Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 13, No. 4, 2025, 393-405.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Image Dehazing Using a Convolutional Autoencoder Network with Integrated Convolutional Block Attention

Homayoun Rastegar and Hassan Khotanlou*

RIV Lab., Department of Computer Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran

Article Info

Article History:
Received 16 April 2025
Revised 26 May 2025
Accepted 04 July 2025

DOI:10.22044/jadm.2025.16079.2723

Keywords:
Image Dehazing, Convolutional Autoencoder Network, Deep Learning, Self-supervised Learning, Convolutional Block Attention.

**Corresponding author: Khotanlou@basu.ac.ir (H. Khotanlou).*

Abstract

One challenge in digital image processing is haze, which is particularly prevalent in humid and rainy environments. Examples of AI-based systems susceptible to this challenge include smart traffic control cameras, autonomous vehicles, Video Assistant Referee (VAR) systems in football stadiums, and security and surveillance cameras. This paper proposes a method to mitigate haze using self-supervised learning (SSL) and deep learning. A Convolutional Autoencoder Network (CAN) with a Convolutional Block Attention Module (CBAM) was developed to reduce haze from images. This method's advantage lies in its reduced number of layers and filters compared to previous models, as well as its utilization of CBAM to emphasize important convolutional channels and image regions. Experiments demonstrate that excessive convolutional filters, while intended to generate diverse features, can hinder a model's ability to dehaze images effectively. Therefore, filter numbers should be carefully limited. A combined loss function was employed to train the proposed architecture, which was evaluated using the NH-haze dataset and the Realistic Single Image Dehazing (RESIDE) dataset. Structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) were utilized for evaluation. Test results indicate that the proposed architecture exhibits higher performance compared to state-of-the-art methods.

1. Introduction

Haze is a meteorological phenomenon consisting of tiny water droplets that primarily forms at higher altitudes. This phenomenon reduces visibility and weakens colors in the environment. Imaging under such conditions results in blurry images with colors far from reality. Haze can even obscure objects in images, disrupting the complete understanding of visuals by intelligent systems such as identification and tracking systems, license plate readers, autonomous vehicles, Video Assistant Referee (VAR), and various others. Typically, classical image quality enhancement and haze removal models require a reference image. However, using machine vision techniques, it is possible to improve image quality and remove haze in a self-learning manner without a reference image. Therefore, haze removal methods can be broadly categorized into two groups: classical information-based methods and self-learning deep methods [1]. Recent research indicates that self-learning deep methods have shown high effectiveness in removing haze from digital images. Self-supervised learning (SSL) is a type of machine learning that relies on the data itself rather than human-generated labels to produce output signals [2]. In this learning approach, there is no need for human image labeling; instead, the model generates an output image that closely resembles the corresponding input image using the input itself. One common and contemporary method for implementing this type of learning is through autoencoder models. Autoencoders are a type of convolutional neural network used to learn efficient features on unlabeled data. An autoencoder consists of an encoder that compresses input data into a latent representation and a decoder that reconstructs the original data from this representation. This network has two unique properties: it compresses and reduces the dimensions of the data in the most effective format, and it creates diverse representations of the input images to enable the model to recognize and learn the image more deeply [3]. For image dehazing, the convolutional autoencoder network requires a hazy input image and a ground truth haze-free image as output. In this paper, we propose a model using self-supervised learning and a convolutional autoencoder model with CBAM to remove haze from digital images.

The proposed model has a smaller number of convolutional filters and is relatively less deep than previous models, which reduces model complexity and the probability of overfitting. One problem faced by researchers and designers in the field of deep neural networks is the unavailability of powerful computer systems for training deep models. This paper demonstrates that this problem can be mitigated to a great extent through the proper design of the deep neural network. It is important to note that the number of filters and depth of the network are not the only factors in solving this problem; the correct arrangement of these parameters is crucial. We also demonstrate that the CBAM mechanism effectively diminishes the influence of less significant convolutional channels during training, reducing computational cost while enhancing the efficiency of the image dehazing method. The contributions of this paper are as follows:

- Proposing a novel convolutional autoencoder architecture with the Convolutional Block Attention Module to effectively reduce haze in digital images.
- Introducing a composite loss function designed to maximize the similarity between the output image of the proposed model and the original image.
- Developing a model with low depth and a small number and size of convolutional filters to minimize computational costs and streamline model execution.

The rest of the paper is organized as follows. Related work is presented in Section 2. Our proposed method for image dehazing is introduced in Section 3. Section 4 is devoted to the experimental results and comparisons. Finally, the paper is concluded in Section 5.

