electronics

MDPI

Article

Adaptive Multi-Feature Attention Network for Image Dehazing

Hongyuan Jing [1,2,*], Jiaxing Chen [1,2], Chenyang Zhang [2], Shuang Wei [2], Aidong Chen [1,2,3] and Mengmeng Zhang [1,2,3,*]

[1] Beijing Key Laboratory of Information Service Engineering, College of Robotics, Beijing 100101, China; 2020220360008@buu.cn (J.C.); aidong@buu.edu.cn (A.C.)
[2] College of Robotics, Beijing Union University, No.4 Gongti North Road, Beijing 100027, China; 2020220360005@buu.edu.cn (C.Z.); 2020240388011@buu.edu.cn (S.W.)
[3] Multi-Agent System Research Centre, Beijing Union University, No. 97 Beisihuan East Road, Beijing 100101, China
* Correspondence: jqrhongyuan@buu.edu.cn (H.J.); zhangmengmeng@buu.edu.cn (M.Z.)

Abstract: Currently, deep-learning-based image dehazing methods occupy a dominant position in image dehazing applications. Although many complicated dehazing models have achieved competitive dehazing performance, effective methods for extracting useful features are still under-researched. Thus, an adaptive multi-feature attention network (AMFAN) consisting of the point-weighted attention (PWA) mechanism and the multi-layer feature fusion (AMLFF) is presented in this paper. We start by enhancing pixel-level attention for each feature map. Specifically, we design a PWA block, which aggregates global and local information of the feature map. We also employ PWA to make the model adaptively focus on significant channels/regions. Then, we design a feature fusion block (FFB), which can accomplish feature-level fusion by exploiting a PWA block. The FFB and PWA constitute our AMLFF. We design an AMLFF, which can integrate three different levels of feature maps to effectively balance the weights of the inputs to the encoder and decoder. We also utilize the contrastive loss function to train the dehazing network so that the recovered image is far from the negative sample and close to the positive sample. Experimental results on both synthetic and real-world images demonstrate that this dehazing approach surpasses numerous other advanced techniques, both visually and quantitatively, showcasing its superiority in image dehazing.

Keywords: dehazing; deep learning; attention mechanism; adaptive feature fusion

check for updates

Citation: Jing, H.; Chen, J.; Zhang, C.; Wei, S.; Chen, A.; Zhang, M. Adaptive Multi-Feature Attention Network for Image Dehazing. *Electronics* **2024**, *13*, 3706. https://doi.org/10.3390/electronics13183706

Academic Editor: Silvia Liberata Ullo

Received: 27 July 2024
Revised: 31 August 2024
Accepted: 6 September 2024
Published: 18 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Extreme weather phenomena, encompassing haze, fog, and dust storms, significantly diminish the visibility of the surroundings, leading to images captured under such conditions experiencing substantial a loss of color fidelity and content details. This image degradation poses a formidable challenge to tasks such as object recognition, image partitioning, movement tracking, and other advanced visual processing endeavors, ultimately resulting in a marked decline in the accuracy of predictions and outcomes derived from these high-level visual tasks. Therefore, as a fundamental vision-processing task, single-image dehazing has attracted significant attention in recent years. The purpose of single-image dehazing is to restore a clean scene from a hazy image. Acknowledging the established understanding, the haze formation process is often mathematically modeled using the Atmospheric Scattering Model (ASM) [1]. This representation is formulated as

$$I(x) = J(x) \cdot t(x) + A \cdot (1 - t(x)), \tag{1}$$

here, $I(x)$ represents the hazy image that is observed, whereas $J(x)$ signifies the corresponding haze-free image. A denotes the global atmospheric light, capturing the environment's overall illumination characteristics, and $t(x)$ denotes the transmission map, a function that quantifies the light's portion reaching the observer without being scattered by particles in

the atmosphere. This model serves as a fundamental basis for analyzing and mitigating the effects of haze on visual data. It is determined by the distance $d(x)$ from the scene to the camera and the atmospheric scattering coefficient β. The transmission map can be formulated as an exponential decay function with respect to the scene-to-camera distance, as given by:

