International Journal of Engineering

Journal Homepage: www.ije.ir

Multi-head Spectral-attentive Residual Generative Adversarial Network: A High-fidelity Generative Adversarial Network Based Model for Image Haze Removal

A. Dey*, S. Biswas, A. Gupta, M. Debnath

Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India

PAPER INFO

Paper history:
Received 28 March 2025
Received in revised form 27 May 2025
Accepted 02 June 2025

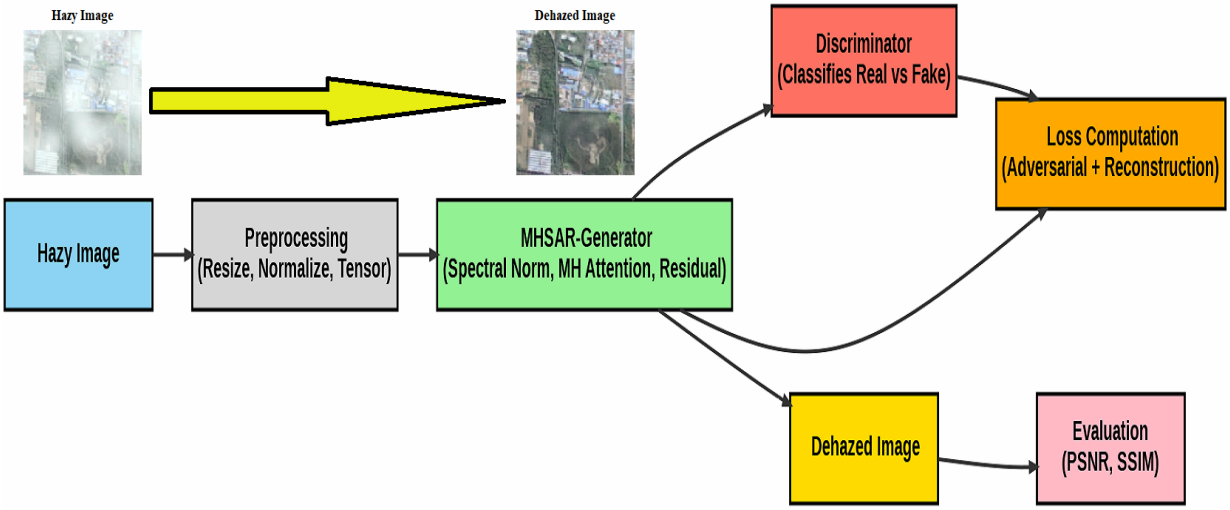
Keywords:
Image Haze Removal
MHSAR-GAN
Single Image Dehazing
Multi-Head Attention
Generative Adversarial Network
Computer Vision

ABSTRACT

Haze severely degrades image quality by suppressing contrast, clarity, and visibility, which is a challenge to tasks in many computer vision applications like autonomous driving, remote sensing, satellite image analysis, video surveillance, and action recognition. These vision-based tasks require clear and highly detailed visual information for efficient analysis and decision-making. This research proposes a novel GAN-based learning method, the Multi-Head Spectral-Attentive Residual Generative Adversarial Network (MHSAR-GAN), to improve the performance of image dehazing. The proposed deep learning-based image haze removal model combines spectral normalization to enhance training stability, multi-head attention for fine-tuning feature selection, and residual learning to retain important structural information to improve single-image dehazing. Depthwise convolutions are also incorporated into the attention mechanism for enhanced spatial feature extraction without added computational complexity. We tested our method on benchmark image dehazing datasets, Haze1K and RESIDE 6K, and compared its performance with state-of-the-art image dehazing models. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were pivotal metrics for assessing the model's performance, ensuring a comprehensive evaluation of image quality and structural fidelity. Experimental results indicate that the proposed MHSAR-GAN achieves superior haze removal with preserved fine-grained image details and clearer visibility compared to existing image dehazing methods in quantitative and qualitative comparisons.

***doi**: 10.5829/ije.2026.39.08b.03*

Graphical Abstract



*Corresponding Author Email: arnabdeycs@gmail.com (A. Dey)

Please cite this article as: Dey A, Biswas S, Gupta A, Debnath M. Multi-head Spectral-attentive Residual Generative Adversarial Network: A High-fidelity Generative Adversarial Network Based Model for Image Haze Removal. International Journal of Engineering, Transactions B: Applications. 2026;39(08):1812-20.

1. INTRODUCTION

Image haze removal is a challenging and vital research area in computer vision technologies. Hazy conditions—caused by particles such as smoke, fog, and pollution significantly degrade image quality by lowering contrast, reducing visibility, and obscuring fine details. This degradation adversely affects downstream tasks such as object detection, scene classification, semantic segmentation, and satellite imagery analysis, which rely on high-fidelity visual inputs. Removing haze from images (1) restores image clarity by eliminating or compensating for the scattering effects introduced by atmospheric pollutants. In particular, optical remote sensing systems and real-time vision systems are highly sensitive to atmospheric interference, where even minor visual distortion can severely hinder decision-making during critical operations. Thus, the main goal of the process is to find a method that gives accurate results under dense fog conditions. Traditionally, histogram equalization, contrast enhancement, and dark channel prior have been used, but they were ineffective in heavy fog conditions (2). Deep learning-based image dehazing typically follows two primary methodologies. The first approach relies on a physical degradation model, wherein neural networks estimate model parameters to reconstruct haze-free images. The second approach, which has gained substantial attention, involves an end-to-end dehazing framework where deep networks directly process hazy images to generate dehazed outputs without explicit parameter estimation. Recent advancements have incorporated models such as random forest regression, CNNs, and GANs, which excel in learning complex feature representations for improved image restoration.