2. Related works

Image dehazing, a crucial task in machine vision and digital image processing, has attracted considerable research attention. Chaitanya and Mukherjee [1] proposed an end-to-end network for single image dehazing using the CycleGAN model and a novel loss function to enhance performance. They claimed that their proposed loss function can approximate the dehazed image to the ground truth. Jeong et al. [4] presented an end-to-end network model incorporating zoomed convolution groups for image dehazing. They asserted that their model can perform complex computational operations efficiently, reducing processing time without compromising performance. Notably, they utilized zoomed convolution groups to extract more effective channels and enhance model efficiency. Choudharya et al. [5] introduced a deep Generative Adversarial Network (GAN) for image dehazing, employing a perceptual loss function to extract high-level features instead of pre-pixel loss functions. Hartantoa and Rahadiantia [6] proposed a method utilizing PDR-Net, pyramid dilated convolution, preprocessing, post processing, and an attention module for image dehazing. Dharejo et al. [7] developed a wavelet Hybrid (Local-Global Combined) Network (WH-Net) for image dehazing, leveraging a convolutional neural network (CNN) in the wavelet domain. They argued that low-level features are more crucial than high-level features for effective dehazing. Li et al. [8] contended that certain digital image dehazing algorithms suffer from a fundamental limitation: their inability to fully extract global image information, leading to incomplete dehazing. To address this, they proposed a hybrid model combining an end-to-end convolutional neural network and a vision transformer to incorporate global features in the dehazing process. Babu et al. [9] employed an end-to-end network comprising a dehazing network, a discriminator network, and a fine-tuning network for image dehazing. They integrated these three models to achieve superior results. First, they processed hazy images using the dehazing network, which estimates the transmission map and atmospheric light alongside parallel convolutional layers. Subsequently, the discriminator network extracted discriminative dehazed images. Finally, the fine-tuning network utilized the discriminator's output to refine the dehazed results. Shakeri et al. [10] proposed an intelligent histogram partitioning method that enhances contrast while preserving image details. Wang et al. [11] introduced DRHNet, an end-to-end image dehazing model that dehazes images by subtracting a learned negative residual map from the hazy image. They claimed that DRHNet effectively extracts and aggregates conceptual and effective information. Additionally, they introduced a novel nonlinear activation function, PRRelu (Reverse Parametric Rectified Linear Unit), to enhance representation learning and accelerate convergence. Babu and Venkatram [12] proposed an Adaptive Bilateral Filter (ABF) with an optimal selection of spatial weight parameters for image dehazing. They emphasized the importance of classifying images into hazy and non-hazy categories to ensure that non-hazy images are not included in the dehazing process, thereby simplifying the procedure. To achieve this, they developed a deep CNN network to classify images into these two categories. Hodges et al. [13] presented a deep CNN model trained on unmatched images for image dehazing, inspired by the Siamese network architecture. Yin et al. [14] proposed a Visual Attention Dehazing Network (VADN) with multi-level refinement and fusion. Their model incorporates a haze attention map that injects supplementary haze information into extracted features. VADN comprises a feature extraction network, a recurrent refinement network, and an encoder-decoder network. First, the feature extraction network extracts features at different levels. Subsequently, the recurrent refinement network generates the haze attention map, which is then fed into the encoder-decoder network for dehazing. Yin et al. [15] developed an end-to-end dehazing network with a spatial/channel attention block to extract more informative features. This block, based on an encoder-decoder architecture with a pyramid pooling operation, is placed at the end of the encoder to aid the decoder in dehazing. Balla et al. [16] proposed a residual convolutional neural network (RCNN) for image dehazing, comprising 14 layers. Their RCNN extracts local features from the hazy image and incorporates them as the fourth color channel alongside the existing three channels. Subsequently, the RCNN is retrained using 4-channel hazy images as input and corresponding haze-free images as output. Hu et al. [17] introduced High-Low level task combination network (HLNet) based on multitask learning. Their network can learn both low-level and high-level tasks, claiming that this approach restores features at different levels, mitigates the impact of Batch Normalization (BN) in the encoder, and results in improved dehazing performance. Yen et al. [18] proposed a deep CNN model advocating for image basis component recovery instead of end-to-end network maps. They suggested decomposing the image into its basic elements and then using a trained CNN architecture to determine whether each element contains haze. If contaminated, it is replaced with its haze-free equivalent. Ren et al. [19] proposed a multi-scale deep neural network for image dehazing, training it to learn the mapping between haze images and their transmission maps. They employed a coarse-scale net for holistic transmission map prediction using the input image and a fine-scale net for refining dehazed image results. Additionally, they introduced a holistic edge-guided network to refine the edges of the estimated transmission map. As discussed, image dehazing from digital images remains a captivating area of research in artificial intelligence. The literature review reveals that researchers are actively exploring solutions and methods based on deep learning and convolutional neural networks to address this challenge. Notably, the results of their research demonstrate the high effectiveness of deep learning and convolutional neural network-based methods in solving such problems. Therefore, this paper aims to investigate a method that can dehaze digital images using deep learning with greater efficiency than existing approaches.

3 Materials and Method

3.1 Convolutional Autoencoder Network

Convolutional Autoencoder Networks (CAN) are developed from simple autoencoder networks that utilize convolutional layers instead of only fully connected layers. In these networks, the input and output layers have the same size. The network has two main parts: the encoder and the decoder. The encoder extracts, compresses, and codes the meaningful features of input images, while the decoder decodes the encoded information to produce a new image. To train these self-supervised learning networks, paired images are required: an initial image and a corresponding expected image. The network adjusts its weights to generate the expected secondary image when given the primary image. Figure 1 shows the basic architecture of a convolutional autoencoder network. As illustrated in the figure, the encoder section progressively compresses information, culminating in the most meaningful and compact representation at its end. Conversely, the decoder section expands this information, returning it to its original size. Applications of CAN include image denoising, image dehazing, and low-light enhancement [20].

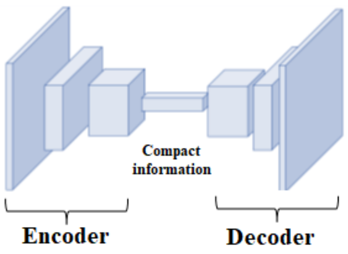

Figure 1. Architecture of the basic convolutional autoencoder network. Each blue box represents a convolutional layer.

3.2 Convolutional Block Attention Module

Drawing on insights from Woo et al. [21], oversized deep neural networks often generate redundant or closely related features, particularly on smaller datasets, which can hinder performance. To address this, they introduced the Convolutional Block Attention Module (CBAM), a computationally efficient framework that integrates both channel-wise and spatial attention mechanisms, inspired by Chen et al. [22]. The channel attention module assigns weights to convolutional channels based on their significance, emphasizing the most relevant ones, while the spatial attention module highlights critical regions within the feature map, enhancing focus on perceptually salient locations. CBAM operates through two sequential submodules: the Channel Attention Module and the Spatial Attention Module, seamlessly embeddable within feed-forward convolutional neural networks. Initially, an input feature map, where C, H, and W represent the number of channels, height, and width, respectively, undergoes processing. The channel attention module employs global average and max pooling to distill spatial information and highlight distinctive features, subsequently producing a channel attention map via shared dense layers. This map refines the input by weighting each channel accordingly. Next, the spatial attention module generates a spatial attention map by compressing the channel-refined features into two 2D maps through pooling, followed by concatenation and convolution, thus prioritizing key spatial regions.

By synergistically combining these processes, CBAM effectively discerns "what" to emphasize through channel attention and "where" to focus via spatial attention. This dual mechanism enables the model to extract more refined and impactful features, optimizing performance while mitigating the inefficiencies of oversized architectures. Figure 2 illustrates the structure of these two modules. The integration of CBAM into convolutional neural networks is straightforward, as it can be applied to any layer without requiring significant architectural changes. This flexibility makes CBAM highly adaptable across various network designs, such as ResNet, VGG, or EfficientNet. Furthermore, CBAM's lightweight design ensures minimal computational overhead, making it suitable for resource-constrained environments like mobile devices or edge computing systems. Experimental results from Woo et al. [21] demonstrate that CBAM consistently improves classification and detection performance across multiple benchmark datasets, including CIFAR-10, CIFAR-100, and ImageNet. The module's ability to suppress irrelevant features while enhancing relevant ones contributes to its effectiveness, particularly in scenarios with limited training data.

3.3 Total Loss Function

In the proposed method, to maximize the similarity between the model's outputs and the original image, a composite loss function is employed. This function blends the Mean Squared Error (MSE) loss, which minimizes pixel-wise differences between the input and reconstructed images, defined as follows:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i)^2 \qquad (1)$$

Additionally, the Structural Similarity Index (SSIM) loss function is incorporated, preserving the image's structural features and expressed as follows:

$$L_{SSIM} = 1 - SSIM\,(y - y) \qquad (2)$$

Furthermore, the L1 Loss function, also known as Mean Absolute Error (MAE), is included to aid in preserving fine textures and structural details of the image, formulated as follows:

$$L_{L1} = \frac{1}{N}\sum_{i=1}^{N}|\,y_i - y_i\,| \qquad (3)$$

Ultimately, these loss functions are integrated into a linear combination:

$$L_{total} = \alpha L_{MSE} + \beta L_{SSIM} + \gamma L_{L1} \qquad (4)$$

Where α, β, and γ are weighting coefficients that determine the relative importance of each loss function.