$$t(x) = e^{-\beta \cdot d(x)}, \tag{2}$$

Given a hazy image, the problem of recovering its clean version is highly ill-posed. Conventional approaches rely on ASM [1] and exploit hand-crafted prior assumptions to solve this challenge. For example, in order to estimate the transmission map, He et al. [2] proposed a Dark Channel Prior that can eliminate the halo phenomenon and blocky situation well. However, when the image tends towards whiteness (especially the sky area), it will cause color distortion and reduce haze removal. Berman et al. [3] proposed a non-local prior in order to represent the recuperative images. Nevertheless, due to its excessive dependence on color classification accuracy, it decreases seriously with the increase of haze concentration. Compared with previous prior-based methods, the IDE [4] method proposed by Ju et al. introduces the light absorption coefficient into the atmospheric scattering model [1], which can boost the visibility of hazy images. However, in distant regions, the preprocessed outcomes often exhibit darker tendencies. Ju et al. proposed an extremely efficient single-image dehazing method named IDGCP [5] based on a gamma correction prior (GCP) to restore high-quality images with only an unknown constant. Huang et al. [6] introduced an innovative image dehazing technique that ingeniously integrates the strengths of various dehazing strategies, including pixel-level, local, non-local, and scene-aware approaches. This comprehensive approach addresses the limitations of relying exclusively on a single prerequisite, thereby enhancing the effectiveness and versatility of image dehazing. These methods try to restrict space to some extent, enhancing image visibility. However, presumptions and prerequisites tailored to particular scenes or atmospheric circumstances may lead to distorted recovered images.

In the past decade, convolutional neural networks (CNNs) [7] have achieved significant advancements in the field of image dehazing. Many researchers have posed a multitude of data-centric approaches. The original method leveraging deep learning techniques, such as All-in-One Dehazing (AOD) [8] and Multi-Scale CNN (MSCNN) [9], has proven CNN's great success. Unlike most previous models, which obtain the transmission rate and the atmospheric light value, Li et al. [8] unified these two parameters into one variable by designing an all-in-one dehazing network. The multi-scale convolutional neural network is proposed by Ren et al. [9] for extracting useful features from hazy images to facilitate scene transmission map estimation. Opposite to minimizing the loss between the dehazed image and the real image, the network is trained to optimize the accuracy of the reconstructed transmission map against its ground truth, enhancing dehazing effectiveness.

However, these methods still follow the traditional dehazing model in Equation (1) and mostly use low-level image features. It has been proven that by stacking convolutional layers to obtain deeper features, the image quality can be improved. To directly obtain restored images, Qin et al. [10] proposed an end-to-end feature fusion attention network (FFA). This method integrates channel-wise and pixel-wise attention mechanisms (CA and PA) to handle different features and uneven pixels, further enhancing the clarity and fidelity of the reconstructed image. However, if different levels of haze are weighted in the same way, the dehazing effect will be reduced. The above methods are easy to over-fit and time-consuming and have poor real-time performance. Some state-of the-art (SOTA) models and our proposed method about their PSNR [11] value and SSIM [12] value are presented in Figure 1.

Although current CNN-based methods have remarkable performance, their model capacity is limited, which depends widely on feature extraction. Indeed, by leveraging an encoder–decoder CNN architecture [7], the model effectively captures the nonlinear relationship between blurred and sharp image pairs, enabling superior feature extraction and overall performance. For instance, Jing et al. [13] proposed a U-Net-based feature at-

tention dehazing network, which adopted feature attention, dense connection, and residual dense block to solve heterogeneity dehazing task. Bianco et al. [14] proposed HR-Dehazer, which utilizes encoder–decoder architecture to train the mapping between hazy images and dehazed images. Feng et al. [15] proposed the URNet dehazing model. They use a hybrid convolution that combines standard convolution and extended convolution in the network's encoder to enlarge the acceptance domain and thus extract image features better. Lu et al. [16] propose a novel framework, combining multi-scale processing, large convolutional kernels, and attention mechanisms. These designs can enhance the feature extraction and learning capabilities of networks. Shen et al. [17] design a two-stage image dehazing network. The first stage is amplitude-guided dehazing, and the second stage is phase-to-structure refinement. Although this method can reduce the probability of information redundancy and enhance the complementarity of features between different layers, it can only capture feature information within a limited domain. Due to the superior performance of self-attention [18], Shi et al. [19] built a Transformer branch combined with Transformer [18] self-attention and built a convolutional neural network branch based on the locally varying attention module. When the two branches are combined, non-local features could be recovered through the Transformer [18] branch, and local features could be obtained through the CNN branch to complete the recovery of the whole image. However, at present, the method has a better effect on facial overdivision, and its role in the field of image defogging remains to be further explored. The above methods can achieve excellent performance but lack an integrating multi-layer feature extraction, fusion, and enhanced focus on intricate details. Therefore, they still have great room for improvement in extracting image features. To optimize feature extraction capabilities and refine the allocation of feature weights, we propose a multi-level feature fusion method and point-wise weighted attention.

