.105554.
- [21] H. Wang, Z. Chen, and C. W. Chen, "Learned video compression via heterogeneous deformable compensation network," *IEEE Transactions on Multimedia*, vol. 26, pp. 1855–1866, 2024, doi: 10.1109/TMM.2023.3289763.
- [22] X. Xiao, N. N. Xiong, J. Lai, C. D. Wang, Z. Sun, and J. Yan, "A local consensus index scheme for random-valued impulse noise detection systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3412–3428, 2021, doi: 10.1109/TSMC.2019.2925886.
- [23] X. Wang, F. Dai, Y. Ma, J. Guo, Q. Zhao, and Y. Zhang, "Near-infrared image guided neural networks for color image denoising," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 3807–3811, doi: 10.1109/ICASSP.2019.8682692.
- [24] J. Wei, Y. Xia, and Y. Zhang, "M 3 Net: a multi-model, multi-size, and multi-view deep neural network for brain magnetic resonance image segmentation," *Pattern Recognition*, vol. 91, pp. 366–378, 2019, doi: 10.1016/j.patcog.2019.03.004.
- [25] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1418–1429, 2020, doi: 10.1109/TCYB.2018.2884715.
- [26] C. Tian, Y. Xu, W. Zuo, B. Zhang, L. Fei, and C.-W. Lin, "Coarse-to-fine CNN for image super-resolution," *IEEE Transactions on Multimedia*, vol. 23, pp. 1489–1502, 2020.
- [27] C. Tian *et al.*, "Lightweight image super-resolution with enhanced CNN," *Knowledge-Based Systems*, vol. 205, 2020, doi: 10.1016/j.knsys.2020.106235.
- [28] M. Tassano, J. Delon, and T. Veit, "An analysis and implementation of the FFDNet image denoising method," *Image Processing On Line*, vol. 9, pp. 1–25, 2019, doi: 10.5201/ipol.2019.231.
- [29] Y. Song, Y. Zhu, and X. Du, "Dynamic residual dense network for image denoising," *Sensors*, vol. 19, no. 17, Sep. 2019, doi: 10.3390/s19173809.
- [30] H. Fei *et al.*, "From multimodal LLM to human-level AI: modality, instruction, reasoning and beyond," in *32nd ACM International Conference on Multimedia*, Oct. 2024, pp. 11289–11291, doi: 10.1145/3664647.3689171.
- [31] F. M. -Martinez, J. L. Abellan, M. E. Acacio, and T. Krishna, "STONNE: enabling cycle-level microarchitectural simulation for DNN inference accelerators," in *2021 IEEE International Symposium on Workload Characterization (IISWC)*, Nov. 2021, pp. 201–213, doi: 10.1109/IISWC53511.2021.00028.
- [32] B. Li, X. Qi, P. H. S. Torr, and T. Lukasiewicz, "Lightweight generative adversarial networks for text-guided image manipulation," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020, pp. 22020–22031.
- [33] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "DPM-solver: a fast ODE solver for diffusion probabilistic model sampling in around 10 steps," in *36th International Conference on Neural Information Processing Systems*, 2022, pp. 5775–5787, doi: 10.52202/068431-0418.
- [34] M. Westerlund, "The emergence of deepfake technology: a review," *Technology Innovation Management Review*, vol. 9, no. 11, pp. 39–52, 2019, doi: 10.22215/TIMREVIEW/1282.
- [35] Santosh, L. Lin, I. Amerini, X. Wang, and S. Hu, "Robust CLIP-based detector for exposing diffusion model-generated images," in *2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Jul. 2024, pp. 1–7, doi: 10.1109/AVSS61716.2024.10672612.
- [36] F. Bianchi *et al.*, "Easily accessible text-to-image generation amplifies demographic stereotypes at large scale," in *2023 ACM Conference on Fairness, Accountability, and Transparency*, Jun. 2023, pp. 1493–1504, doi: 10.1145/3593013.3594095.
- [37] J. Oppenlaender, "The creativity of text-to-image generation," *Proceedings of the 25th International Academic Mindtrek Conference*, 2022, pp. 192–202, doi: 10.1145/3569219.3569352.
- [38] S. Schulhoff *et al.*, "Ignore this title and HackAPrompt: exposing systemic vulnerabilities of LLMs through a global scale prompt hacking competition," in *2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 4945–4977, doi: 10.18653/v1/2023.emnlp-main.302.
- [39] W. Peebles and S. Xie, "Scalable diffusion models with transformers," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2023, pp. 4172–4182, doi: 10.1109/ICCV51070.2023.00387.
- [40] Y. Tian, Z. Tu, H. Chen, J. Hu, C. Xu, and Y. Wang, "U-DiTs: downsample tokens in u-shaped diffusion transformers," in *38th Conference on Neural Information Processing Systems (NeurIPS 2024)*, 2024, pp. 51994–52013, doi: 10.52202/079017-1648.
- [41] R. S. Thakur, S. Chatterjee, R. N. Yadav, and L. Gupta, "Image de-noising with machine learning: a review," *IEEE Access*, vol. 9, pp. 93338–93363, 2021, doi: 10.1109/ACCESS.2021.3092425.
- [42] A. Khmag, "Additive Gaussian noise removal based on generative adversarial network model and semi-soft thresholding approach," *Multimedia Tools and Applications*, vol. 82, no. 5, pp. 7757–7777, 2023, doi: 10.1007/s11042-022-13569-6.
- [43] M. Ding *et al.*, "CogView: mastering text-to-image generation via transformers," in *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, Dec. 2021, pp. 19822–19835.
- [44] L. Yang, H. Qian, Z. Zhang, J. Liu, and B. Cui, "Structure-guided adversarial training of diffusion models," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 7256–7266, doi: 10.1109/CVPR52733.2024.00693.
- [45] M. Aghajarian, J. E. McInroy, and S. Muknahallipatna, "Deep learning algorithm for Gaussian noise removal from images," *Journal of Electronic Imaging*, vol. 29, no. 4, 2020, doi: 10.1117/1.jei.29.4.043005.