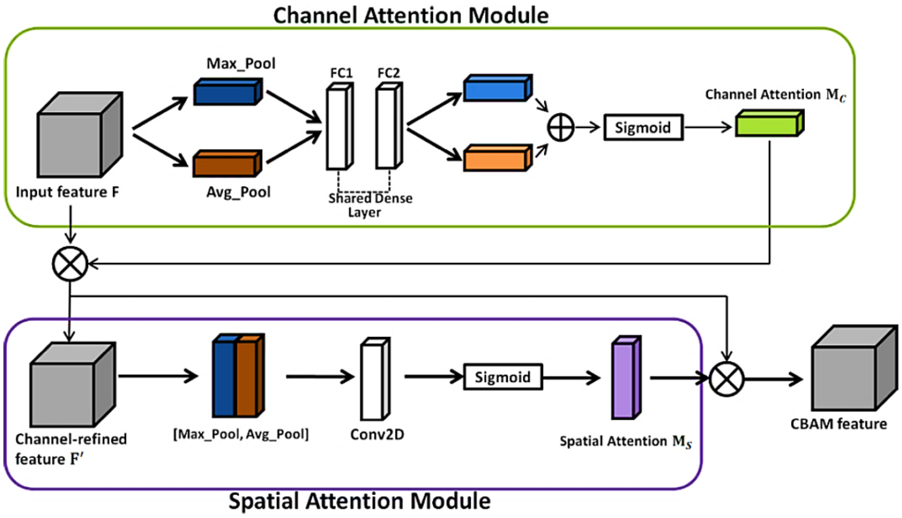

Figure 2. The Convolutional Block Attention Module integrates two complementary components: the channel attention module, positioned at the upper tier and the spatial attention module, situated at the lower tier.

3.4 Proposed Architecture

This paper proposes a self-supervised learning method for image dehazing. Autoencoder networks are a well-established and appealing approach in this field. Through extensive investigations and testing, a novel autoencoder network architecture incorporating CBAM has been developed. This architecture effectively removes haze from hazy images, producing results closely resembling the original image. Notably, this architecture achieves high efficiency in haze removal while employing a limited number of convolutional filters in its convolutional layers and having relatively low depth. This design choice reduces the network's complexity, mitigating overfitting and decreasing its time complexity. The model utilizes a composite loss function to ensure accurate and effective training. This combined loss function incorporates MSE, SSIM, and L1 loss functions, each contributing to the optimization process by focusing on different aspects of the predicted output, ensuring both pixel-wise accuracy and structural coherence in the results. Figure 3 depicts the proposed architecture of the autoencoder network for image dehazing. As shown in Figure 3, the proposed architecture comprises 13 layers, including six convolutional layers, six ReLU layers, and one dropout layer in the encoder section. Following each ReLU layer in the encoder, a CBAM is implemented to select the most relevant feature channels and influential regions within those features. The decoder section consists of 13 layers, comprising seven convolutional layers and six ReLU layers. As stated, the encoder and decoder sections have equal sizes. Furthermore, the number of convolutional filters in this architecture is limited to between 13 and 31, resulting in lower complexity and fewer convolutional filters compared to previous architectures. According to Figure 3, the convolutional filters of the first two layers have a size of 11. This choice is motivated by the fact that image dehazing is related to the color features of the image. By using a filter size of 11, we aim to retain low-level image features during the training process of the convolutional autoencoder network.

3.3 Implementation details

We implemented and trained the proposed architecture using Python software. Each experiment was performed using 1000 epochs and batch sizes of 64. Adam optimizer and training rate of 0.001 were used to train the network. The datasets are divided into 70% and 30% of the training and test sets, respectively.

3.4 datasets

To train and test the designed convolutional autoencoder network architecture, we used NH-haze and Realistic Single Image Dehazing (RESIDE) datasets. The NH-haze dataset contains 55 images of the outdoor environment in two sections, clear and hazy. This dataset is the first dataset that consists of non-homogeneous images. Although this dataset has few images, it is one of the most challenging datasets for image dehazing models due to the close distance to the haze and the high concentration of haze [23]. RESIDE dataset includes indoor and outdoor parts. The indoor section contains 1449 images of exterior views and the outdoor section contains 492 images from inside the building in two groups: clear and hazy [24]. Examples of the used datasets images are shown in Figure 4.

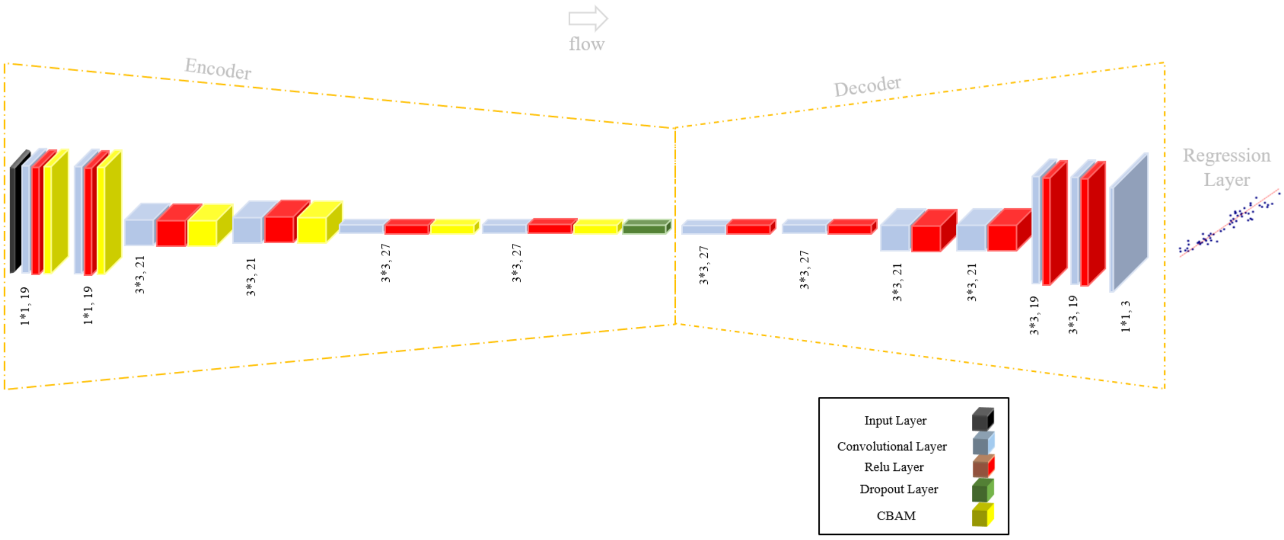

Figure 3. Our proposed architecture of Convolutional Autoencoder network with CBAM for image dehazing.