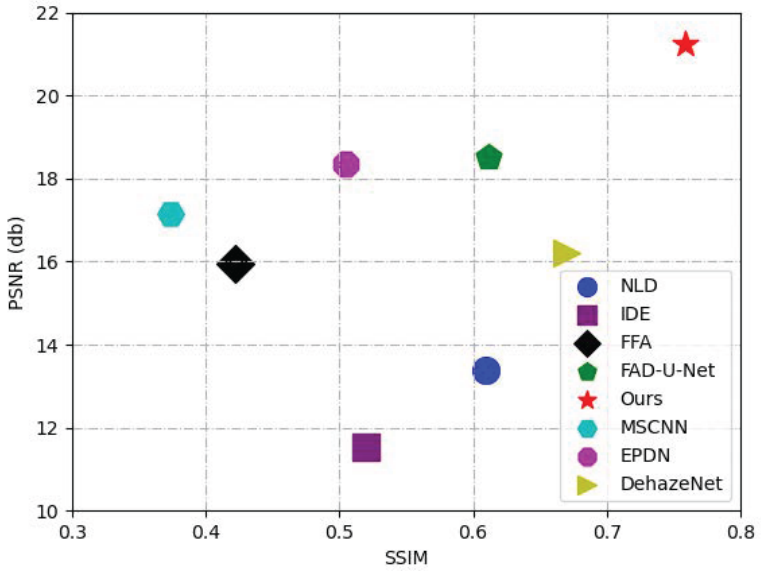

Figure 1. The best PSNR and SSIM of compared methods and our method.

In this work, we propose an adaptive multi-feature attention network (AMFAN) that simultaneously applies to the removal of haze across both consistent and varying degrees of haze in images. A sketch of the main ideas is shown in Figure 2. Inspired by the method in which Jing et al. [13] add dense connections in both the encoder and decoder parts to improve the information utilization between each non-adjacent layer, we utilize the basic framework of presented haze reduction approach by Jing et al. [13] and simultaneously design an attention multi-layer feature fusion module (AMLFF) and a point-wise weighted attention module (PWA) in the decoder part, which can obtain balanced weight distribution of detailed information and frame information and the point-to-point weights map of attention strategy. Therefore, in the challenging non-homogeneous

haze case, our method largely surpasses some previous methods. To ensure a balanced evaluation of both objective and subjective image quality, the training process incorporates not only image quality losses like PSNR [11] and SSIM [12] but also a loss function that captures visual perception. This approach comprehensively assesses the restored images' fidelity and perceptual appeal.

To summarize, the key contributions outlined in this paper encompass the following aspects.

- We propose an adaptive multi-feature attention network. The point-wise attention module (PWA) for enhancing the fusion strategy and the attention multi-layer feature fusion (AMLFF) module for balancing features across different layers are two key components of this network.
- To enhance multi-scale feature fusion, we propose the point-wise attention (PWA) module, which integrates both global and localized features from the feature maps by focusing on important regions.
- We introduce a feature fusion block (FFB) based on the idea of AFF by using the point-wise weighted attention module (PWA) to better balance the multi-scale features of the inputs. In addition, we use the proposed feature fusion block (FFB) to design an attention multi-layer feature fusion module (AMLFF) to fuse features from different layers adaptively and balance the weights between feature maps well.

In this article, the first part introduces the research methodology employed in this paper, the second section discusses the development of image dehazing methods, the third section elaborates on the architecture of the introduced network framework and the loss functions used, and the fourth part proves the effectiveness and advancement of the proposed method through extensive experiments.

