Complex Engineering Systems

Research Article

Open Access

A study on non-uniform image dehazing algorithm based on serialized integrated attention and multi-dimensional transformer

Tianhao Bai, Ji Qiu

College of Physics and Electronic Information, Guangxi Minzu University, Nanning 530006, Guangxi, China.

Correspondence to: Dr. Ji Qiu, College of Physics and Electronic Information, Guangxi Minzu University, University East Road No. 188, Nanning 530006, Guangxi, China. E-mail: qiuji@gxmzu.edu.cn; ORCID: 0000-0002-2384-501X

Abstract

To address the issues of detail loss and blurred restoration in the non-uniform image dehazing process of existing non-uniformly hazy images, this paper presents a novel non-uniform image dehazing algorithm based on serialized integrated attention and multi-dimensional Transformer. This approach aims to restore clear, detailed scenes from heavily hazy images. Firstly, a serialized integrated attention module is established to capture image features. This module amalgamates spatial and channel attention mechanisms and is applied to the shallow-layer network. It effectively concentrates on the local features of the image in both spatial and channel dimensions. Secondly, a multi-dimensional Transformer module is incorporated into the deep-layer network to extract global information and reduce information loss during feature extraction. Finally, feature network fusion is carried out to adaptively fuse the feature maps of the shallow layer and the deep layer. This allows the model to take into account local and global information, combine the detailed local features of the shallow layer with the broad global information of the deep layer, and capture fine-grained details while integrating the image context. The experimental results clearly demonstrate the effectiveness of the proposed algorithm. On the I-HAZE, O-HAZE, and NH-HAZE non-uniform haze datasets, the algorithm achieves Peak Signal-to-Noise Ratio values of 22.86, 25.86, and 22.06, along with Structural Similarity Index Measurement values of 0.8731, 0.7799, and 0.7796, respectively. Moreover, the effectiveness of this algorithm is verified on real-world hazy images. Compared with other dehazing algorithms, our proposed method outperforms them in both visual effects and objective metrics.

1. INTRODUCTION

In modern society, computer vision and image processing technologies are extensively applied across various domains, from security surveillance to autonomous driving, and even to everyday entertainment. High-quality image information is indispensable in all these applications. However, under adverse weather conditions such as haze, image quality often deteriorates, causing a loss of detail and posing significant challenges for related applications. Therefore, research on image dehazing algorithms holds substantial practical significance.

Currently, in the domain of dehazing, traditional approaches mainly include image enhancement and restoration, which were among the earliest techniques used for image dehazing. Image enhancement dehazing algorithms are designed to improve image quality by specifically reducing or eliminating the degradation caused by weather conditions such as fog and haze. These algorithms typically enhance aspects of the image such as contrast, brightness, color, and detail, making the image visually clearer for human observation and analysis. Examples include histogram equalization algorithms [1] and the color attenuation prior [2]. Conversely, image restoration dehazing algorithms focus on a deep understanding of the physical mechanisms underlying image degradation and use this understanding to establish atmospheric scattering models. These models often rely on certain prior knowledge, such as the assumption that scene depth information changes gradually over large areas or that local regions of the image exhibit consistency. By accurately estimating the parameters of these models, researchers can infer the image content as it would appear under ideal, fog-free conditions, thereby restoring the hazy image. For example, Dark Channel Prior (DCP) [3] is based on the observation that, in a fog-free outdoor image, the majority of pixels in its dark channel (the minimum value across the color channels for each pixel) are very close to zero. This prior knowledge can be used to estimate the transmission map in the atmospheric scattering model, which then allows for the recovery of a fog-free image. However, this method may not perform well on images with dense fog or more complex fog effects and tends to produce halo effects in bright areas such as the sky. Liu *et al.* proposed a multi-purpose dehazing framework for nighttime hazy images [4]. They mainly constructed a non-linear model based on the Retinex theory to describe various adverse degradation situations of nighttime hazy images. A prior dehazing method was used to remove the haze in the illumination component. However, if the prior assumptions do not match the actual image conditions, such as the presence of special lighting or abnormal haze distribution in the scene, the dehazing effect may not be satisfactory, and problems such as halos and color distortion may occur.

In recent years, with the rise of deep learning technologies, significant advancements have been made in deep learning-based image dehazing algorithms [5]. Deep learning models learn the mapping relationship for image dehazing by training on large datasets, enabling them to automatically extract features from images and optimize dehazing effects [6]. These methods offer greater flexibility and accuracy, allowing them to better adapt to various complex scenes and weather conditions. Convolutional Neural Networks (CNN) are the most commonly used models in this domain. They extract feature information from images through multiple layers of convolution and pooling operations [7]. By training CNN models on hazy images, these models can detect and remove haze-related features, thereby restoring clear images. Additionally, attention mechanisms have become a popular technique for improving the effectiveness of dehazing on hazy images [8]. Li *et al.* proposed a non-uniform dehazing algorithm based on improved ConvNeXt [9]. Although the ImageNet-pre-trained ConvNeXt model was adopted to supplement knowledge, when dealing with small non-uniform haze datasets, there may still be an overfitting problem. This leads to a decline in the generalization ability of the model on new, unseen data, resulting in unstable dehazing effects. The dehazing module obtains more image information through the Feature Fusion Group (FFG) [10], which reduces color distortion and artifacts in the dehazed

images. Then, the obtained image information is passed to the Deep Normalized Corrected Convolution Block (DNCC) to reduce covariate shift, making the model easier to train.

However, CNN has its own limitations. Its receptive field is relatively limited. Although the receptive field can be expanded by stacking multiple convolutional layers, it still struggles to capture the global information of images when dealing with complex image dehazing tasks. It is difficult for CNN to effectively grasp the distribution law of fog in the entire image and the long-range dependency relationships between various objects in the scene. Different from CNN, the Transformer architecture, with its self-attention mechanism at the core, has certain advantages in feature extraction. The self-attention mechanism allows the model to calculate the dependency relationships between each position in the sequence in parallel, breaking the limitation of locality, thus enabling the model to conduct global-perspective modeling and analysis of the image.

To this end, recently, Song *et al*. proposed the DehazeFormer de-fogging framework[11]. On the commonly used Standard Objective Testing Set (SOTS) -indoor dataset, it outperformed Feature Fusion Attention Network (FFANet) with only 25% of the parameters and 5% of the computational cost, surpassing most de-fogging algorithms. However, this algorithm also has certain drawbacks, such as high computational complexity and a high degree of dependence on data. In order to give full play to the advantages of both CNN and Transformer and overcome their respective shortcomings, hybrid algorithms of Transformer and CNN emerged. Subsequently, Wang *et al*. proposed a de-fogging algorithm based on the fusion of dual-attention convolution and Transformer for image de-fogging[12]. However, it has deficiencies in the restoration of image de-fogging details. Recently, Wang *et al*. proposed the GridFormer image de-fogging algorithm[13]. This research focused on the problems of image blurring and low contrast in severe weather, and innovatively proposed a residual dense Transformer model with a grid structure. Through the unique grid-structure design, this model effectively integrates multi-scale information and improves the feature-extraction ability. Jiang *et al*. introduced the Mutual Retinex method, which combines transformers and CNNs to enhance image quality by capturing both global and local features[14]. This approach improves image enhancement performance, particularly in challenging conditions, but its higher computational cost and the complexity of training may limit its practical application in real-time systems. Zheng *et al*. proposed a dehazing network named T-Net to address the problem of single image dehazing[15]. Stacked T-Nets use a recursive strategy to explore complex feature relationships in the image and improve the de-fogging effect. However, the T-Net framework is complex, and the training and testing costs are high in devices or scenarios with limited computing resources. Qiu *et al*. introduced the Multi-scale Attention Refinement (MSAR) module to correct the error of the Taylor expansion of softmax-attention, enabling the model to more accurately process image information during the de-fogging process, reducing the error caused by approximation and improving the de-fogging accuracy[16]. Although there is the MSAR module, the Taylor expansion approximation may still have errors that cannot be completely eliminated in images with complex textures under extreme fog conditions, resulting in less-than-ideal de-fogging effects, and the restoration of details and color accuracy may be affected (references). Moreover, evaluation metrics are also a key point of research. Commonly used metrics such as Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measurement (SSIM) are used to evaluate the performance of image de-fogging algorithms.