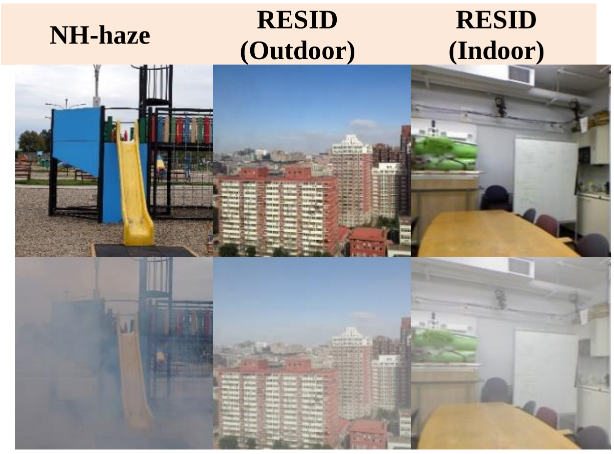

Figure 4. Examples of the used datasets. Below are hazy images and above are corresponding clean images.

4. Ablation Studies

This section discusses the results obtained from implementing the proposed method on various datasets. The proposed method's performance is evaluated using PSNR and SSIM metrics. Initially, the proposed model's results on the NH-haze and RESIDE datasets are presented visually. Subsequently, the model's performance, as measured by these metrics, is compared to state-of-the-art methods.

4.1 Results on the RESIDE Dataset

This section investigates the performance of the proposed model for image dehazing on images from the RESIDE dataset. Table 1 presents the average PSNR and SSIM values of the proposed model when tested on the RESIDE dataset, with and without CBAM. The error optimization process in this section utilized the Adam optimizer. As evident from these results, the proposed method demonstrates high efficiency in image dehazing.

4.2 Results on the NH-haze Dataset

This section presents the results of the proposed model for image dehazing on images from the NH-haze dataset. Figure 7 shows the subjectively dehazed images. Table 2 presents the average PSNR and SSIM values of the proposed model on the NH-haze dataset, with and without CBAM. Due to the presence of thick haze and close-up shots in this dataset, the proposed method achieves good performance by effectively reducing haze in these image types.

Table 1. The average PSNR and SSIM values of the proposed model on RESIDE dataset without and with CBAM.

	Without CBAM		With CBAM	
Section	PSNR	SSIM	PSNR	SSIM
Outdoor	37.87	0.9863	38.96	0.9891
Indoor	38.48	0.9867	38.54	0.9886

Table 2. The average PSNR and SSIM values of the proposed model on NH-haze dataset without and with CBAM.

	Without CBAM		With CBAM	
dataset	PSNR	SSIM	PSNR	SSIM
NH-haze	19.90	0.7249	20.21	0.7311

4.3 Effect of the Loss Function

This section examines the impact of the proposed loss function on the efficiency of the proposed architecture with CBAM for image dehazing. Table 3 shows the model's efficiency in image dehazing with and without the proposed loss function (using the Adam objective function) on the datasets used in this study. Figure 5 illustrates the effect of varying numerical values for the coefficients of the proposed loss function. Note that, due to space constraints, only a portion of the results from the outdoor section of the RESIDE dataset are shown.

Table 3. The average PSNR and SSIM values of the proposed model with CBAM on used datasets without and with proposed Loss function.

	Adam & CBAM		Proposed Loss Function & CBAM	
Section	PSNR	SSIM	PSNR	SSIM
Outdoor	38.96	0.9891	43.71	0.9986
Indoor	38.54	0.9886	41.42	0.9941
NH-haze	20.21	0.7311	22.06	0.7532

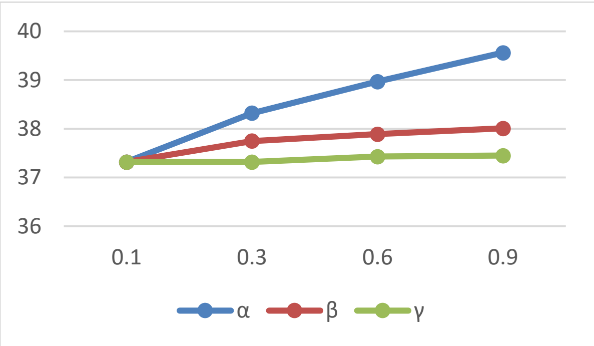

Figure 5. Effect of different numerical values for the coefficients of the proposed loss function.

As depicted in Figure 5, the numerical values of the coefficient associated with the loss function have the greatest impact, while the numerical values of the coefficient associated with the regularization term have the least impact on the efficiency of the proposed method and parameter optimization. This may be because, in image dehazing, accurately matching the color characteristics of the generated image to the original image is paramount. The emphasis on color fidelity ensures that the dehazed images maintain visual realism, which is critical for applications like autonomous driving and surveillance. Figures 6, 7, and 8 visually and subjectively demonstrate the results of image dehazing on the outdoor and indoor sections of the RESIDE dataset and the NH-haze dataset using the proposed architv ecture with CBAM and the proposed loss function. These figures highlight the model's ability to effectively remove haze while preserving fine details and natural color gradients. Quantitative metrics, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), further validate the superior performance of the proposed method compared to baseline approaches.

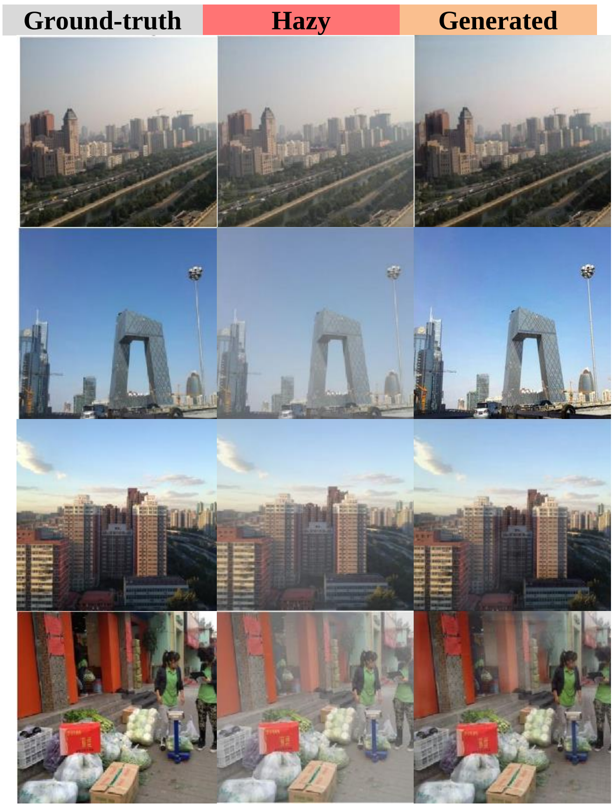

Figure 6. The results of image dehazing on Outdoor section of the RESIDE dataset using proposed method.