The Multi-Head Spectral-Attentive Residual GAN (MHSAR-GAN) is an advanced generative adversarial network that aims to improve the quality of image translation by means of a robust methodology that overcomes the common setbacks to traditional GAN methods. One problem in standard GANs is instability during training; for instance, this may lead to mode collapse and a deteriorated quality of images. MHSAR-GAN attempts to solve these by implementing spectral normalization for its generator and discriminator, which ensures stable training conditions and maintains better gradient flow. Feature-and-context preservation is yet another challenge conventional GANs face, especially in cases that require intricacies of fine detail reconstruction. By incorporating multi-head self-attention with depthwise convolutions, MHSAR-GAN can locate and preserve both global scene structures and local textures. With the further refinement of generator residual blocks using adaptive instance normalization and SiLU activation - keeping contrast and structural integrity of generated images - the discriminator uses spectral normalization among its many structural levels to enforce the adversarial training toward improved sharpness and detail preservation. With a perfect balance of all the aforementioned features, MHSAR-GAN presents an advanced GAN model exhibiting high fidelity haze removal while overcoming the limitations of conventional GAN-based dehazing methods.

This research is noteworthy because it contributes the following:

(a) Develop MHSAR-GAN, a novel and efficient GAN-based framework for image haze removal.

(b) Incorporates spectral normalized residual multi-head self-attention and depthwise convolutions, enriching feature refinement and image dehazing.

(c) Adaptive Instance Normalization and SiLU activation for improved contrast preservation and texture reconstruction.

(d) Evaluation of the proposed MHSAR-GAN model on benchmark Haze1K and RESIDE 6K datasets and state-of-the-art dehazing methods.

2. RELATED WORKS

In this section, we discuss prior research efforts in the field of haze removal. Over time, much research has been done on various dehazing techniques, from traditional image processing techniques to latest deep learning-based models. The techniques are meant to clear an image of the haze using different forms of mathematical principles and learning-based strategies.

One of the earliest and widely recognized techniques for haze removal is the Dark Channel Prior (DCP) which was introduced by He et al. (3). This method relies on the assumption that in most haze-free images, at least one color channel has some pixels with very low intensity, except in sky regions. Using this prior knowledge, DCP estimates the transmission map and restores clear images. This method performs very poor on dense haze conditions and needs post-processing steps to retrieve artifacts and color distortions (4). Cai et al. (5) used DehazeNet, a convolutional neural network (CNN), instead of handcrafting priors to learn the mapping between hazy and clear images. The model learns relevant features automatically and improves over traditional models to a large extent. DehazeNet incorporated nonlinear activation functions and depth-wise feature extraction to enhance dehazing performance (6). Ren et al. (7) presented a Gated Fusion Network (GFN) with a selective fusion mechanism to dynamically maintain a balance between fine and coarse details of an image depending on the haze density. This gating mechanism preserves structural details that are important, allowing the method to perform well in a variety of environmental conditions. Qu et al. (8) employs additional perceptual loss functions and skip

connections into the Pix2Pix Dehazing framework to allow the retention of details in the output image. In contrast to the existing CNN-based methods that have focused only on pixel-wise accuracy, the Enhanced Pix2Pix Dehazing Network (EPDN) relies on perceptual similarity for increasing contrast and clarity in images. Cui et al. (9) introduced an efficient and effective network (EENet) with improved spatial-spectral learning for single image dehazing. This gives the model relevance to focus on important features and improve dehazing performance. To solve the problem of real-time processing, Patil et al. (10) present an unsupervised recurrent learning framework for real-world video dehazing, leveraging temporal coherence to enhance visibility restoration while eliminating the need for paired training datasets—marking a significant advancement in real-time haze removal from dynamic scenes. Su et al. (11) introduces a new dehazing approach that effectively bridges the synthetic-real domain gap by integrating multiple prior-based pseudo-labels with adaptive weighting and domain transfer mechanism, achieving superior performance on real scene single image dehazing. Zhao et al. (12) introduced a transformer enhanced texture attention model named TransDehaze for end-to-end single image dehazing. Chen et al. (13) proposed an unsupervised single-image dehazing method using a self-guided inverse-retinex GAN, significantly improving texture restoration and contrast enhancement in real-world hazy images. The model performs particularly well for outdoor scenes.

Different researchers have also applied GAN architectures to specific applications such as remote sensing. One such example is the GAN model proposed by Ning et al. (14), a multi-scale attention with deformable transformers is utilized to effectively tackle the challenge of remote sensing image dehazing. Similarly, Shen et al. (15), introduced spectral constraints to ensure consistency between different spectral bands and further improve clarity in multi-modal remote sensing images. Other methods include PMDNet (16), Optimized Dehazing (17), Multi-stream Feature Aggregation Network (18-22), Vision Transformer based DehazeFormer (23-26), DCP and Inverse Image (27), and Improved DCP (28). The structural similarity index proposed by Khosravi et al. (26) exploits maximally stable extremal regions to improve the detectability of the content of the image. Dehazing techniques have evolved significantly, from traditional handcrafted priors to sophisticated deep learning models. CNN methods have been very promising in feature extraction, whereas GAN methods could further improve de-hazing results by generating visually realistic and perceptually consistent images. Attention mechanisms, as well as, recent contrastive learning methods are likely to further advance the solutions in these arenas and make them applicable in quite a number of areas, such as remote sensing, video surveillance, and real-time environmental monitoring.