Based on the above analysis, this paper proposes a non-uniform image dehazing algorithm based on Serialized Integrated Attention and Multi-dimensional Transformer (SIA-MT), addressing the challenge of recovering detailed information obscured by haze in different regions of real-world non-uniform hazy images. Unlike existing algorithms, the proposed method combines CNN and Transformer[17] to jointly process hazy images for dehazing purposes. It primarily constructs a shallow feature extraction network using spatial and channel dual attention serialized convolution based on CNN, along with a deep feature extraction network centered on the fusion of Transformer modules and Swin Transformer[18] modules to extract both shallow and deep features from hazy images. Furthermore, a multi-feature fusion reconstruction network is designed to obtain

the dehazed image.

2. METHODS

2.1. Peptide topology data representation

2.1.1 Swin transformer

The Transformer architecture achieves global modeling capabilities through its global self-attention mechanism. However, it is less effective at handling multi-scale features. In the original Transformer, multi-head self-attention (MHSA) is applied across all spatial positions, resulting in a computational complexity of $H \times W$ for processing an image of size $O = 2 \times (HW)^2 \times d$. The computational complexity of the original Transformer increases quadratically with the image resolution. Consequently, when processing high-resolution images, the original Transformer becomes less efficient and requires more resources. Moreover, single-scale feature representation has inherent limitations.

To reduce computational complexity and effectively handle multi-scale features, the Swin Transformer employs self-attention within smaller windows rather than across the entire image as in the original Transformer. This approach ensures that the computational complexity of self-attention is fixed as long as the window size remains constant. The computational complexity for the entire image scales linearly with the size of the image. The operations in a Swin Transformer layer are as follows: Given an input feature map, a linear layer projects it into the self-attention queries Q, keys K, and values V matrices, and performs window-based grouping. Swin Transformer applies MHSA within each window, with different partitions for adjacent windows. Thus, the self-attention computation is given in

$$\text{Attention}\,(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{SoftMax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}} + \boldsymbol{B}\right)\boldsymbol{V} \tag{1}$$

Where d represents the number of channels, and B denotes the relative positional bias term. Finally, the output is projected to the final self-attention output through a linear layer.

2.2. Overall framework of the dehazing networks

The non-uniform image dehazing algorithm of serial integrated attention and multi-dimensional Transformer proposed in this paper has an overall structure as shown in Figure 1.

The network primarily consists of four feature extraction network hierarchies and a Gated Fusion Sub-network (GFS)[19]. Among them, the feature extraction network hierarchy includes a concatenated comprehensive attention module and a fusion Transformer module. The concatenated comprehensive attention module is located in the first two layers of the feature extraction stage, aiming to extract shallow features, better preserve and recover image details, and handle both channel and spatial dimensional features to better address non-uniformity. The multi-dimensional Transformer module combines the Transformer and Swin Transformer for deep feature extraction in both channel and spatial dimensions. Leveraging the advantages of global self-attention in Transformer, it enhances global channel features along the channel dimension, while the window-based self-attention of the Swin Transformer reduces computational load and facilitates local information interaction, thereby aggregating global spatial features. The GFS, serving as a fusion mechanism for features from different layers, can adaptively control the fusion ratio of features at different levels, enhancing fine-grained details and complementing key features across layers. Moreover, it integrates the high-resolution spatial information from shallow features and the semantic information from deep features to enhance the model's feature representation capability.
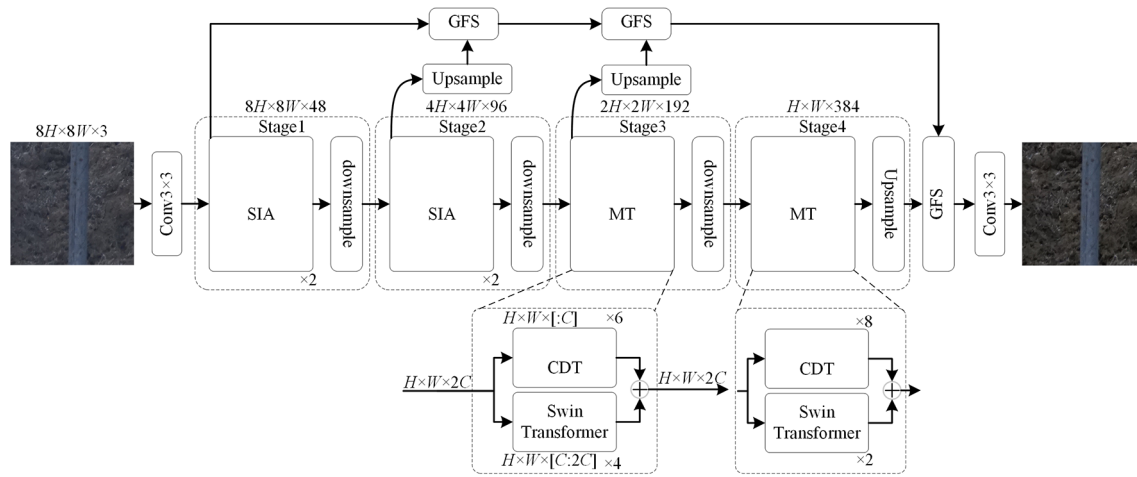
Figure 1. Overall framework of the dehazing network.

In the SIA-MT algorithm, convolutional attention is employed in the shallow network stages where the image spatial scale is large and the number of channels is small, while Transformer is used in the deeper network stages where the image spatial scale is smaller and the number of channels is larger. This approach effectively enhances local fine-grained details in shallow features and global semantic information in deep features, while maintaining good computational efficiency. Additionally, the GFS compensates for missing feature information and strengthens the model's expressive power.

To address the problem of blur in non-uniform image dehazing, this paper attempts to incorporate attention-based recurrent networks to handle haze information. The algorithm first introduces the hazy image into a CNN-based shallow feature extraction network for local feature extraction, generating distinctive feature attention maps in both spatial and channel dimensions. After a series of CNN processes, the resulting feature maps are then input into the Transformer-based deep feature extraction network for global feature extraction.

2.3. Serial integrated attention

In real-world scenarios, hazy images often exhibit non-uniform characteristics with both randomness and uncertainty in their distribution. In contrast, traditional dehazing algorithms primarily address artificially synthesized images with uniform haze. Due to the inherent differences between synthesized and real-world non-uniform haze, traditional methods often fall short in handling the latter effectively. To better restore image details obscured by non-uniform haze in real-world conditions, an attention mechanism was designed under the shallow feature extraction network, integrating spatial and channel attention mechanisms in a series, referred to as serial integrated attention (SIA), to enhance shallow feature extraction.