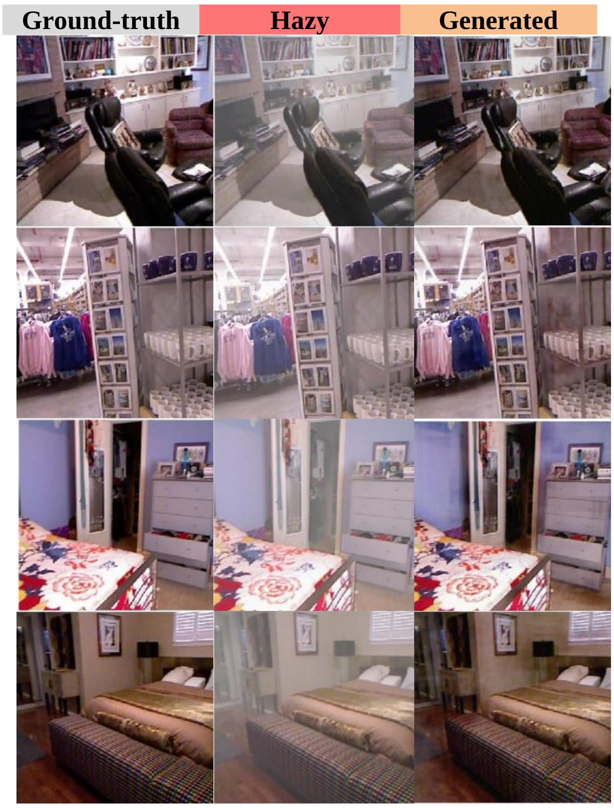

Figure 7. The results of image dehazing on Indoor section of the RESIDE dataset using proposed method

Figure 8. The results of image dehazing on NH-haze dataset using proposed method.

4.4 Effects of Convolutional Filters

This section investigates the influence of the number and size of convolutional filters in the convolutional autoencoder network architecture on image dehazing. Experiments conducted to determine the optimal convolutional autoencoder architecture for image dehazing revealed that increasing the number and size of convolutional filters in a controlled manner is beneficial, while excessive increases or decreases can reduce the architecture's efficiency. This may be because image dehazing relies on low-level features and image color features. Therefore, the complexity and depth of the extracted features should be carefully controlled. Figure 9 illustrates the effect of increasing the number of convolutional filters on the performance of the convolutional autoencoder network in image dehazing on the outdoor section of the RESIDE dataset. As shown, excessively increasing the number of filters can diminish the convolutional autoencoder's efficiency in image dehazing. The best performance in this figure is observed when the number of filters ranges from 19 (the first filter in the encoder section) to 27 (the last filter in the encoder section). It is important to note that this number of filters was chosen empirically. The numerical intervals in this figure represent the number of filters in the first convolutional layer of the encoder section and the number of filters in the last convolutional layer of the encoder section. Furthermore, the balance between filter count and computational efficiency ensures practical deployment in real-time applications. These findings underscore the importance of empirical tuning to optimize dehazing performance while maintaining resource efficiency.

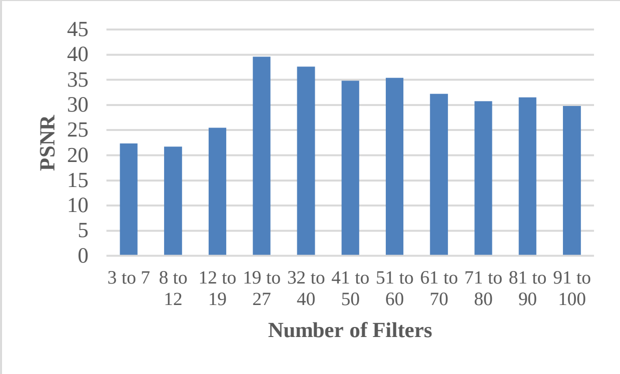

Figure 9. The results of image dehazing on outdoor section of the RESIDE dataset using proposed method with different number of convolutional filters.

Similarly, the size of the convolutional filter significantly affects the efficiency of the convolutional autoencoder architecture in image dehazing. According to experiments, the optimal filter size for the proposed architecture is within the range of 1 to 3. Filter sizes larger than this range can lead to the loss of low-level features and color information, ultimately decreasing model performance. Figure 10 shows the effect of increasing the size of convolutional filters on the efficiency of the convolutional autoencoder architecture in image dehazing on the indoor section of the RESIDE dataset. It is important to note that the numerical intervals in this figure represent the size of filters in the first convolutional layer of the encoder section and the number of filters in the last convolutional layer of the encoder section.

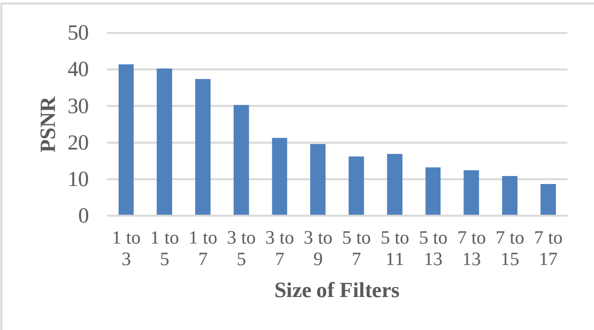

Figure10. The results of image dehazing on indoor section of the RESIDE dataset using proposed method with different size of convolutional filters.

4.5 Comparison with Other Methods

This section compares the proposed method's results with those of previous research models. Figures 11 and 12 provide a visual and subjective comparison between the proposed method and other methods using a test image from the indoor and outdoor sets of the RESIDE dataset, respectively. Figure 13 presents a subjective comparison between the proposed method and other methods using a test image from the NH-haze dataset. The compared methods are described below.

As shown in Figures 11, 12, and 13, Guo et al. [25] proposed a method called "Image Dehazing Transformer with Transmission-Aware 3D Position Embedding" (Dehamer) for image dehazing. Their approach combines CNNs and transformers to enhance image dehazing capabilities. They leverage the transformer's modeling power and the CNN's local representation capacity. Additionally, they introduced a transformer-based module for effectively transferring and integrating meaningful information.

Xu et al. [26] developed a Feature Fusion Attention Network (FFA-net) for dehazing. Their end-to-end feature fusion attention network utilizes a convolutional layer, three group structures, a concatenate layer, a Channel Attention (CA) module, a Pixel Attention (PA) module, and two convolutional layers.

Liu et al. [27] proposed an Attention-based Multi-scale Network for dehazing, named GridDehaze-net. Their architecture comprises three sections: preprocessing, backbone, and post-processing. In the preprocessing section, they employ a convolutional layer without an activation function and a residual dense block (RDB), which generates 16 feature maps for training. The backbone utilizes attention-based mechanisms to manage and refine input information for training. Each RDB consists of five convolutional layers, the first four layers increasing feature maps, and the last layer using attention to integrate them.

Dong et al. [28] proposed a Multi-scale Boosted Dehazing Network (MSBDN) based on an encoder-decoder network. Their architecture includes CBlock modules, distillation modules, attention modules, short skip connections, improvement modules, and local skip connections. The encoder section uses convolutional layers to extract image features and enhances this information using CBlocks and attentions. Local skip connections are used for data integration and downsampling. Finally, attention, CBlock modules, and upsampling convolutions are used to reproduce information and upsample in the decoder section.