3. PROPOSED METHOD

Image dehazing is a challenging task that involves removing haze from images to enhance visibility and recover details. The proposed Multi-Head Spectral-Attentive Residual Generative Adversarial Network (MHSAR-GAN) leverages a carefully designed generator and discriminator to achieve high-quality dehazing results. Before feeding images into the proposed MHSAR-GAN architecture, the dataset undergoes a structured preprocessing pipeline. Hazy and clear images are first loaded and resized to 128×128 pixels to maintain consistency. They are then converted into tensors and normalized to the [-1,1] range, ensuring stable GAN training. A DataLoader is used for efficient batch processing, and the data is transferred to a GPU (if available) to accelerate computations. Finally, the preprocessed images are fed into the MHSAR-GAN generator, serving as input for the dehazing process. The Generator architecture is shown in Figure 1.

The following discussion elaborates on the architectures of both Generator and Discriminator, along with their importance in achieving superior performance.

3. 1. Generator (G) The generator is responsible for transforming a hazy image $x \in \mathbb{R}^{H\times W\times C}$ into a dehazed output $\hat{x} \in \mathbb{R}^{H\times W\times C}$. The architecture consists of three key components: encoding, residual blocks with multi-head self-attention, and decoding.

3. 1. 1. Encoding Stage The encoding stage extracts hierarchical features from the input image using a series of convolutional layers. The first layer applies a 7×7 convolution with spectral normalization (SN) and instance normalization (IN) followed by a non-linearity (σ), capturing low-level structural information. Two subsequent 3×3 convolutions with stride 2 progressively downsample the feature maps, enhancing spatial abstraction. The encoding process is formulated

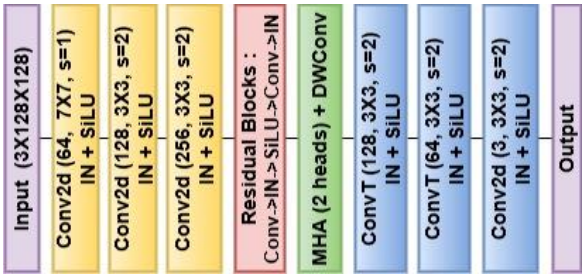


Figure 1. MHSAR-GAN Generator (G) Architecture

as is given below. The initial convolution operation is represented as given in Equation 1 and the downsampling layers are represented as Equation 2.

- Initial convolution:

$$f_1 = \sigma\big(\text{IN}(\text{SN}(\mathbf{W}_1 * x) + b_1)\big) \tag{1}$$

where, $*$ denotes convolution, σ is the activation function, $\mathbf{W}_1$ is the convolution kernel and b_1 is the bias term.

- Downsampling layers:

$$f_{i+1} = \sigma\big(\text{IN}(\text{SN}(\mathbf{W}_i * f_i) + b_i)\big) \tag{2}$$

here, i indexing each downsampling convolution. The algorithm for Generator and Discriminator, are given in Algorrithm 1 and 2 respectively. In the given Generator and Discriminator algorithm, SN denotes Spectral Normalization, IN represents Instance Normalization, BN represents Batch Norm, Conv denotes Convolution, ConvT denotes Transpose Convolution, and LReLU represents Leaky ReLU activation, s signify the stride.

3. 1. 2. Residual Blocks for Feature Refinement The model incorporates two (adjustable) residual blocks to facilitate robust feature learning. Each residual block refines the feature representation while maintaining stability using a residual scaling factor of 0.1 as given in Equation 3. This prevents excessive gradient magnitudes, ensuring a more stable optimization process. The use of residual learning aids in preserving fine details during dehazing. A series of residual blocks refine features while preserving essential details. Each block consists of two convolutions and a residual scaling factor $\alpha = 0.1$:

$$r_i = f_i + \alpha \cdot \sigma\Big(\text{IN}\Big(\text{SN}\Big(\mathbf{W}_2 * \sigma\big(\text{IN}(\text{SN}(\mathbf{W}_1 * f_i) + b_1)\big)\Big) + b_2\Big)\Big) \tag{3}$$

Algorithm 1. Generator (G)

Require: $x \in \mathbb{R}^{H\times W\times C}$
Ensure: $\hat{x} \in \mathbb{R}^{H\times W\times C}$

Encoding:

$x_1 = \sigma\Big(\text{IN}\Big(\text{SN}\big(\text{Conv}(x, 7\times 7, \text{s} = 1)\big)\Big)\Big)$

$x_2 = \sigma\Big(\text{IN}\Big(\text{SN}\big(\text{Conv}(x_1, 3\times 3, \text{s} = 2)\big)\Big)\Big)$

$x_3 = \sigma\Big(\text{IN}\Big(\text{SN}\big(\text{Conv}(x_2, 3\times 3, \text{s} = 2)\big)\Big)\Big)$