The proposed SIA module utilizes global average pooling and global max pooling operations to compress spatial and channel information while focusing on the weighted information of different spatial and channel dimensions. This approach compresses the size of the feature maps and removes redundant information, thereby effectively simplifying the model's complexity. The SIA attention mechanism is illustrated in Figure 2. The spatial attention mechanism in the SIA module first weights each spatial position of the input image, focusing on the important regions in the image. By calculating the importance in the spatial dimension, spatial attention can highlight the key areas in the image (such as foreground objects or parts with rich details), thus helping the network avoid processing background noise or irrelevant information. Based on the output of the spatial attention mechanism, the channel attention mechanism further adjusts the weights of different channels, selectively enhancing the feature channels with rich information content or stronger semantic information. It assigns different weighted values according to the importance of the channels, enabling the network to focus
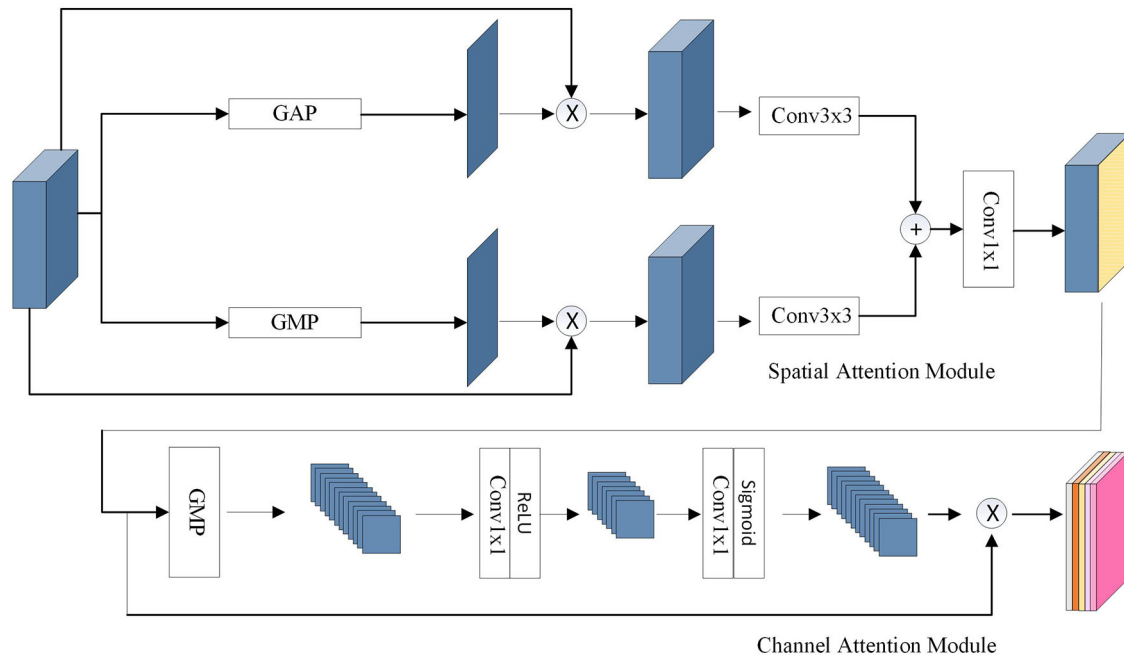
Figure 2. Serial integrated attention (SIA).

on those most discriminative feature channels.

Placing spatial attention first helps to screen out the most important regions in the image initially, preventing the waste of computational resources on irrelevant areas in subsequent operations. This approach is particularly suitable for the image dehazing task. Since haze is usually unevenly distributed in different regions of the image, focusing through spatial attention at the early stage can assist subsequent steps in more accurately restoring details. Under the influence of spatial attention, channel attention can further enhance valuable channel features within important regions, ensuring a more detailed and hierarchical image restoration.

The specific design process of the spatial attention module is as follows: The input is a feature map with a size of $H \times W \times C$, where H and W represent the height and C width of the image, respectively, and C is the number of channels. This feature map contains the preliminary feature information extracted by the shallow-layer network. First, global average pooling and global max pooling are performed on the input feature map in the channel dimension, respectively. By compressing the channel size, we obtain the weighted feature maps corresponding to the average value and the maximum value of all channels at each spatial position, which is convenient for learning spatial features later. Then, the weighted feature maps of the average and maximum values are multiplied pixel-by-pixel with the input feature map respectively to get two output feature maps both of size $H \times W \times C$, after which important spatial regions are assigned higher weights and unimportant regions are weakened. Finally, the two output feature maps are optimized by 3×3 convolutions respectively and then added pixel-by-pixel to obtain the final feature map that combines the advantages of global average pooling and global max pooling, highlighting important spatial regions and significant features in the image and effectively improving the dehazing effect when passed to subsequent network layers.

The specific design process of the channel attention module is that the input is the feature map output from the spatial attention module. First, global max pooling in the spatial dimension is performed on the input feature map, and the obtained weighted feature map is of size $1 \times 1 \times C$, where the value of each channel represents the most significant feature of that channel. Next, a 1×1 convolution and Rectified Linear Unit

(ReLU) function activation are carried out on the weighted feature map to adjust the weights between channels, enhancing the network's attention to important channels and also the model's expressive ability. Then, it goes through 1×1 convolution and Sigmoid function again to further adjust the weights between channels, ensuring that the channel attention mechanism can perform refined weighting, and then maps the weighted value of each channel to the range of [0, 1] to represent the importance of that channel. Finally, the channel-weighted feature map ($1 \times 1 \times C$) is multiplied element-by-element with the original input feature map ($H \times W \times C$). This step weights the input feature map through the weight of each channel in the channel-weighted feature map, highlighting the features of important channels and suppressing unimportant channels.

Supposing $F_X \in R^{H \times W \times C}$ is the input of the serial integrated attention module, this module can be expressed as:

$$F_{c1} = \text{Conv}\left(\left(G_{Avg}\left(F_x\right) \otimes F_x\right), k = 3\right) \tag{2}$$

$$F_{c2} = \text{Conv}\left(\left(G_{Max}\left(F_x\right) \otimes F_x\right), k = 3\right) \tag{3}$$

$$CF_c = \text{Conv}\left(\left(F_{c1} \oplus F_{c2}\right), k = 1\right) \tag{4}$$

$$PF_c = \partial\left(\text{Conv}\left(\varphi\left(\text{Conv}\left(G_{\max}\left(CF_c\right), k = 1\right)\right), k = 1\right)\right) \tag{5}$$

$$F_{cp} = CF_c \otimes PF_c \tag{6}$$

Where ∂ represents the Sigmoid activation function, φ indicates the ReLU function, F_x stands for the input feature map, $\otimes$ denotes the element-wise multiplication, and $\oplus$ signifies the element-wise addition.

2.4. Multi-dimensional transformer module

In the deep feature extraction architecture, as consecutive downsampling operations are performed, the spatial dimensions are progressively reduced, thereby increasing the receptive field of the features and enriching the captured global semantic information. To efficiently extract these global characteristics, the introduction of the Transformer module has become an effective strategy, leveraging its inherent capability for global coarse-grained processing. The core component of the Transformer module is the MHSA mechanism, which does not directly compute self-attention at the pixel level on fine-grained data, but instead cleverly employs a self-attention strategy along the channel dimension. This design significantly reduces computational overhead and enhances overall efficiency while still effectively capturing and integrating global feature information.

The Swin Transformer adopts a hierarchical structural design, where the resolution of feature maps gradually decreases and the number of channels increases as the network deepens, allowing the model to capture multi-scale features from local to global. Therefore, in the deep feature extraction network, an innovative approach combining Transformer in the channel dimension (CDT) with Swin Transformer is employed, referred to as Multi-dimensional Transformer (MT), to extract feature information, as illustrated in Figure 3.

Firstly, $F_t \in R^{H \times W \times 2C}$ given the input feature $F_t \in R^{H \times W \times 2C}$, a 1×1 convolution and a 3×3 depthwise separable convolution are applied to generate the query, representing query-related feature information, the key, representing feature information that matches the query, and the value. The results are then fed into the channel-dimension MHSA calculation. After a series of computations, the output is finally added to the feature F_t input into the Swin Transformer, resulting in the F_{st} feature map.