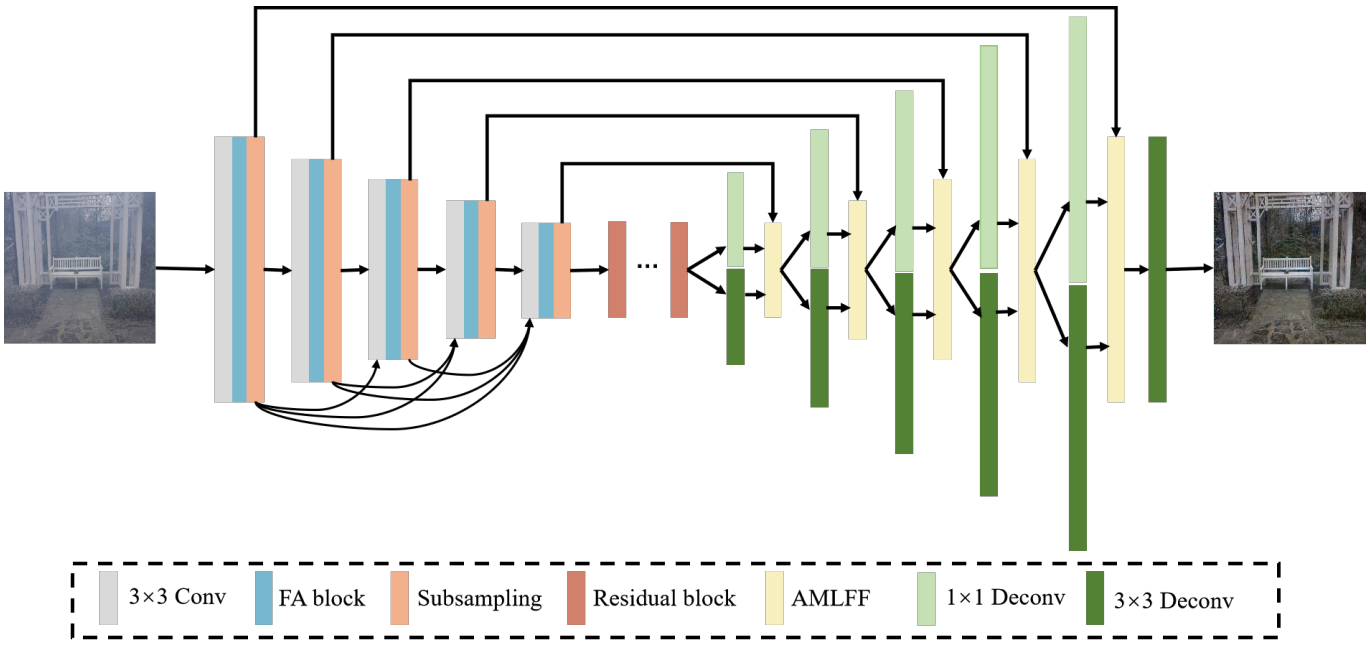

Figure 2. Architecture of the proposed adaptive multi-feature attention network. Given a hazy image, we utilize a 3 × 3 convolution layer to construct the features. The encoder component projects the feature map into a multidimensional space. In the decoder module, the AMLFF module uses the PWA module to balance the weights of different features. At the same time, the AMLFF module also fuses multiple features to obtain the fused feature map.

2. Related Work

Recently, two primary networks have emerged for image dehazing: the traditional prior-based dehazing network and the deep-learning-based dehazing network. The former

approach employs constraints and atmospheric scattering models, whereas the latter relies solely on convolutional neural networks and latter encompasses direct, indirect, and hybrid end-to-end dehazing networks.

2.1. Traditional Dehazing Methods

The prior-based dehazing method involves manually distilling the statistical disparity between hazy and clear images as a priori knowledge. Subsequently, this prior knowledge is used to inversely derive the final clear image from the parameters A and t in the given Equation (1). The prior-based method is the pioneer of image dehazing. He et al. [2] argue that in non-sky regions, certain pixels exhibit at least one color channel with significantly diminished intensity. They refer to this observation as the Dark Channel Prior [2] and propose a dark-channel-based dehazing algorithm based on this prior. However, this algorithm neglected the disparity in image features between the sky and non-sky regions, thereby lacking segmentation of the sky region in hazy images. Meng et al. [20] employed regularization algorithms to constrain the boundaries of the sky region. However, this approach can lead to a loss of image detail and create artifacts in the boundaries. Relying solely on low-level features often results in the restored image exhibiting excessive color saturation or inaccuracies. Fattal et al. [21] assumed that the scene's chromaticity was independent of the transmittance and consequently employed statistical methods to estimate the transmittance of the scene. But this approach does not achieve satisfactory results for images that possess limited color characteristics. Berman et al. [3] introduced a non-local prior algorithm that rearranges image pixel clusters according to haze density, yet it is prone to leaving residual haze artifacts. Numerous experiments have shown that variations in environmental conditions can significantly impact prior conditions, which restricts the effectiveness of the aforementioned methods to only a very limited set of scenarios.

2.2. Deep-Learning-Nased Methods

Undeniably, deep learning techniques have attained noteworthy accomplishments, spanning from natural language processing and computer vision to the realm of image dehazing. Broadly speaking, these techniques can be classified into three distinct categories: indirect end-to-end dehazing networks, direct end-to-end dehazing networks, and hybrid end-to-end dehazing networks.

Indirect end-to-end dehazing networks: The indirect end-to-end methods often combine convolutional neural networks with the atmospheric scattering model [1] to estimate the transmission map and atmospheric light value and then inversely derive the dehazed image to complete the image dehazing task. Cai et al. [22] proposed a trainable end-to-end network called DehazeNet for estimating the transmission rate, which utilizes existing assumptions in image dehazing and convolutional feature extraction and restores clear images through pixel-by-pixel operations. However, this method has the issue of incomplete feature extraction. Since most datasets for training dehazing networks are synthetic hazy images, Chen et al. [23] proposed a Principled Synthetic-to-Real Dehazing (PSD) method to enhance the generalization and robustness of the network. Due to the fact that indirect, end-to-end dehazing methodologies rooted in the atmospheric scattering model [1] primarily concentrate on estimating the transmission map, Liang et al. [24] proposed a dehazing algorithm that utilizes the well-known Dark Channel Prior and reduces the problem of selecting region sizes through a pyramid-based multi-scale image. This approach adopts a pyramid structure to obtain image features and derives the final clear image based on the atmospheric scattering model [1]. However, indirect end-to-end methods often rely on the atmospheric scattering model [1], which has significant limitations.