Residual Blocks:

for i = 1 to n do

$x_3 = x_3 + 0.1 \cdot \text{SN-Conv} \rightarrow \text{IN} \rightarrow \text{SiLU} \rightarrow \text{SN-Conv} \rightarrow \text{IN}(x_3)$

Multi-Head Attention (MHA):

$, K = \text{SN}\big(\text{Conv}(x_3, 1\times 1)\big), Q = \text{SN}\big(\text{Conv}(x_3, 1\times 1)\big)$

$V = \text{SN}\big(\text{Conv}(x_3, 1\times 1)\big)\ A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)$

$x_4 = AV$

$x_5 = \gamma \cdot \text{DepthwiseConv}(x_4) + \beta x_3$

Decoding:

$x_6 = \sigma\Big(\text{IN}\big(\text{ConvT}(x_5, 4\times 4, \text{s} = 2)\big)\Big)$

$x_7 = \sigma\Big(\text{IN}\big(\text{ConvT}(x_6, 4\times 4, \text{s} = 2)\big)\Big)$

$\hat{x} = \tanh\Big(\text{SN}\big(\text{Conv}(x_7, 7\times 7, \text{s} = 1)\big)\Big)$

Return $\hat{x}$

Algorithm 2. Discriminator (D)

Require: $x \in \mathbb{R}^{H\times W\times C}$
Ensure: $D(x) \in \mathbb{R}^{H'\times W'}$

Feature Extraction:

$x_1 = \text{LReLU}\Big(\text{SN}\big(\text{Conv}(x, 4\times 4, \text{s} = 2)\big)\Big)$

$x_2 = \text{LReLU}\Big(\text{BN}\Big(\text{SN}\big(\text{Conv}(x_1, 4\times 4, \text{s} = 2)\big)\Big)\Big)$

$x_3 = \text{LReLU}\Big(\text{BN}\Big(\text{SN}\big(\text{Conv}(x_2, 4\times 4, \text{s} = 2)\big)\Big)\Big)$

$x_4 = \text{LReLU}\Big(\text{BN}\Big(\text{SN}\big(\text{Conv}(x_3, 4\times 4, \text{s} = 1)\big)\Big)\Big)$

Classification:

$D(x) = \text{SN}\big(\text{Conv}(x_4, 4\times 4, \text{s} = 1)\big)$

Return $D(x)$

3. 1. 3. Multi-Head Attention (MHA) Mechanism To enhance feature dependencies over large spatial regions, a multi-head attention mechanism is integrated. Given an input feature map x, attention mechanism computes:

1. Query (Q), key (K), and value (V) matrices:

$$Q = W_q * x, \quad K = W_k * x, \quad V = W_v * x \tag{4}$$

2. Scaled dot-product attention

$$A = \text{Softmax}\left(\frac{QK^{\text{T}}}{\sqrt{d}}\right) \tag{5}$$

3. Weighted aggregation:

$$\hat{x} = AV \tag{6}$$

4. Multi-head representation:

$$\text{MHA}(x) = W_o[\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_h] \tag{7}$$

In Equations 4-7, h denotes the count of attention heads, and W_o projects the concatenated heads back to the original feature space. The attended feature maps x_4 undergo depthwise convolution, followed by a weighted residual connection with x_3. This transformer attention improves the model's ability to capture long-range dependencies crucial for removing non-uniform haze patterns.

3. 1. 4. Decoding Stage The decoding phase restores the original resolution of the dehazed image. Here, Two transposed convolution layers (4×4) progressively

upsample the feature maps. A final 7×7 spectral-normalized convolution, followed by a tanh activation, generates the refined output $\hat{x}$. The use of spectral normalization ensures stable training by controlling the Lipschitz constant of the generator. The decoding phase involves upsampling the features and reconstructing the output. The transposed convolution and the final reconstruction are mathematically expressed as Equations 8 and 9, respectively.

- Transposed convolution:

$$f_{i+1} = \sigma\left(\mathrm{SN}\left(\mathbf{W}_i^{\mathrm{T}} * f_i\right) + b_i\right) \quad (8)$$

- Final reconstruction:

$$\hat{x} = \tanh(\mathrm{SN}(\mathbf{W}_{\text{final}} * f_{\text{last}}) + b_{\text{final}}) \quad (9)$$

3. 2. Discriminator (D) The discriminator serves as an adversarial critic, distinguishing real dehazed images from fake ones generated by G. Here s denotes the stride and F signify the number of filters. The architecture follows a hierarchical feature extraction strategy.

The discriminator architecture is shown in Figure 2.

3. 2. 1. Feature Extraction Stage The input image undergoes four layers of convolution with spectral normalization (SN) and leaky ReLU activations. Each convolutional layer uses a 4×4 kernel, progressively reducing spatial dimensions while increasing feature abstraction. Here, batch normalization (BN) is applied after the first convolutional layer to stabilize training.

Hierarchical convolutional layers gradually reduce spatial dimensions. In Equation 10, $f_0 = x$ and the kernel size is 4×4.

$$f_i = \mathrm{LeakyReLU}(\mathrm{SN}(\mathbf{W}_i * f_{i-1}) + b_i) \quad (10)$$

3. 2. 2. Classification stage A final convolution layer reduces the feature maps to a lower-dimensional representation, providing a scalar output $D(x)$ as given in Equation 11. This output represents the probability of whether the input is a real or fake dehazed image.

$$D(x) = \sigma(\mathbf{W}_d * f_{\text{last}} + b_d) \quad (11)$$

here, $D(x)$ represents the probability of the image being real.