2.5. Gated fusion subnetwork

The GFS is a feature convolution module that integrates features from different levels[19]. Its core idea is to learn the importance weights of features at various levels and linearly combine these features according to the learned weights, thereby achieving effective feature fusion. Current dehazing algorithms often fail to effectively
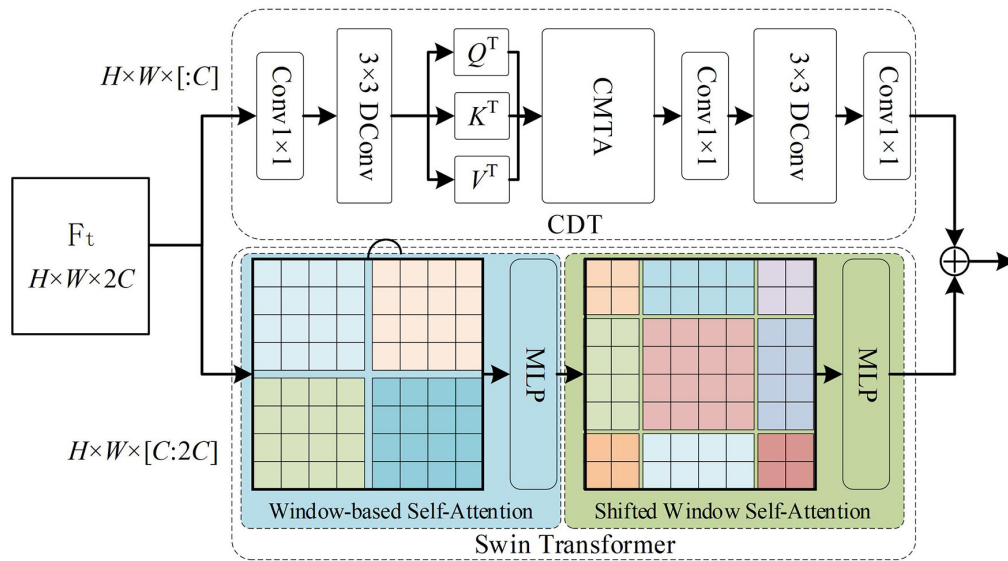
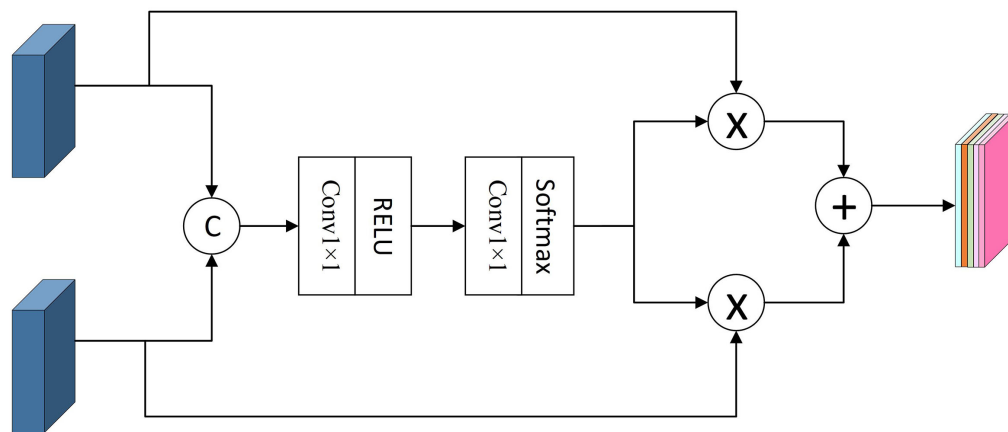**Figure 3.** Multi-dimensional transformer module(MTM).

Figure 4. Gated fusion subnetwork(GFS).

utilize multi-level feature information, leading to excessive feature redundancy and subsequently reducing the performance of dehazing on non-uniform haze images. In response, this paper introduces the GFS to intelligently fuse multi-level feature information, fully leveraging the advantages of each level to enhance the overall performance and output quality of the network. In the task of image dehazing, the GFS effectively combines low-level and high-level feature information, thereby improving image clarity and visual effects, as illustrated in Figure 4.

Firstly, two different levels of feature information are separately input and fused through concatenation. The input features are processed with a convolutional layer 1×1 that adjusts channel parameters, followed by a ReLU activation function to introduce non-linearity into the network. The features are then passed through a subsequent point-wise convolution 1×1 to transform the channels, resulting in channel features containing information from both levels. Next, a Softmax activation function is used to adaptively assign weight information to different channels. These weights are then used to perform adaptive fusion with the features extracted from different layers. Finally, the resulting outputs are summed to further enhance the feature representation.

2.6. Loss function

The loss function plays a central role in machine learning and deep learning, serving as a crucial metric for measuring the discrepancy between the model's predictions and the actual results. It acts not only as a benchmark for evaluating model performance but also as a foundational element in optimizing model parameters, implementing the backpropagation algorithm, and conducting model selection and comparison. In this paper, a combination of pixel-wise reconstruction loss [20] and perceptual loss [21] is employed to enhance the visual quality of the images.

2.6.1 Pixel reconstruction loss function

The pixel-wise reconstruction loss function primarily measures the loss between the dehazed image obtained from the dehazing network and the true clear image. This loss function is used to constrain the dehazing network, ensuring that the dehazed image is closer to the true clear image [20], which is given by

$$L_s = \frac{1}{N} \sum_{i=1}^{N} \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \tag{7}$$

where N represents the total number of pixels, $x = J_i^g - J_i^d$ and J_i^g denote the values at the i pixel of the true image, and J_i^d denotes the value at the i pixel of the dehazed image.

2.6.2 Perceptual loss function

To ensure visual perceptual similarity between the dehazed image and the true image, a perceptual loss function [21] is introduced, as given in

$$L_p = \frac{1}{C_j H_j W_j} \left\| \varphi_j \left(J^d \right) - \varphi_j \left(J^g \right) \right\|_2^2 \tag{8}$$

where φ_j represents the feature maps extracted from the j layer of the VGG16 model.

2.6.3 Overall loss function

To better train the performance of the dehazing network model, a combined loss function is chosen. The pixel reconstruction loss function and the perceptual loss function are appropriately combined to constrain the dehazing network. The total loss function is defined as

$$L = L_s + 0.01 L_p \tag{9}$$

3. RESULTS AND DISCUSSION

3.1. Analysis of benchmark dataset

In this study, we primarily utilized the real-world non-homogeneous haze datasets I-HAZE [22], O-HAZE [23], and NH-HAZE [24], as well as the synthetically generated haze dataset RESIDE [25] for training the dehazing network model. During model training, due to the limited number of images in the real non-uniform hazy dataset, this study selects 30 pairs from the I-HAZE dataset and 45 images from the O-HAZE dataset. These images are uniformly divided into 64 parts, resulting in 1,920 pairs of images in the cut I-HAZE dataset and 2,880 pairs in the cut O-HAZE dataset. The NH-HAZE dataset consists of 55 pairs, which are uniformly divided into 16 parts, resulting in 880 pairs of images in the cut NH-HAZE dataset.