Direct end-to-end dehazing networks: Direct end-to-end methods do not calculate atmospheric light values and transmission rates. Instead, they directly extract image features to remove haze and obtain clean images, which provides better universality and robustness. The entire network framework proposed by Zhao et al. [25] consists of a feature extraction module, a Pyramid Global Context-based U-Net module, a deconvolution layer, and a

pyramid pooling module. It utilizes pyramid pooling operations to learn point-to-point attention related to remote haze. However, due to insufficient image feature extraction and a loss of spatial information, direct end-to-end methods often have issues with the recovery of unclear image details. To address this problem, Zhang et al. [26] proposed a dense residual dilation dehazing network that utilizes a dense connection block to integrate both low-level and high-level feature representations, thus fully utilizing multi-scale features for better dehazing. Jing et al. [13] introduced a feature attention dehazing network based on U-Net and a dense connection model, which uses a feature attention module to replace one convolution operation in each encoder layer to better preserve details. To mitigate information loss, dense connections are incorporated between the encoder and decoder components while enhancing the utilization of features across different layers. For performance enhancement and to augment the network's depth and width, Luo et al. [27] designed a novel large kernel convolution dehazing block composed of decomposed depthwise large kernel convolution blocks and channel-enhanced feed-forward networks, rather than directly increasing the size of convolution kernels, to reduce the parameters and computational overhead. Fu et al. [28] proposed a new two-dimensional discrete wavelet transform dehazing network called DW-GAN. This method first applies wavelet transforms in the DWT branch [28], capable of retaining a richer assortment of high-frequency image details.

Hybrid end-to-end dehazing networks: Hybrid end-to-end networks often have multiple branches that utilize different sub-networks to interact and achieve the clarification of blurry images. To alleviate the distortion phenomenon in sky regions caused by the Dark Channel Prior (DCP) method [2], Zhao et al. [29] introduced a refinement feature branch based on the DCP [2] method. However, this method still relies on priors and has poor generalization. In their work, Yang et al. [30] fused haze imaging constraints and image prior learning within a multi-branch network architecture, grounded in the DCP [2] approach, utilizing CNN to further ascertain the transmission map and Dark Channel Prior [2] parameters, thus combining deep learning methods with traditional image restoration methods. Li et al. [31] introduced the Neural Radiance Field (NeRF) image restoration method into image restoration algorithms, establishing a simple and effective image restoration NeRF model called Dehazing-NeRF. One subnetwork in this method utilizes NeRF to synthesize colors and light rays at any position and direction in the image, while another subnetwork estimates parameters in the atmospheric scattering model [2] in an unsupervised manner. The final network completes the task of image restoration by combining the atmospheric scattering model [2] without relying on manual design. To better restore the image structure and details in dense fog regions, Guo et al. [32] established a dual-branch semi-supervised attention network (SCANet), fusing two different network structures, enabling the network to better restore fog-covered regions while improving the convergence speed.

3. Proposed Method

The subsequent section delves into the intricacies of our adaptive multi-feature attention network tailored for dehazing purposes, illustrated in Figure 2.

As shown in Figure 2, our adaptive multi-feature attention network is established based on the encoder–decoder structure. Specifically, we stack a layer of 3×3 convolution and a feature attention (FA) [10] module at each encoder layer. The 3×3 convolution is used to refine the extracted features, and the FA [10] module allows the network to prioritize salient and informative features while diminishing the influence of irrelevant or disruptive data. Finally, after a downsampling layer, it is passed to the next layer. To reduce feature loss during image compression, we introduce dense connections [33] between the different layers of the encoder. In the high-dimensional part of the network, we choose to keep the residual module of FAD-U-Net [13] to realize the removal of image haze. In the decoder part, we have made a lot of innovations on the basis of FAD-U-Net [13] and established a new attention module to accomplish adaptive integration of image features across varying

levels and further restore clear images. At the end, a layer of 3×3 convolution is used to obtain the final clear image.

The implementation details of the decoder are described in detail in this section. Firstly, we specify the detail of point-wise weighted attention module that can aggregate the local and global information of high-level feature maps. This configuration enhances the network's capability to prioritize haze-impacted areas but also effectively enhances feature fusion, improving the network's dehazing performance. Then, we elaborate on the attention multi-layer feature fusion module, which is beneficial to effectively balance features from different levels. Finally, we detail the loss functions employed throughout the training process.