Figure 2. MHSAR-GAN Discriminator (D) Architecture

3. 3. Significance of the Proposed GAN Model The proposed GAN model enhances image dehazing by integrating spectral normalization, residual blocks, and self-attention mechanisms:

1. **Spectral Normalization:** Ensures stable training of both G and D by controlling weight magnitudes.
2. **Residual Learning:** Helps retain important details, making dehazed images more realistic.
3. **Self-Attention:** Enables the generator to consider long-range dependencies, reducing artifacts and improving dehazing quality.
4. **Hierarchical Discrimination:** The discriminator effectively differentiates real and fake images, enforcing structural consistency in dehazed outputs.

4. EXPERIMENTAL RESULTS

In this section, we present comprehensive experiments to rigorously evaluate the effectiveness of our proposed GAN method. The proposed work is implemented within a Docker-based Ubuntu environment with Python, and is executed on an NVIDIA Tesla T4 GPU (16GB VRAM) and 2 vCPUs of Intel Xeon architecture.

4. 1. Dataset Details We evaluated our method on publicly available benchmark haze removal datasets Haze1K, and RESIDE 6K.

Haze1K: The Haze1K dataset is a reference in establishing improved techniques for dehazing satellite imagery. It comprises 1,200 pairs of hazy and corresponding clear images, supplemented with Synthetic Aperture Radar (SAR) data acquired from Earth observation satellites such as GF-2 and GF-3. The dataset promotes research into haze removal, which is critical in enhancing the accuracy of remote sensing applications. Haze1K has specifically created three different haze levels-thick, moderate, and thin-that would mimic real atmospheric conditions in which the models are trained and hence help generalize scenarios across varying visibility scenarios. Given the inherent differences in capture timing between SAR and optical images, precise alignment is applied to facilitate meaningful comparisons. Haze1K enables the integration of SAR data as an auxiliary input to refine dehazing outcomes by preserving structural integrity.

RESIDE 6K: The RESIDE dataset, which stands for Realistic Single Image Dehazing, comprises paired hazy and clear images to facilitate dehazing research. Specifically, the RESIDE 6K has 6,000 image pairs allocated for training and 1,000 image pairs for testing. The training set is evenly divided into 3,000 indoor scene pairs (ITS) and 3,000 outdoor scene pairs (OTS), with all images standardized to a resolution of 400×400 pixels. Likewise, the test set maintains this balance, containing 500 indoor and 500 outdoor

pairs, ensuring diverse scene coverage without any resizing.

4. 2. Evaluation Metrics Here PSNR and SSIM are two main metrics.

Peak signal-to-noise ratio (PSNR): It is a key metric in engineering and image processing that quantifies the fidelity of a reconstructed image or video relative to its original form. It represents the ratio between the maximum possible signal power and the power of noise that affects accuracy. Since signals often have a wide dynamic range, PSNR is expressed in decibels (dB) on a logarithmic scale, making it a widely accepted standard for evaluating lossy compression effects on images and videos. Mathematically, PSNR is computed using the Mean Squared Error (MSE), which measures the average squared differences between the original noise-free image and its distorted counterpart. Given an $m \times n$ grayscale image I and its distorted version K, the MSE is defined as illustrated in Equation 12:

$$\text{MSE} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(I(i,j) - K(i,j)\right)^2 \quad (12)$$

The PSNR in decibels (dB) is then given by Equation 13:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{L^2}{\text{MSE}}\right) \quad (13)$$

Here, L denotes the maximum pixel intensity of the image. In the case of an 8-bit image, $L = 255$, and in case of an B-bit image, $L = 2^B$.

At higher PSNR value signifies better image quality, indicating lower noise interference in the reconstruction. In the case of color images, PSNR is computed either separately for each channel (e.g., in RGB or YCbCr color spaces) or as an average across all channels, ensuring a comprehensive evaluation of quality degradation.

Structural Similarity Index (SSIM): It is a widely used metric for assessing the perceived quality of digital images and videos. Unlike traditional error-based metrics like Mean Squared Error (MSE) or PSNR, SSIM evaluates image quality based on structural similarity, incorporating human visual perception principles. It considers three key factors: luminance, contrast, and structure, making it more aligned with human perception. Mathematically, SSIM is computed for two image windows x and y of size $N \times N$ as:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

Here, μ_x and μ_y are the mean intensities of image windows x and y, representing luminance. σ_x^2 and σ_y^2 are the variances of x and y, capturing contrast. σ_{xy} is the covariance between x and y, measuring structural similarity. C_1 and C_2 are small constants to stabilize the division, preventing instability when denominators are close to zero.

SSIM is computed locally over image patches and can be averaged to obtain a single quality score for the entire image. It is particularly effective in scenarios involving compression, denoising, and enhancement, where preserving structural details is crucial.