Among these, images from the I-HAZE and O-HAZE datasets are randomly selected in a proportion close to 15:1 to form the training and testing sets for the dehazing model. For the NH-HAZE dataset, the majority of the image pairs are randomly selected as the training set, with a smaller portion of images used as the test set for the dehazing model.
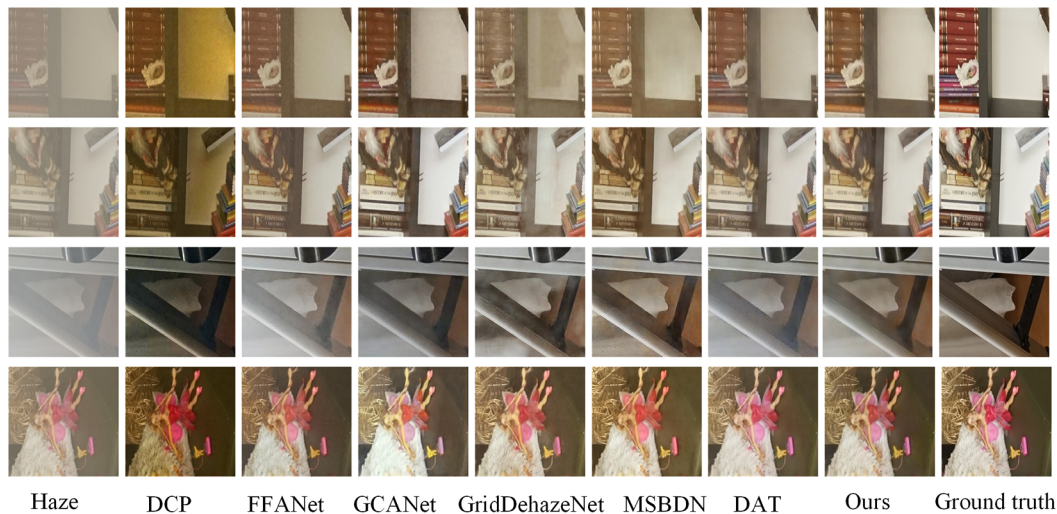
|Haze|DCP|FFANet|GCANet|GridDehazeNet|MSBDN|DAT|Ours|Ground truth|

Figure 5. I-HAZE dataset test results.

Finally, the experimental results of the dehazing model on these test sets were compared both subjectively and objectively with the results of existing dehazing algorithms, including DCP[3], FFANet[26], Gated Context Aggregation Network (GCANet)[19], GridDehazeNet[27], Multi-Scale Boosted Dehazing Network (MSBDN)[28], Ultrahighdefinition image Dehazing (UD)[29], and Dual Attention and Transformer (DAT)[12].

3.2. Subjective evaluation on real non-uniform haze datasets

To verify the effectiveness of the proposed dehazing model on real non-uniform haze images, this study conducted tests on the I-HAZE, O-HAZE, and NH-HAZE test datasets. The dehazing results on these different datasets were subjected to a subjective visual evaluation to assess the model's performance in enhancing the visual clarity of hazy images.

3.2.1 Analysis of I-HAZE dataset test results

The testing results of different algorithms on the I-HAZE dataset are shown in Figure 5.

From Figure 5, it can be observed that while the DCP dehazing algorithm achieves some dehazing effect, the overall distortion is relatively severe. For instance, the images in the second column have a generally dark color tone.

The FFANet and GridDehazeNet dehazing algorithms leave a small amount of haze residue in the test images. For example, the images in the third column show slight haze, and the first image in the fifth column exhibits more noticeable haze. The GCANet dehazing algorithm shows poor dehazing performance, with artifacts visible in the red regions of the fourth image in the fourth column.

The MSBDN and UD dehazing algorithms introduce localized color distortions in the test images, such as the bookshelf color in the second image of the sixth column, and the bookshelf and canvas colors in the second and third images of the seventh column. The DAT dehazing algorithm also leaves residual haze in some of the test images, as seen in the first image of the eighth column, where haze remains in the left half of the image. In contrast, the test images from the proposed algorithm are visually closer to the real clear images, demonstrating significantly better dehazing performance.
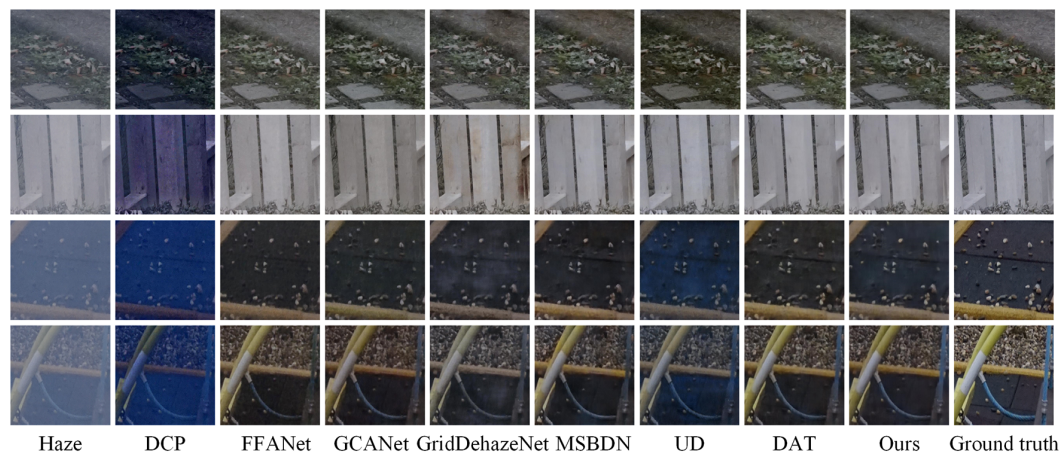
Figure 6. O-HAZE dataset test results.

3.2.2 Analysis of O-HAZE dataset test results

The test results of different dehazing algorithms on the O-HAZE dataset are shown in Figure 6.

From Figure 6, it can be observed that the DCP dehazing algorithm produces images with an overall bluish tint, leading to significant color distortion. The FFANet and GCANet algorithms also show some haze remnants, as seen in the lower right corner of the fourth image in the third and fourth columns. GridDehazeNet's dehazed images exhibit minor haze remnants in certain areas, such as the central region of the first, third, and fourth images in the fifth column.

While the MSBDN algorithm effectively removes haze, its performance in detail recovery is not as strong as the proposed algorithm. For instance, the central region of the first image in the sixth column and the first image in the ninth column demonstrate less effective detail recovery compared to the proposed method. The UD algorithm also shows dehazing effects, but some images suffer from color distortion, such as the presence of blue hues in the third and fourth images in the seventh column.

Both the DAT algorithm and the proposed algorithm achieve successful dehazing, but the proposed method excels in detail recovery compared to the DAT algorithm. For example, in the eighth and ninth columns, the recovery of stone details is more pronounced in the results of the proposed method when compared to the real label images in the tenth column.

3.2.3 Analysis of NH-HAZE dataset test results

To further validate the dehazing effect of the proposed algorithm on hazy images and enhance the generalization ability of the dehazing network model, testing was conducted on the NH-HAZE dataset. The testing results are shown in Figure 7.