Yi et al. [29] proposed a Multi-scale Topological Network (MSTN) for image dehazing. Unlike other methods that rely on attention modules for information integration and selection, MSTN allows the integration of features with different sizes. It utilizes a multi-branch structure where i represents the number of rows (network depth) and j represents the scale of the model. Each branch is responsible for extracting features at different scales. Finally, skip connections are used to separate or interact features with different scales.

Another method called Dehazing Using Progressive Feature Fusion (SGID-PFF) for image dehazing was proposed by Bai et al. [30]. This method, like others, is based on CNNs and integrates information from different parts for image dehazing. The architecture includes four main parts: Depth Estimation Generator, Depth Estimation Discriminator, Dehaze Generator, and Dehaze Discriminator. The Dehaze Generator is responsible for generating haze-free images. The Depth Estimation Generator helps estimate the depth of feature maps accurately, and the Dehaze Generator utilizes these maps. Finally, the Dehaze Discriminator identifies the difference between correct and fake information produced. Bai and colleagues' architecture resembles U-net, consisting of 9 layers in the encoder section and 9 layers in the decoder section, using leakyReLU as the activation function in all layers.

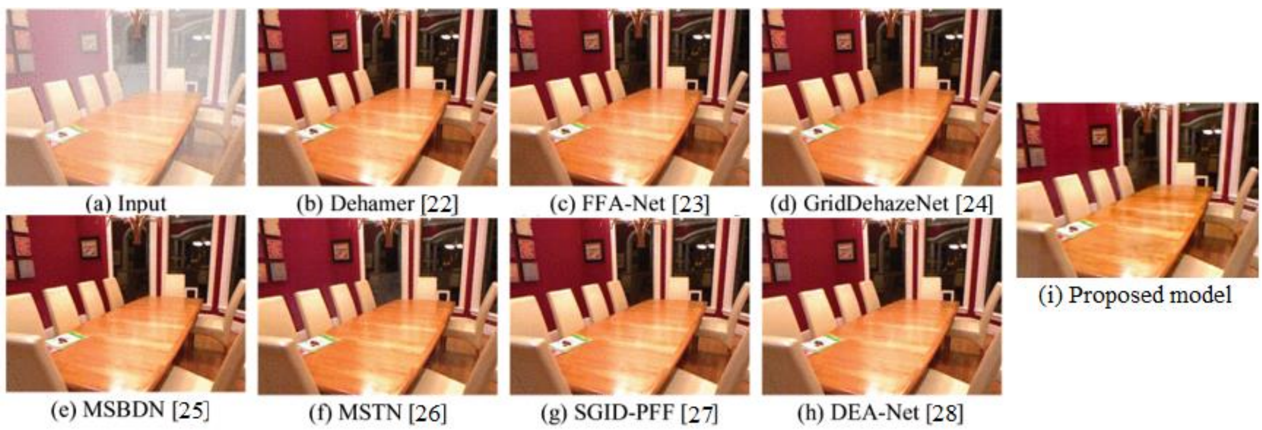

Figure11. Visual comparison of an image dehazing from the Indoor testing image.

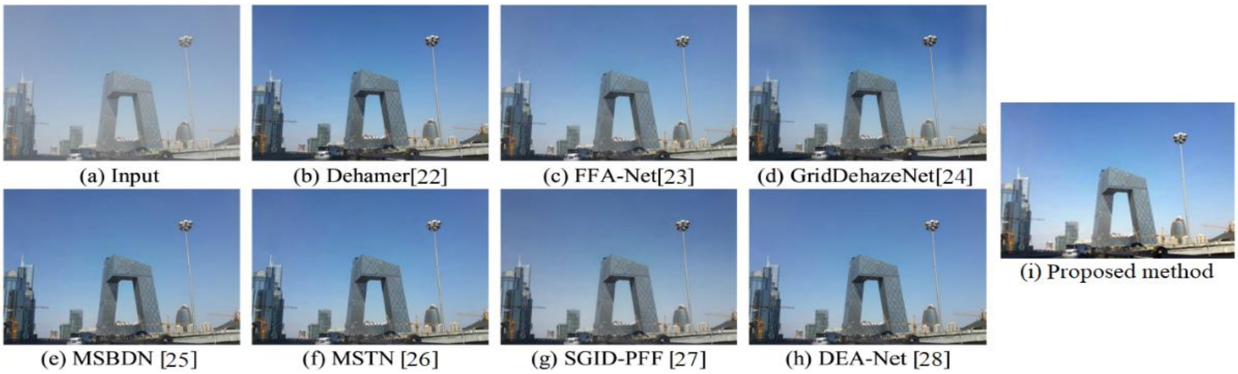

Figure 12. Visual comparison of an image dehazing from the Outdoor testing image.

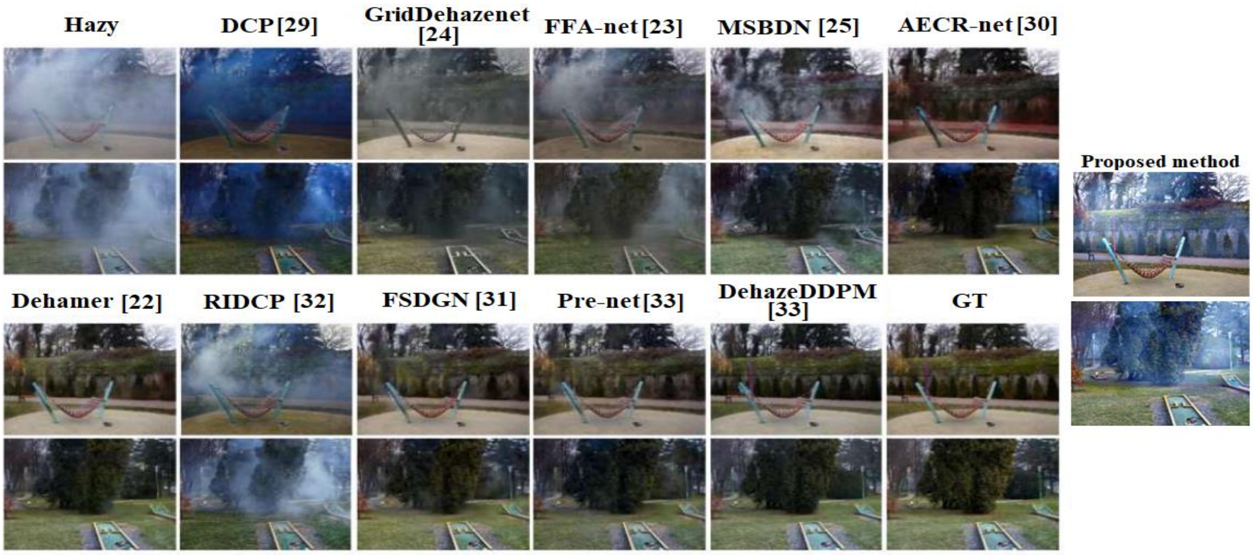

Figure 13. Visual comparison of an image dehazing from the NH-hazy testing image.