4. 3. Result Analysis Table 1 presents a comparative analysis of image dehazing methods on the Haze1K dataset across different fog densities, namely thin, moderate, and thick fog using PSNR and SSIM metrics. The results clearly highlight the superior performance of the proposed MHSAR-GAN model, which consistently outperforms prior state-of-the-art methods in terms of both PSNR and SSIM. Traditional methods such as DCP (3) exhibit bad performance, particularly under moderate and thick fog conditions, due to their reliance on handcrafted priors, which struggle with complex haze patterns. Deep learning-based methods, such as DehazeNet (5) and SAR-Opt-cGAN (19), demonstrate noticeable improvements over traditional techniques by learning data-driven features. However, DehazeNet (5), despite having the lowest number of trainable parameters (0.009M), delivers relatively poor performance. More advanced models like Trinity-Net (6) and FFA-Net (25), with 9.82M and 4.45M parameters respectively, show good results due to the integration of refined feature extraction and attention mechanisms. FFA-Net (25) demonstrates stronger perceptual quality, particularly in moderate haze. Nonetheless, these models are outperformed by AUNet (24) and DehazeFormer (23), which achieve higher PSNR and SSIM scores but at the cost of increased complexity with 7.417M and 9.68M parameters, respectively. Remarkably, our proposed MHSAR-GAN surpasses all existing methods, achieving the highest PSNR and SSIM scores across all haze levels while maintaining a lightweight structure of only 3.60M parameters. This improvement can be attributed to its novel Multi-Head self-attention Residual GAN, which effectively preserves structural details while enhancing perceptual quality. Notably, under thick fog conditions, the proposed MHSAR-GAN achieves a remarkable PSNR of 22.5850 and an SSIM of 0.9157, significantly improving visibility and fine details where other image dehazing methods struggle.

These results underscore the effectiveness of the proposed approach, demonstrating its robustness across varying haze intensities. The test results on thick, thin and moderate haze are shown in Figures 3, 4 and 5 using proposed MHSAR-GAN model. These figures collectively highlight the robustness of the proposed dehazing model across varying haze densities.

Comparison of various benchmark image dehazing methodologies on the RESIDE 6K data are provided in Table 2 taking PSNR and SSIM as evaluation metrics.

TABLE 1. Quantitative Evaluation of Image Dehazing Methods on HAZE1K

Method	#Params	Thin		Thick		Moderate	
	(in M)	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Actual	--	12.7712	0.7241	8.5893	0.4215	12.5867	0.7399
DCP (3)	--	13.1517	0.7246	10.2513	0.5850	9.7830	0.5735
DehazeNet (5)	0.009M	19.7529	0.8950	14.3321	0.7064	18.1250	0.8552
Trinity-Net (6)	9.82M	19.3708	0.7816	17.1281	0.7723	19.6918	0.7062
SAR-Opt-cGAN (19)	--	20.1950	0.8419	19.6553	0.7573	21.6616	0.7941
AUNet (24)	7.417M	23.0170	0.9060	19.9090	0.847	24.3270	0.9290
FFA-Net (25)	4.45M	20.1410	0.8582	19.1255	0.7976	22.5586	0.9132
DehazeForner (23)	9.68M	21.9274	0.8843	20.2133	0.8049	24.4407	0.9268
MHSAR-GAN (Ours)	**3.60M**	**24.0842**	**0.9426**	**22.5850**	**0.9157**	**26.0070**	**0.9384**

Figure 3. Test result on **Thick haze** on Haze1K dataset

Figure 4. Test result on **Thin haze** on Haze1K dataset

Figure 5. Test result on **Moderate haze** on Haze1K dataset

TABLE 2. Image Dehazing Results on RESIDE 6K dataset

Method	Params	PSNR	SSIM
DCP (3)	--	17.88	0.8160
DehazeNet (5)	0.009M	21.02	0.8700
PSD (20)	6.21M	15.47	0.8149
GridDehazeNet (21)	0.956M	25.65	0.9371
RefineDNet (22)	--	20.75	0.8656
DehazeFormer (23)	9.68M	28.96	0.9655
Proposed Model	**3.60M**	**29.70**	**0.9311**

Prior dehazing methods such as DCP (3) and PSD (20) exhibit relatively low PSNR values of 17.88 and 15.47 respectively, demonstrating their limitations under different haze conditions. PSD (20) has 6.21M trainable parameters. Figures 6 and 7 illustrates the test results of the proposed dehazing network on the RESIDE 6K indoor and outdoor test image dataset.

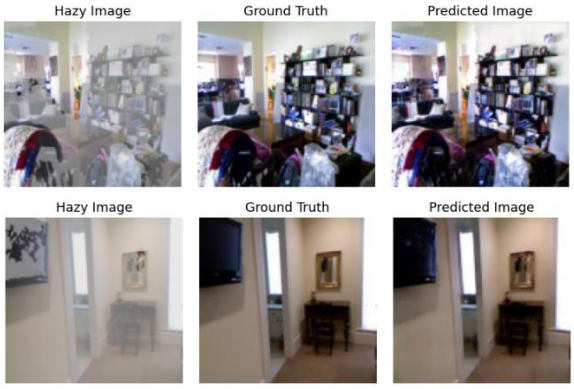


Figure 6. Test Results on RESIDE 6K dataset (Indoor scene)

Figure 7. Test result on RESIDE 6K dataset (Outdoor scene)

On the other hand, the performance of learning-based architectures such as DehazeNet (5) and RefineDNet (22) is comparatively better as evidenced by the corresponding PSNR values of 21.02 and 20.75. DehazeNet [5] and GridDehazeNet (21) exhibit the lowest trainable parameters; however their performance is moderate in terms of image haze removal. More advanced deep-learning methods like GridDehazeNet (21) and Dehazeformer (23) demonstrate significant gains, with Dehazeformer achieving high PSNR of 28.96 and an SSIM of 0.9655. The proposed MHSAR-GAN model achieves competitive performance with a PSNR of 29.70 and an SSIM of 0.9311, demonstrating its effectiveness in preserving structural details while maintaining high-quality restoration.