From Figure 7, the DCP and UD dehazing algorithms exhibit color distortion in the test images. For example, the test images in the second and seventh columns show severe color distortion compared to the ground truth images. The GCANet, GridDehazeNet, and MSBDN dehazing algorithms result in haze residue in the test images, with poor dehazing performance. For instance, in the fourth, fifth, and sixth columns, the zoomed-in regions of the images exhibit residual haze. The FFANet dehazing algorithm shows haze remnants, as seen in the zoomed-in area of the fourth image in the third column. The DAT dehazing algorithm struggles with the recovery of texture details. For example, in the fourth image of the eighth column, the restoration of the chair's details is blurry. In contrast, the proposed algorithm demonstrates superior texture detail recovery compared
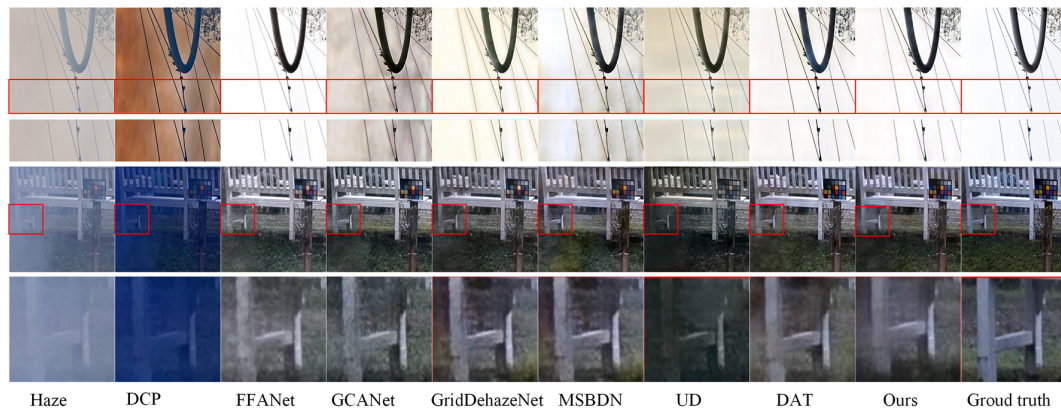
Figure 7. NH-HAZE dataset test results.

Table 1. Comparison of objective metrics for different algorithms on real non-uniform hazy datasets

Algorithm	I-HAZE		O-HAZE		NH-HAZE	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
DCP	10.06	0.574	12.54	0.6807	10.82	0.6556
FFANet	21.25	0.8428	24.57	0.758	19.97	0.7442
GCANet	21.4	0.8406	24.33	0.7599	19.94	0.7333
GridDehazeNet	20.45	0.8256	23.04	0.7448	20.74	0.7507
MSBDN	21.61	0.8576	24.27	0.7651	20.05	0.7521
UD	19.19	0.8181	22.46	0.7387	17.24	0.6764
DAT	**23.02**	<u>0.865</u>	<u>25.59</u>	<u>0.7771</u>	<u>21.22</u>	<u>0.7734</u>
Ours	<u>22.86</u>	**0.8731**	**25.86**	**0.7799**	**22.06**	**0.7796**

to the DAT algorithm, with the best dehazing effect overall.

3.3. Objective evaluation on real non-homogeneous haze image datasets

To verify the dehazing performance of various algorithms, an objective evaluation of the image restoration effects was conducted. The objective evaluation metrics used were Peak Signal to Noise Ratio (PSNR) and SSIM. The objective evaluation results of different dehazing algorithms on various datasets are shown in Table 1.

In Table 1, the optimal objective metric values are highlighted in bold, and the second-best values are underlined. As observed from this Table, the proposed algorithm achieves the best objective metrics on the I-HAZE, O-HAZE, and NH-HAZE real non-homogeneous haze datasets. Specifically, the PSNR of the proposed algorithm on the I-HAZE dataset is 22.86 dB, slightly lower than the optimal value, but its SSIM is 0.8731, which is 0.0081 higher than the second-best value. On the O-HAZE and NH-HAZE non-homogeneous haze datasets, the proposed algorithm demonstrates superior performance, with the best objective metric values compared to other benchmark algorithms.

3.4. Subjective visual evaluation on synthetic uniform haze dataset

To further validate the effectiveness of the dehazing algorithm, we conducted tests on both indoor and outdoor images from the SOTS dataset. The test results are shown in Figure 8.

From Figure 8 above, it can be observed that the DCP dehazing algorithm results in color distortion in the test images, such as the overall darkening of images in the second column. The FFANet and GridDehazeNet algorithms leave haze residue in the test images, with poor dehazing performance. For instance, the fourth image in the third column has haze remnants, and the fourth, fifth, and sixth regions in the fifth column still exhibit haze. The GCANet dehazing algorithm shows limited dehazing effectiveness in the sky areas of the test images, and indoor images suffer from some color distortion, such as the fourth image in the fourth column and the floor area in the second image of the fourth column. The MSBDN and UD dehazing algorithms result
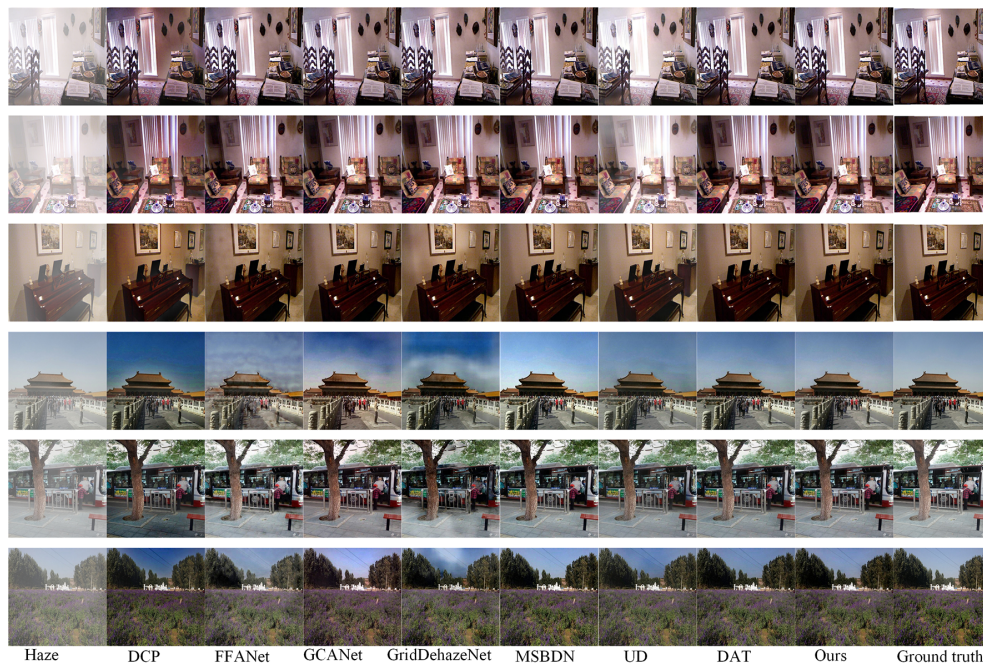
Figure 8. SOTS dataset test results.

Table 2. Comparison of image metrics for different algorithms on the SOTS dataset

Algorithm	SOTS-indoor		SOTS-outdoor	
	PSNR/dB	SSIM	PSNR/dB	SSIM
DCP	8.26	0.5579	9.75	0.6686
FFANet	30.12	0.9681	19.83	0.8225
GCANet	29.73	0.9626	21.77	0.8968
GridDehazeNet	29.07	0.9749	18.31	0.8450
MSBDN	28.31	0.9663	23.22	0.9125
UD	22.63	0.8946	22.00	0.9066
DAT	31.36	0.9767	22.68	0.9130
Ours	**32.13**	**0.9784**	**24.19**	**0.9189**

in overly bright images after dehazing, evident in the first and fourth images of the sixth column and the first and second images of the seventh column. The DAT dehazing algorithm exhibits slightly weaker detail recovery compared to the ground truth images, such as poor recovery in the ground area of the fifth image in the eighth column. In contrast, the proposed algorithm produces test images that are closer to the ground truth images, demonstrating better dehazing results and more noticeable detail recovery.

3.5. Objective evaluation of dehazing results on synthetic haze datasets

To further validate the effectiveness of the proposed algorithm for dehazing haze images, objective evaluation metrics PSNR and SSIM were used to analyze the test results on synthetic haze datasets, as shown in Table 2. The optimal objective metric values are highlighted in bold, and the second-best values are underlined.