3.1. Point-Wise Weighted Attention

Considering that different scale features are capable of distinguishing weight information and the non-uniform dispersal of haze patches, we design a point-wise weighted attention module (PWA) at pixel level to improve the weight of the feature fusion strategy, as shown in the Figure 3b. Since spatial feature information is crucial to the integrity of the image, we use adaptive scaling features by exploiting channel attention to obtain spatial inter-dependencies between channels. In this module, Initially, we integrate the global spatial information across channels into the channel descriptors, utilizing global average pooling [34]; the operation formula is as follows:

$$g_c = H_p(F_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \tag{3}$$

$$g_w = H_p(F_w) = \frac{1}{H \times C} \sum_{i=1}^{H} \sum_{j=1}^{C} X_w(i,j) \tag{4}$$

$$g_h = H_p(F_h) = \frac{1}{C \times W} \sum_{i=1}^{C} \sum_{j=1}^{W} X_h(i,j) \tag{5}$$

where $X_c(i,j)$ represents the value at the position (i,j) of theC-th channel; $X_w(i,j)$ represents the value at (i,j), the position of the W-th channel; $X_h(i,j)$ represents the value at the position (i,j) of the H-th channel; and H_p is the global average pooling function. After the average pooling on the channels, the shape of the feature map changes from $C \times H \times W$ to $C \times 1 \times 1$. Meanwhile, the input feature map is transposed to obtain $H \times C \times W$ and $W \times H \times C$ feature maps, average pooling is performed on the transposed images, and the shape of the feature maps is changed to $W \times 1 \times 1$ and $H \times 1 \times 1$, respectively.

Subsequently, the three obtained feature maps are sequentially passed through two layers of convolution and a ReLU activation function [9], followed by a Sigmoid function, with the aim of deriving weights for different channels.

$$F_c = Conv(\delta(Conv(g_c(X_c)))) \tag{6}$$

$$F_w = Conv(\delta(Conv(g_w(X_w)))) \tag{7}$$

$$F_h = Conv(\delta(Conv(g_h(X_h)))) \tag{8}$$

where g_c, g_w, g_h denote the average pooling operation on X_c, X_w, X_h, respectively; $Conv$ represents a kernel size of 1×1 convolution; and δ denotes the Rectified Linear Unit. Then, the three obtained attention maps with weights of different channels result in a point-wise weighted attention feature map enriched with global contextual data through multiplication.

$$F_{globalatt} = F_c \otimes F_w \otimes F_h \tag{9}$$

where F_c represents an attention feature map of size $C \times 1 \times 1$, F_w represents an attention feature map of size $1 \times W \times 1$, and F_h represents an attention feature map of size $1 \times 1 \times H$. To emphasize the importance of hazy regions and high-frequency feature details, we design an attention module capable of obtaining local contextual content.

The attention module has two convolutional layers a ReLU activation layer, and a BN layer [35] behind each convolutional layer. In the BN layer between each convolution layer, the point-by-point weighted attention map with local context information can be obtained.

$$F_{localatt} = \beta(Conv(\delta(\beta(Conv(X_c))))) \tag{10}$$

where X_c represents the feature map of the input, $Conv$ denotes a kernel size of a 1×1 convolution, β is the Batch Normalization (BN) [35], and δ denotes the Rectified Linear Unit. The size of the feature map is not changed from $C \times H \times W$ to $1 \times H \times W$ but keeps the original size of the input feature map $C \times H \times W$.