The proposed image dehazing model has fewer trainable parameters (3.60M) compared to DehazeFormer (9.68M) while achieving higher PSNR (29.70 vs 28.96) but slightly lower SSIM (0.9311 vs 0.9655). Thus, the slightly lower SSIM is an acceptable trade-off, given the clear advantage in PSNR and model compactness.

The proposed model balance between perceptual quality and efficiency makes it a strong contender for real-world dehazing applications. Improved dehazing significantly boosts performance in autonomous driving, remote sensing, video surveillance, object detection, scene classification, and semantic segmentation by enabling more effective feature extraction.

5. CONCLUSION AND FUTURE SCOPE

The study highlights how deep learning techniques, especially GANs, can address image haze removal. Although most conventional CNN-based dehazing methods are moderately successful, GANs are quite powerful for the reconstruction of fine details and creation of much more visually coherent outputs. Our contribution is the Multi-Head Spectral-Attentive Residual Generative Adversarial Network (MHSAR-GAN), which uses spectral normalization, multi-head self-attention, and residual scaling in an attempt to increase haziness removal capacity through better feature extraction and more stable training. The evaluation of the Haze1K and RESIDE 6K datasets validates the model's ability at different haze densities and verifies major improvements in visibility restoration and structural preservation. Despite all these developments, challenges like handling extreme haze weather and color fidelity remain open areas of exploration. Future work may include refining attention mechanisms, boosting computational efficiency, and exploring self-supervised learning to reduce reliance on paired training data for broader adaptability in real-world single image dehazing applications.

6. REFERENCES

1. Khatun A, Mostafiz R, Shorif SB, Uddin MS, Hadi MA. Dehazing using Generative Adversarial Network-A Review. SN Computer Science. 2024;6(1):20. 10.1007/s42979-024-03571-0
2. Ali AM, El-Rabaie E-SM, Ramadan KF, Abd El-Samie FE, El-Shafai W. CLEAR-Net: A Cascaded Local and External Attention Network for Enhanced Dehazing of Remote Sensing Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2025. 10.1109/jstars.2025.3534984
3. He K, Sun J, Tang X. Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence. 2010;33(12):2341-53. 10.1109/tpami.2010.168
4. Tao T, Xu H, Guan X, Zhou H. MCADNet: A Multi-Scale Cross-Attention Network for Remote Sensing Image Dehazing. Mathematics (2227-7390). 2024;12(23). 10.3390/math12233650
5. Cai B, Xu X, Jia K, Qing C, Tao D. Dehazenet: An end-to-end system for single image haze removal. IEEE transactions on image processing. 2016;25(11):5187-98. 10.1109/TIP.2016.2598681
6. Chi K, Yuan Y, Wang Q. Trinity-net: Gradient-guided swin transformer-based remote sensing image dehazing and beyond. IEEE Transactions on Geoscience and Remote Sensing. 2023;61:1-14. 10.1109/tgrs.2023.3285228
7. Ren W, Ma L, Zhang J, Pan J, Cao X, Liu W, et al., editors. Gated fusion network for single image dehazing. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. 10.1109/CVPR.2018.00343
8. Qu Y, Chen Y, Huang J, Xie Y, editors. Enhanced pix2pix dehazing network. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. 10.1109/CVPR.2019.00835
9. Cui Y, Wang Q, Li C, Ren W, Knoll A. EENet: An effective and efficient network for single image dehazing. pattern recognition. 2025;158:111074. 10.1016/j.patcog.2024.111074
10. Patil PW, Randive SN, Gupta S, Rana S, Venkatesh S, Murala S. Unpaired recurrent learning for real-world video de-hazing. Pattern Recognition. 2025:111698. 10.1016/j.patcog.2025.111698
11. Su Y, Wang N, Cui Z, Cai Y, He C, Li A. Real scene single image dehazing network with multi-prior guidance and domain transfer. IEEE Transactions on Multimedia. 2025. 10.1109/TMM.2025.3543063
12. Zhao X, Xu F, Liu Z. TransDehaze: transformer-enhanced texture attention for end-to-end single image dehaze. The Visual Computer. 2025;41(3):1621-35. 10.1007/s00371-024-03458-4
13. Chen H, Chen R, Li Y, Li H, Li N. Unsupervised single-image dehazing via self-guided inverse-retinex GAN. Multimedia Systems. 2025;31(2):139. 10.1007/s00530-025-01713-9
14. Ning J, Yin J, Deng F, Xie L. MABDT: Multi-scale attention boosted deformable transformer for remote sensing image dehazing. Signal Processing. 2025;229:109768. 10.1016/j.sigpro.2024.109768
15. Shen H, Zhong T, Jia Y, Wu C. Remote sensing image dehazing using generative adversarial network with texture and color space enhancement. Scientific Reports. 2024;14(1):12382. 10.1038/s41598-024-63259-6
16. Pushpalatha D, Prithvi P. PMDNet: A multi-stage approach to single image dehazing with contextual and spatial feature preservation. Journal of Visual Communication and Image Representation. 2025;107:104379. 10.1016/j.jvcir.2024.104379
17. Ayoub A, El-Shafai W, El-Samie FEA, Hamad EK, El-Rabaie E-SM. Video and image quality improvement using an enhanced optimized dehazing technique. Multimedia Tools and