As shown in Table 2, the proposed algorithm consistently achieves the best objective metrics. Specifically, in the SOTS-indoor dataset, the test images of the proposed algorithm have a PSNR of 32.13 dB, which is 0.77 dB higher than the second-best value, and an SSIM of 97.84, exceeding the second-best value by 0.0007. In the SOTS-outdoor dataset, the test images of the proposed algorithm achieve a PSNR of 24.19 dB, which is 0.97 dB higher than the second-best, and an SSIM of 0.9189, surpassing the second-best by 0.0069.
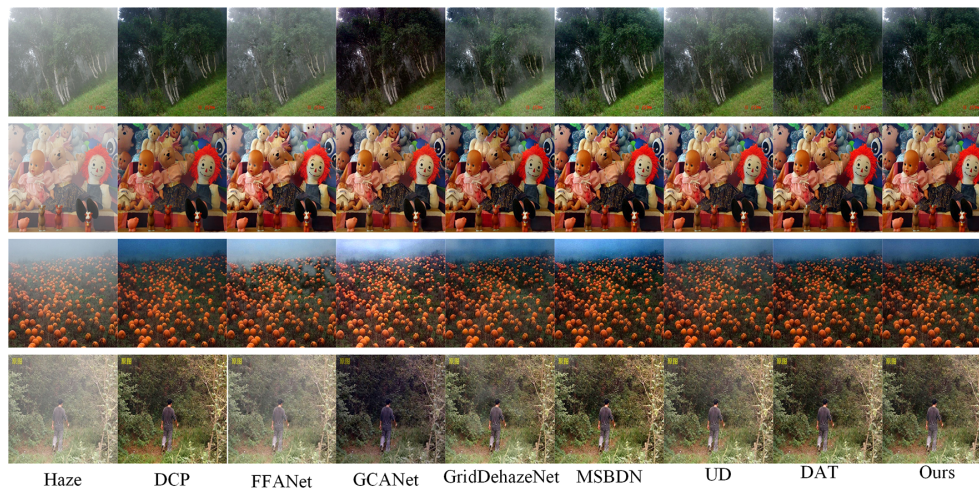
Haze   DCP   FFANet   GCANet   GridDehazeNet   MSBDN   UD   DAT   Ours

Figure 9. Test results on real-world hazy images.

3.6. Visual effects of real-world hazy images

Due to the random and uneven distribution of haze in real-world scenes, with varying thicknesses, we further validated the performance of the proposed algorithm on real non-uniform haze images. To evaluate this, we assessed four randomly selected dehazed images from real scenes without ground truth labels, as shown in Figure 9.

From Figure 9, it is evident that FFANet, GridDehazeNet, and UD dehazing algorithms leave haze residue in the test images, with poor dehazing performance in images such as the first images in columns 3, 5, and 7. The GCANet dehazing algorithm results in color distortion, such as excessive whiteness in the sky area of the third image in column 4. The DCP, MSBDN, and DAT dehazing algorithms do not perform as well in detail recovery compared to the proposed algorithm, with examples such as the face of the doll in the second image of columns 2, 6, and 8 showing less effective recovery. In summary, the subjective visual evaluation of the dehazed results on real non-uniform haze images demonstrates that the proposed algorithm outperforms other comparative algorithms in both haze removal and texture detail recovery.

3.7. Ablation study

To demonstrate the effectiveness of each module in the proposed algorithm, ablation experiments were conducted using different configurations trained and tested on the real non-uniform haze dataset I-HAZE. The differences between the experimental results and the ground truth images were computed, and both subjective visual assessments and objective metrics (PSNR, SSIM) were used for comparison. The specific ablation experiments are as follows:

- Using a multi-dimensional Transformer network for deep feature extraction, referred to as "MT";
- Using a channel attention mechanism for shallow feature extraction to focus on haze features in different channels, with a multi-dimensional Transformer for deep feature extraction, referred to as "MT+CA";
- Using a spatial attention mechanism for shallow feature extraction to focus on haze concentration in different regions, with a multi-dimensional Transformer for deep feature extraction, referred to as "MT+SA";
- Using a serial comprehensive attention mechanism for shallow feature extraction, focusing on different haze regions, combined with a multi-dimensional Transformer for deep feature extraction, which is the proposed non-uniform image dehazing algorithm, referred to as "Ours."

The ablation results are shown in Figure 10, with the red box highlighting the enlarged regions.
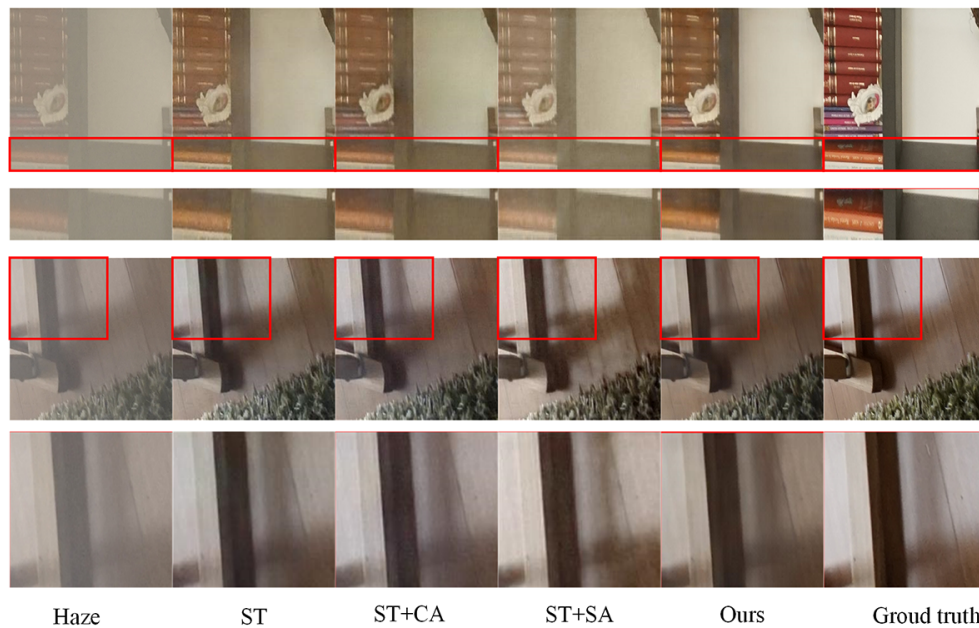
Haze	ST	ST+CA	ST+SA	Ours	Groud truth

Figure 10. Ablation study test results.

Table 3. Test results of different modules on the I-HAZE dataset

Module	I-HAZE	
	PSNR/dB	**SSIM**
MT	22.64	0.8561
MT+CA	22.66	0.8635
MT+SA	22.08	0.866
Ours	**22.86**	**0.8731**

The first column shows hazy images; the second to fifth columns display the dehazing results, and the sixth column shows the ground truth images. From the second row, it is evident that all methods achieve some dehazing effect. However, the combination of serial comprehensive attention and multi-dimensional Transformer provides better color restoration, with the proposed algorithm's results being closer to the ground truth images, while other methods show color distortion. The fourth row demonstrates that the proposed algorithm also effectively recovers details, such as those on wooden boards in non-uniform haze images.

To further validate the effectiveness of the proposed algorithm for dehazing non-uniform haze images, objective evaluation metrics were used to assess the results of the ablation experiments. Table 3 presents a comparison of the objective evaluation metrics for the four ablation experiments. The optimal objective metric values are highlighted in bold.

From Table 3, the results on the I-HAZE dataset indicate that the proposed algorithm achieves the best PSNR and SSIM scores. This demonstrates that the combination of the serial comprehensive attention mechanism and multi-dimensional Transformer in the proposed non-uniform dehazing algorithm not only effectively removes haze from non-uniform images but also achieves the best performance in these objective metrics.