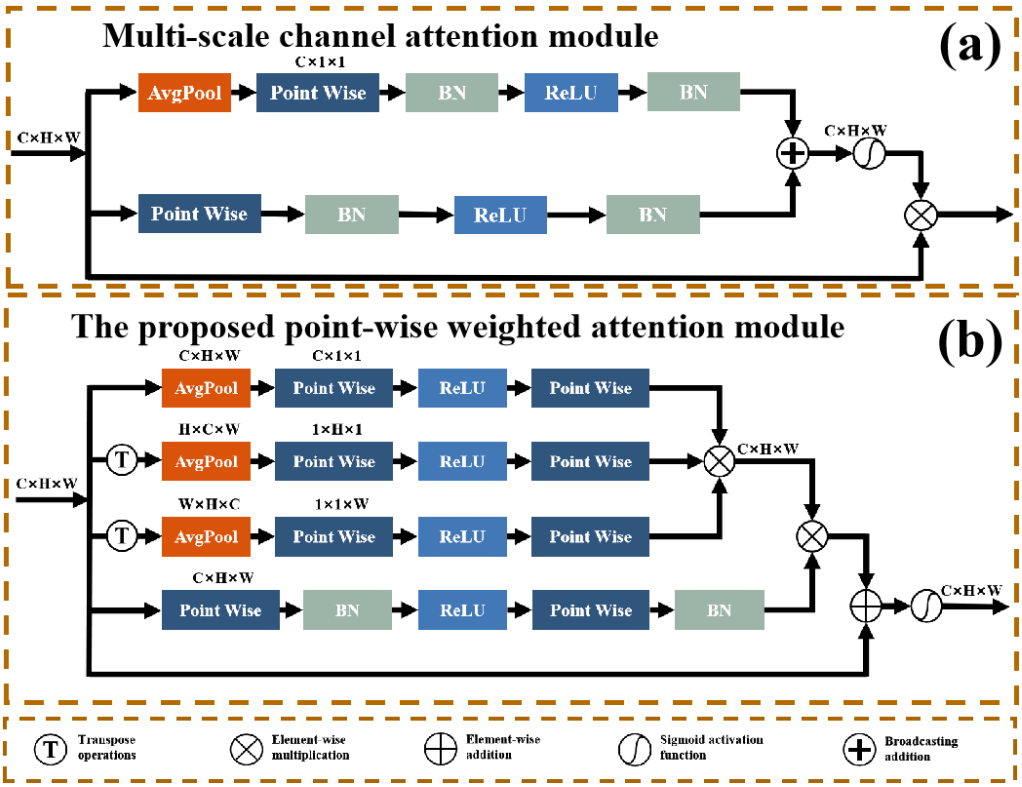

Figure 3. (**a**) Multi-scale channel attention module (MS-CAM). (**b**) Illustration of the proposed point-wise weighted attention (PWA) in our proposed method. AvgPool represents the global average pooling operation.

Then, the attention map that we obtained is multiplied by local and global information. Finally, the feature map obtained is added point-wise to the original input feature map, and the final element-wise weighed attention feature map is obtained by the sigmoid activation layer.

$$F^* = S((F_{globalatt} \oplus F_{localatt} \otimes X_c) \tag{11}$$

where X_c represents the feature map of the input, $F_{globalatt}$ is the point-wise weighted attention feature map with global context information, $F_{localatt}$ denotes the point-wise weighted attention map with local context information, and $S(\cdot)$ represents the Sigmoid activation function. The point-wise weight attention module (PWA) does not only obtain the channel weight or pixel weight but also combines the channel weight and the pixel weight, so our proposed point-wise weight attention module has stronger robustness.

3.2. Adaptive Multi-Layer Feature Fusion

High-level features have low-sampling spatial resolution, which is beneficial for haze removal but compresses more semantic context information. Compared to high-level features, low-level features possess higher resolution, which significantly aids in the localization of haze regions. Therefore, in order to fuse the feature maps from the encoder and decoder, we propose an attention multi-layer feature fusion module (AMLFF) based on a point-wise weighted attention module (PWA) to redistribute the weights of the feature maps of different scales and then obtain the fused features. The architecture of the attention multi-layer feature fusion module (AMLFF) is shown in the Figure 4. Our proposed AMLFF achieves feature-level fusion through our proposed feature fusion block (FFB), which exploits the idea of AFF. The following formula shows the principle of AFF:

$$Z = M(X \cup Y) \otimes X + (1 - M(X \cup Y)) \otimes Y \tag{12}$$

where Z is the fused feature, M represents MS-CAM, and $\cup$ denotes the initial feature integration. As shown in Figures 4 and 5, our proposed attention multi-layer feature fusion module (AMLFF) is multi-input compared with AFF, while AFF is two-input. In addition, the AMLFF is based on our proposed point-wise weighted attention module (PWA), while AFF is based on MS-CAM. The architecture of MS-CAM is presented in Figure 3a.

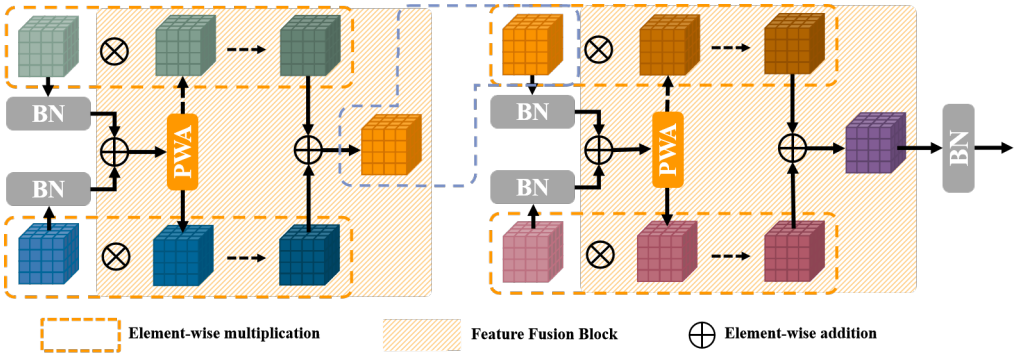

Figure 4. Structures of the proposed adaptive multi-layer feature fusion (AMLFF). AMLFF based on point-wise weighted attention (PWA) is aimed to redistribute the weights of the feature maps of different scales and then obtain the fused features.