- [46] Z. M. Hussien and A. H. Al-Asadi, "Deep generative adversarial networks for noise reduction in medical images: a review," *Journal of Education and Science*, vol. 33, no. 3, pp. 24–34, 2024, doi: 10.33899/edusj.2024.148937.1448.
- [47] Z. Chen, K. Pawar, M. Ekanayake, C. Pain, S. Zhong, and G. F. Egan, "Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges," *Journal of Digital Imaging*, vol. 36, no. 1, pp. 204–230, 2023, doi: 10.1007/s10278-022-00721-9.

BIOGRAPHIES OF AUTHORS



Firas Abdulrahman Yousif    received the B.Sc. and M.Sc. degrees in Computer Engineering from Northern Technical University, Mosul, Iraq, in 2004 and 2009, respectively. He is currently a lecturer in the Department of Computer Science, College of Education, as well as the head of the cultural relations division in the Department of Scholarships and Cultural Relations, Presidency of the University of Al Hamdaniya. His research interests include the applications of artificial intelligence and computer networks. He can be contacted at email: firasabdulrahman@uohamdaniya.edu.iq.



Shamil Kh. Ramadhan    holds a diploma in Electronics (2003), a bachelor's degree in Computer Technology Engineering (2006), and a master's degree in Computer Technology Engineering (2022) from Northern Technical University in Iraq. He is currently an assistant lecturer at Northern Technical University, Mosul, Iraq. He is interested in computer engineering, artificial intelligence, computer vision, image processing, human-computer interaction, pattern recognition, and neural networks. He can be contacted at email: shamil.ramadhan@ntu.edu.iq.



Ahmed A. Mostfa    received his M.Sc. degree in Electrical and Computer Engineering from Virginia Commonwealth University, United States, in 2016. Since 2017, he has been serving as a lecturer in the Department of Computer Science at the University of Al Hamdaniya. He held the position of chair of the Department of Continuing Education. He currently moved to Ninevah University as a hosted lecturer and Computer Center Service. He has authored or co-authored more than five journal papers, contributing to advancements in various fields of research. His primary research interests include the application of artificial intelligence, Android development, and robotics. He can be contacted at email: mostfa@uohamdaniya.edu.iq.