Chen et al. [31] proposed DEA-Net, another method for image dehazing based on CNNs. This model, like other proposed models, includes three parts: encoder, transformation module, and decoder. The key element of this network is the use of attention blocks to integrate information. They replaced the convolutional layers with Detail Enhancement Convolution layer (DEConv) and claimed it produces stronger representations compared to previous convolutional layers.

He et al. [32] proposed a dark channel prior (DCP) method for image dehazing. This method assumes that, except for sky pixels, all other pixels have a minimum RGB intensity value. Therefore, they aim to intensify these values.

Wu et al. [33] proposed AECR-Net for image dehazing using an autoencoder network. They also developed a loss function that aligns ground-truth images with clear images while creating a large distance from hazy images. Wu et al. [34] proposed a framework for image dehazing using a frequency and spatial dual guidance network (FSDGN). They proposed a CNN architecture that combines convolutional layer features and Fourier transform features for image dehazing. Yu et al. [35] proposed Real Image Dehazing network via high-quality Codebook Priors (RIDCP) for image dehazing. They designed a new autoencoder architecture. Initially, they used VQGAN, a pretrained network, to extract features and codebooks, then generated a haze-free image with a new decoder architecture. Yu et al. [36] proposed the DehazeDDPM framework based on Denoising Diffusion Probabilistic Mode. DehazeDDPM removes haze from the image in two steps. The first step utilizes the Atmospheric Scattering Model (ASM) to guide the data distribution towards clear data. The second step reproduces the data lost due to haze.

Table 4 presents an objective comparison of the proposed method with other methods on the used datasets. As shown in Table 4, the proposed method achieved first place on the RESIDE (indoor) and RESIDE (outdoor) datasets and second place on the NH-Haze dataset with a small margin.

Table 4. Objective comparison between proposed methods.

Dataset	Work	PSNR	SSIM
RESIDE(Indoor)	Dehamer [22]	36.63	0.9881
	FFA-Net [23]	36.39	0.9886
	GridDehazeNet [24]	32.16	0.9836
	MSBDN [25]	32.61	0.9799
	MSTN [26]	35.64	0.9894
	SGID-PFF [27]	38.31	0.9860
	DEA-net [28]	41.16	0.9937
	Proposed method	41.42	0.9941
RESIDE(Outdoor)	Dehamer [22]	35.07	0.9849
	FFA-Net [23]	33.57	0.9837
	GridDehazeNet [24]	30.75	0.9811
	MSBDN [25]	34.71	0.9849
	MSTN [26]	32.91	0.9831
	SGID-PFF [27]	30.16	0.9803
	DEA-net [28]	36.14	0.9882
	Proposed method	43.71	0.9986
NH-HAZE	DCP [29]	12.72	0.4410
	GridDehaze [24]	13.80	0.5370
	FFA-net [23]	18.13	0.6515
	MSBDN [25]	17.97	0.6647
	AECR-net [30]	19.88	0.7170
	Dehamer [22]	20.66	0.6844
	RIDCP [31]	13.52	0.552
	FSDGN [32]	19.99	0.7306
	DehazeDDPM [33]	22.28	0.7309
	Proposed method	22.06	0.7532

Compared to the approaches reviewed in this section, particularly those in [31] and [36], which have demonstrated comparable performance, the proposed method exhibits several strengths:

1. The model requires significantly fewer parameters, making it more suitable for deployment on resource-constrained devices.
2. The proposed architecture exhibits greater robustness across diverse weather and lighting conditions, as demonstrated by our additional qualitative results.
3. The integration of attention mechanisms enhances the interpretability of feature selection during the dehazing process.
4. The model maintains high performance while reducing computational costs, making it more suitable for real-time applications.

5. Conclusion

One of the key challenges leading to the failure of digital images and diagnostic systems based on digital image processing is the presence of haze in these images. To address this issue, researchers have proposed numerous methods leveraging machine vision and deep learning. In this paper, we present a novel deep learning-based approach for image dehazing. Our method employs self-supervised learning and introduces a new convolutional encoder architecture incorporating the Convolutional Block Attention Module (CBAM), along with a custom loss function to optimize learning parameters. We evaluated the effectiveness of the proposed architecture on multiple datasets, demonstrating its high performance in dehazing digital images. Furthermore, the results indicate that a well-designed convolutional encoder network—without excessive complexity—can partially alleviate the reliance on high-performance hardware for digital image dehazing, addressing a critical need in the research community.

References

[1] B. S. Chaitanya, S. Mukherjee. "Single image dehazing using improved cycleGAN." *Journal of Visual Communication and Image Representation*, Vol. 74, pp. 103014, 2021.

[2] S. Zhang, X. Zhang, S. Wan, W. Ren, L. Zhao, L. Shen. "Generative adversarial and self-supervised dehazing network", *IEEE Transactions on Industrial Informatics*, Vol. 20, no. 3, pp. 4187-4197, March 2024.

[3] G. D'Angelo, F. Palmieri. "Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction", *Journal of Network and Computer Applications,* Vol. 173, pp. 102890, Jan. 2021.

[4] C. Y. Jeong, K. Moon, M. Kim. "An end-to-end deep learning approach for real-time single image dehazing", *Journal of Real-Time Image Processing*, Vol. 20, pp. 12, 2023.

[5] R. R. Choudhary, K. K. Jisnu, G. Meena. "Image dehazing using deep learning techniques", *Procedia Computer Science*, Vol. 167, pp. 1110-1119, 2020.

[6] C. A. Hartanto, L. Rahadianti. "Single image dehazing using deep learning", *JOIV: International Journal on Informatics Visualization*, Vol. 5, pp. 76-82, 2021.

[7] F. A. Dharejo, Y. Zhou, F. Deeba, M. A. Jatoi, M. A. Khan, G. A. Mallah, X. Wang. "A deep hybrid neural network for single image dehazing via wavelet transform", *Optik*, Vol.231, pp. 166462, 2021.

[8]S. Li, Q. Yuan, Y. Zhang, B. Lv, F. Wei. "Image dehazing algorithm based on deep learning coupled local and global features", *Applied Sciences*, Vol. 12, pp. 8552, 2022.

[9] G. H. Babu, V. K. Odugu, N. Venkatram, B. Satish, K. Revathi, B. J. Rao. "Development and performance evaluation of enhanced image dehazing method using deep learning networks", *Journal of Visual Communication and Image Representation*, Vol. 97, pp. 103976, 2023.

[10] M. Shakeri, M. H. Dezfoulian, H. Khotanlou. "Density-based histogram partitioning and local equalization for contrast enhancement of images". *Journal of AI and Data Mining*, Vol. 6, pp. 1–12, 2018.

[11] C. Wang, Z. Li, J. Wu, H. Fan, G. Xiao, H. Zhang. "Deep residual haze network for image dehazing and deraining", *IEEE Access*, Vol. 8, pp. 9488-9500, 2020.

[12] G. H. Babu, N. Venkatram. "ABF de-hazing algorithm based on deep learning CNN for single I-Haze detection", *Advances in Engineering Software*, Vol. 175, pp. 103341, 2023.