4. CONCLUSIONS AND FUTURE WORK

To tackle the problems of detail loss and blurry restoration results in existing dehazing algorithms, this paper incorporates the non-uniform image dehazing algorithm based on serialized integrated attention and multi-dimensional transformer. By merging spatial and channel attention mechanisms seamlessly, the algorithm

accurately targets crucial local image features in both the spatial and channel dimensions, providing a robust basis for subsequent feature processing and image restoration. Building upon this foundation, a multi-dimensional Transformer module is introduced into the deep feature extraction network. Leveraging its strong global modeling capability, this module effectively captures global image information, notably minimizing information loss and overcoming the limitations of traditional algorithms in dealing with global features. By deploying these two modules with unique advantages into the shallow and deep networks respectively and enabling them to cooperate, the proposed algorithm is able to extract comprehensively and profoundly the image's feature information, achieving more efficient dehazing of non-uniform images. Extensive experimental results confirm the superior performance of the proposed algorithm. Compared to other algorithms, the dehazed images achieve optimal SSIM and PSNR values, and exhibit enhanced dehazing efficiency. Notwithstanding these advancements, the time-consuming nature of the algorithm restricts its application in real-time scenarios, leaving considerable room for improvement. In the future, we will continue to conduct in-depth research, focusing on optimizing the algorithm structure and reducing computational complexity, aiming to enhance the algorithm's efficiency and better meet the demands of various practical application scenarios, thereby further unlocking its potential in the field of image processing.

DECLARATIONS

Authors' contributions
Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Bai, T.
Contributed to approach validation and writing-original draft preparation: Bai, T.; Qiu, J.
Contributed to the investigation, supervision, and writing-review and preparation: Qiu, J.
Both authors read and approved the final manuscript.

Availability of data and materials
The code and data used to support the findings of this study are available from the corresponding author upon request.

Financial support and sponsorship

Conflicts of interest
Both authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright

REFERENCES

1. Chiu, C. C.; Ting, C. C. Contrast enhancement algorithm based on gap adjustment for histogram equalization. *Sensors.* **2016**, *16*, 936. DOI

2. Qiao, S.; Li, Q.; Wang, Y. An improved color attenuation priori dehazing algorithm and its hardware implementation. In 2019 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC); 12-14 June 2019; Xi'an, China. pp.1-3. DOI

3. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **2011**, *33*, 2341-53. DOI

4. Liu, Y.; Yan, Z.; Tan, J.; Li, Y. Multi-purpose oriented single nighttime image haze removal based on unified variational Retinex model. *IEEE. Trans. Circuits. Syst. Video. Technol.* **2023**, *33*, 1643-57. DOI

5. Berman, D.; Treibitz, T.; Avidan, S. Single image dehazing using haze-lines. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **2020**, *42*, 720-34. DOI

6. Li, S.; Yuan, Q.; Zhang, Y.; Lv, B.; Wei, F. Image dehazing algorithm based on deep learning coupled local and global features. *Appl. Sci.* **2022**, *12*, 8552. DOI

7. Ilesanmi, A. E.; Ilesanmi, T. O. Methods for image denoising using convolutional neural network: a review. *Complex. Intell. Syst.* **2021**, *7*, 2179-98. DOI

8. He, S.; Chen, Z.; Wang, F.; Wang, M. Integrated image defogging network based on improved atmospheric scattering model and attention feature fusion. *Earth. Sci. Inform.* **2021**, *14*, 2037-48. DOI

9. Li, Z.; Gao, T.; Chen, C.; Wen, Y. Non-uniform dehazing algorithm based on improved ConvNeXt. In Proceedings of the 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT); 29-31 March 2024; Nanjing, China, pp.613-9. DOI

10. Lin, C.; Rong, X.; Yu, X. MSAFF-net: multiscale attention feature fusion networks for single image dehazing and beyond. *IEEE. Trans. Multimedia.* **2023**, *25*, 3089-100. DOI

11. Song, Y.; He, Z.; Qian, H.; Du, X. Vision transformers for single image dehazing. *IEEE. Trans. Image. Process.* **2023**, *32*, 1927-41. DOI

12. Wang, K. P.; Zang, Z. J.; Yang, Y, Fei, S. M.; Wei, J. Y. Non-homogeneous dehazing algorithm based on fusion of dual attention and transformer. *J. Beijing. Univ. Posts. Telecommun.* **2024**, *47*, 30-7. DOI

13. Wang, T.; Zhang, K.; Shao, Z.; et al. GridFormer: residual dense transformer with grid structure for image restoration in adverse weather conditions. *Int. J. Comput. Vision.* **2024**, *23*, 20. DOI

14. Jiang, K.; Wang, Q.; An, Z.; Wang, Z.; Zhang, C.; Lin, C. Mutual Retinex: combining transformer and CNN for image enhancement. *IEEE. Trans. Emergy. Top. Comput. Intell.* **2024**, *8*, 2240-52. DOI

15. Zheng, L.; Li, Y.; Zhang, K.; Luo, W. T-net: deep stacked scale-iteration network for image dehazing. *IEEE. Trans. Multimedia.* **2023**, *25*, 6794-807. DOI

16. Qiu, Y.; Zhang, K.; Wang, C.; Luo, W.; Li, H.; Jin, Z. MB-TaylorFormer: multi-branch efficient transformer expanded by Taylor formula for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2023, pp. 12802-13. Available from: https://openaccess.thecvf.com/content/ICCV2023/html/Qiu_MB-TaylorFormer_Multi-Branch_Efficient_Transformer_Expanded_by_Taylor_Formula_for_Image_ICCV_2023_paper.html [Last accessed on 18 Mar 2025]

17. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention is all you need. *arXiv.* **2017**, *30*, 6000-10. DOI

18. Liu, Z.; Lin, Y.; Cao, Y.; et al. Swin transformer: hierarchical vision transformer using shifted windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10-17 October 2021; Montreal, QC, Canada, pp.10012-22. DOI

19. Chen, D.; He, M.; Fan, Q.; et al. Gated context aggregation network for image dehazing and deraining. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV); 7-11 January 2019; Waikoloa, HI, USA, pp.1375-83. DOI

20. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV); 7-13 December 2015; Santiago, Chile. DOI

21. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer vision - ECCV 2016. Cham: Springer International Publishing; 2016. pp. 694-711. DOI

22. Ancuti, C.; Ancuti, C. O.; Timofte, R.; De Vleeschouwer, C. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. In: Blanc-talon J, Helbert D, Philips W, Popescu D, Scheunders P, editors. Advanced Concepts for Intelligent Vision Systems. Cham: Springer International Publishing; 2018. pp. 620-31. DOI

23. Ancuti, C. O.; Ancuti, C.; Timofte, R.; De Vleeschouwer, C. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. *arXiv.* **2018**. DOI

24. Ancuti, C. O.; Ancuti, C.; Timofte, R. NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 14-19 June 2020; Seattle, WA, USA, pp.444-5. DOI

25. Li, B.; Ren, W.; Fu, D.; et al. Benchmarking single image dehazing and beyond. *IEEE. Trans. Image. Process.* **2018**, *28*, 492-505. DOI

26. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-net: feature fusion attention network for single image dehazing. *AAAI.* **2020**, *34*, 11908-15. DOI

27. Liu, X.; Ma Y.; Shi, Z.; Chen, J. GridDehazeNet: attention-based multi-scale network for image dehazing. In Proceedings of the I 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 27 October 2019-2 November 2019; Seoul, Korea (South), pp.7314-23. DOI

28. Dong, H.; Pan, J.; Xiang, L.; et al. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 13-19 June 2020; Seattle, WA, USA. pp. 2157-67. DOI

29. Zheng, Z.; Ren, W.; Cao, X.; et al. Ultra-high-definition image dehazing via multi-guided bilateral learning. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 20-25 June 2021; Nashville, TN, USA, pp.16180-9. DOI