Applications. 2025;84(20):22681-99. 10.1007/s11042-024-19263-z

18. Wu J, Tao H, Xiao K, Chu J, Leng L. Multi-stream feature aggregation network with multi-scale supervision for single image dehazing. Engineering Applications of Artificial Intelligence. 2025;139:109486. 10.1016/j.engappai.2024.109486
19. Grohnfeldt C, Schmitt M, Zhu X, editors. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium; 2018: IEEE. 10.1109/igarss.2018.8519215
20. Chen Z, Wang Y, Yang Y, Liu D, editors. PSD: Principled synthetic-to-real dehazing guided by physical priors. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. 10.1109/CVPR46437.2021.00710
21. Liu X, Ma Y, Shi Z, Chen J, editors. Griddehazenet: Attention-based multi-scale network for image dehazing. Proceedings of the IEEE/CVF international conference on computer vision; 2019. 10.1109/ICCV.2019.00741
22. Zhao S, Zhang L, Shen Y, Zhou Y. RefineDNet: A weakly supervised refinement framework for single image dehazing. IEEE Transactions on Image Processing. 2021;30:3391-404. 10.1109/tip.2021.3060873
23. Song Y, He Z, Qian H, Du X. Vision transformers for single image dehazing. IEEE Transactions on Image Processing. 2023;32:1927-41. 10.1109/tip.2023.3256763
24. Du Y, Li J, Sheng Q, Zhu Y, Wang B, Ling X. Dehazing network: Asymmetric unet based on physical model. IEEE Transactions on Geoscience and Remote Sensing. 2024;62:1-12. 10.1109/tgrs.2024.3359217
25. Qin X, Wang Z, Bai Y, Xie X, Jia H, editors. FFA-Net: Feature fusion attention network for single image dehazing. Proceedings of the AAAI conference on artificial intelligence; 2020. 10.1609/aaai.v34i07.6865
26. Khosravi M, Hassanpour H. A novel image structural similarity index considering image content detectability using maximally stable extremal region descriptor. International Journal of Engineering-Transactions B: Applications. 2017;30(2):172. 10.5829/idosi.ije.2017.30.02b.03
27. Zhou X, Bai L, Wang C. Single image dehazing algorithm based on dark channel prior and Inverse Image. International Journal of Engineering Transactions A: Basics. 2017;30(10):1471-8. 10.5829/ije.2017.30.10a.07
28. Nasrabad FA, Hassanpour H, Amiri SA. Adaptive image dehazing via improving dark channel prior. International Journal of Engineering Transactions B: Applications. 2019;32(2):253-9. 10.5829/ije.2019.32.02b.10

Persian Abstract

چکیده

مه با کاهش کنتراست، وضوح و قابلیت دید، کیفیت تصویر را به شدت کاهش می دهد، که این امر چالشی برای بسیاری از کاربردهای بینایی کامپیوتر مانند رانندگی خودکار، سنجش از دور، تجزیه و تحلیل تصاویر ماهواره ی، نظارت تصویری و تشخیص حرکت است .این وظایف مبتنی بر بینایی برای تجزیه و تحلیل و تصمیم گیری کارآمد، به اطلاعات بصری واضح و بسیار دقیقی نیاز دارند .این تحقیق یک روش یادگیری مبتنی بر GAN جدید، شبکه رقابتی مولد باقیمانده با توجه طیفی چند سر(MHSAR-GAN)، را برای بهبود عملکرد مه زدایی تصویر پیشنهاد می دهد .مدل پیشنهادی حذف مه تصویر مبتنی بر یادگیری عمیق، نرمال سازی طیفی را برای افزایش پایداری آموزش، توجه چند سر برای تنظیم دقیق انتخاب ویژگی و یادگیری باقیمانده را برای حفظ اطلاعات ساختاری مهم برای بهبود مه زدایی تک تصویر ترکیب می کند .پیچش های عمقی نیز در مکانیسم توجه برای استخراج ویژگی های مکانی بهبود یافته بدون پیچیدگی محاسباتی اضافی گنجانده شده اند .ما روش خود را بر روی مجموعه داده های مه زدایی تصویر معیار، Haze1K و RESIDE 6K آزمایش کردیم و عملکرد آن را با مدل های مه زدایی تصویر پیشرفته مقایسه کردیم .نسبت سیگنال به نویز پیک (PSNR) و شاخص شباهت ساختاری (SSIM) معیارهای محوری برای ارزیابی عملکرد مدل بودند که ارزیابی جامعی از کیفیت تصویر و صحت ساختاری را تضمین می کردند .نتایج تجربی نشان می دهد که MHSAR-GAN پیشنهادی در مقایسه با روش های مه زدایی تصویر موجود در مقایسه های کمی و کیفی، به حذف مه بهتری با حفظ جزئیات ریز تصویر و وضوح بیشتر دست می یابد.