[13] C. Hodges, M. Bennamoun, H. Rahmani. "Single image dehazing using deep neural networks", *Pattern Recognition Letters*, Vol. 128, pp. 70–77, 2019.

[14] S. Yin, X. Yang, Y. Wang, Y. H. Yang. "Visual attention dehazing network with multi-level features refinement and fusion", *Pattern Recognition*, Vol. 118, pp 108021, 2021.

[15] S. Yin, Y. Wang, Y. H. Yang. "A novel image-dehazing network with a parallel attention block", *Pattern Recognition*, Vol. 102, pp. 107255, 2020.

[16] P. K. Balla, A. Kumar, R. Pandey. "A 4-channelled hazy image input generation and deep learning-based single image dehazing", *Journal of Visual Communication and Image Representation*, Vol. 100, pp. 104099, 2024.

[17] K. Hu, F. Wu, Z. Zhan, J. Luo, H. Pu. "High-low level task combination for object detection in foggy weather conditions", *Journal of Visual Communication and Image Representation*, Vol. 98, pp 104042, 2024.

[18] C. H. Yeh, C. H. Huang, L. W. Kang, M. H. Lin. "Single image dehazing via deep learning-based image restoration", in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov 2018, pp. 1609–1615.

[19] W. Ren, J. Pan, H. Zhang, X. Cao, M. H. Yang. "Single image dehazing via multi-scale convolutional neural networks with holistic edges", *International Journal of Computer Vision*, Vol. 128, pp. 240–259, 2020,

[20] Y. Zhang. "A better autoencoder for image: Convolutional autoencoder", *ICONIP17-DCEC*, Mar 2018. Available online: http://users.cecs.anu.edu.au/Tom.Gedeon/conf/ABCs2018/paper/ABCs2018_paper_58.pdf (accessed on 23 March 2017), 2018.

[21] S. Woo, J. Park, J. Y. Lee, I. S. Kweon. "CBAM: Convolutional block attention module", in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[22] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, T. S. Chua. "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5659–5667.

[23] C. O. Ancuti, C. Ancuti, R. Timofte. "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 444–445.

[24] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, Z. Wang. "Benchmarking single-image dehazing and beyond", *IEEE Transactions on Image Processing*, Vol. 28, pp. 492–505, 2018.

[25] C. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, C. Li. "Image dehazing transformer with transmission-aware 3D position embedding", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2022, pp. 5812–5820.

[26] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia. "FFA-Net: Feature fusion attention network for single image dehazing", in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, pp. 11908–11915.

[27] X. Liu, Y. Ma, Z. Shi, J. Chen. "GridDehazeNet: Attention-based multi-scale network for image dehazing", in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019, pp. 7313–7322.

[28] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang. "Multi-scale boosted dehazing network with dense feature fusion", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020, pp. 157–2167.

[29] Q. Yi, J. Li, F. Fang, A. Jiang, G. Zhang. "Efficient and accurate multi-scale topological network for single image dehazing", *IEEE Transactions on Multimedia*, Vol. 24, pp. 3114–3128, 2022.

[30] H. Bai, J. Pan, X. Xiang, J. Tang. "Self-guided image dehazing using progressive feature fusion", *IEEE Transactions on Image Processing*, Vol. 31, pp. 1217–1229, 2022.

[31] Z. Chen, Z. He, Z.-M. Lu. "DEA-net: Single image dehazing based on detail-enhanced convolution and content-guided attention", arXiv: 2301.04805, 2023.

[32] K. He, J. Sun, X. Tang. "Single image haze removal using dark channel prior", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, pp. 2341–2353, 2010

[33] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, L. Ma. "Contrastive learning for compact single image dehazing", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10551–10560.

[34] R. Q. Wu, Z. P. Duan, C. L. Guo, Z. Chai, C. Li. "Ridcp: Revitalizing real image dehazing via high-quality codebook priors", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22282–22291.

[35] H. Yu, N. Zheng, M. Zhou, J. Huang, Z. Xiao, F. Zhao. "Frequency and spatial dual guidance for image dehazing", in *European Conference on Computer Vision*, 2022, pp. 181–198.

[36] H. Yu, J. Huang, K. Zheng, M. Zhou, F. Zhao. "High-quality image dehazing with diffusion model", arXiv: 2308.11949, 2023.

حذف مه از تصویر با استفاده از یک شبکه خودرمزگذار با بلوک توجه یکپارچه پیچشی

همایون رستگار و حسن ختن لو*

گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه بوعلی سینا، همدان، ایران.

ارسال ۲۰۲۵/۰۴/۱۶؛ بازنگری ۲۰۲۵/۰۵/۲۶ پذیرش ۲۰۲۵/۰۷/۰۴

چکیده:

یکی از چالش‌های پردازش تصویر دیجیتال، وجود مه است که به‌ویژه در محیط‌های مرطوب و بارانی شایع می‌باشد. نمونه‌هایی از سیستم‌های مبتنی بر هوش مصنوعی که تحت تأثیر این چالش قرار می‌گیرند، شامل دوربین‌های کنترل هوشمند ترافیک، خودروهای خودران، سیستم‌های کمک‌داوری ویدئویی (VAR) در استادیوم‌های فوتبال و دوربین‌های نظارتی و امنیتی می‌شود. این مقاله روشی برای کاهش مه با استفاده از یادگیری خودنظارتی (SSL) و یادگیری عمیق پیشنهاد می‌دهد. یک شبکه خودکدگذار پیچشی (CAN) همراه با ماژول توجه بلوک پیچشی (CBAM) توسعه داده شده است تا مه را از تصاویر حذف کند. مزیت این روش در تعداد کمتر لایه‌ها و فیلترها در مقایسه با مدل‌های قبلی و همچنین استفاده از CBAM برای تأکید بر کانال‌های پیچشی و نواحی مهم تصویر است. آزمایش‌ها نشان می‌دهند که فیلترهای پیچشی بیش‌ازحد، با وجود هدف تولید ویژگی‌های متنوع، می‌توانند توانایی مدل در حذف مؤثر مه از تصاویر را مختل کنند. بنابراین، تعداد فیلترها باید به دقت محدود شود. یک تابع زیان ترکیبی برای آموزش معماری پیشنهادی به کار گرفته شد که روی مجموعه داده‌های NH-haze و مجموعه داده واقع‌گرایانه حذف مه از تصویر تکی (RESIDE) ارزیابی گردید. برای ارزیابی، معیار شباهت ساختاری (SSIM) و نسبت سیگنال به نویز پیک (PSNR) استفاده شد. نتایج آزمایش‌ها نشان می‌دهند که معماری پیشنهادی در مقایسه با روش‌های پیشرفته موجود، عملکرد بالاتری دارد.

کلمات کلیدی: حذف مه، شبکه خودرمزگذار پیچشی، یادگیری عمیق، یادگیری خودناظر، بلوک توجه یکپارچه پیچشی.