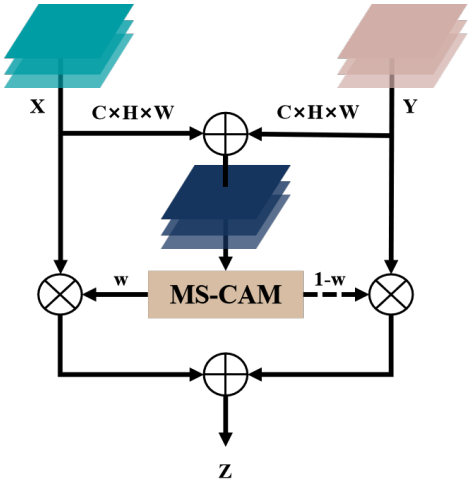

Figure 5. Structures of attention feature fusion.

Using the idea of AFF, we use our proposed point-wise weighted attention module (PWA) to design a feature fusion block (FFB). The following is the formula of the feature fusion block:

$$F(X, Y) = P(X \cup Y) \otimes X + (1 - P(X \cup Y)) \otimes Y \tag{13}$$

where P represents our proposed point-wise weighted attention module (PWA), and $\cup$ is the initial feature concatenation.

Inspired by dehaze-flow [36], the fusion of multiple sets of convolutional features helps improve the convergence of the network. Therefore, the proposed attention multi-layer feature fusion module (AMLFF) uses the up-sampled features from a decoder with a kernel size of 3×3, the up-sampled features from a decoder with a kernel size of, and the encoder down-sampled features as input. It is believed that generating a variety of dehazing results for adaptive fusion can achieve better results. The attention multi-layer feature fusion module (AMLFF) processing can be expressed as follows:

$$AMLFF(X, Y, Z) = F(\beta(F(\beta(X), \beta(Y))), \beta(Z)) \tag{14}$$

where $AMLFF(\cdot)$ is the result of adaptive multi-layer features fusion, $F(\cdot)$ represents feature fusion block (FFB), and β is the Batch Normalization (BN).

The attention multi-layer feature fusion module (AMLFF) can not only effectively balance the weight of the decoder input but also improve the robustness of the network.

3.3. Loss Function

We utilized three loss functions as the optimization objective of the proposed network: the reconstruction loss L_{rec}, the image structure similarity loss L_{ssim} [37], and the contrastive regularization L_{cr},as Equation (14).

$$L_{total} = \alpha_{rec} L_{rec} + \beta_{ssim} L_{ssim} + \lambda_{cr} L_{cr}, \tag{15}$$

Reconstruction loss. This loss can facilitate feature selection in model optimization and has proven that the training effect of L_1 loss is better than that of L_2 loss [38].

$$L_{rec} = ||J_{gt} - D(I_{haze})||_1, \tag{16}$$

where J_{gt} denotes the corresponding clean image, I_{haze} stands for input hazy image, and $D(\cdot)$ is the proposed adaptive multi-feature attention network.

SSIM loss. SSIM [12] has a good correlation with the human visual system, so we utilize the Structural Similarity Index (SSIM) loss to further enhance the visual quality of the defogged images. Hence, we define the SSIM loss as follows:

$$L_{ssim} = \frac{1}{N} \sum_{i=1}^{N} (1 - SSIM_i(D(I_{haze}, J_{gt}))), \tag{17}$$

where N represents the number of pixels and $SSIM(\cdot)$ represents the SSIM value of every pixel.

Contrastive regularization. Contrastive learning aims to learn a representation to get closer to "positive" samples in latent space and stay away from the representation between "negative" samples. Therefore, contrastive regularization (CR), which is capable of generating a better-quality image, is incorporated into the loss function as:

$$L_{cr} = \rho(V(I_{haze}), V(J_{gt}), V(D(I_{haze}, \omega))). \tag{18}$$

where ρ represents a loss function that calculates some differences or similarity between three feature representations, $V(\cdot)$ [39] is the fixed pre-trained model, e.g., VGG-19, and

$D(\cdot)$ is the proposed adaptive multi-feature attention network. Specifically, for the non-ablation mode, where ablation equals False, ρ is defined as

$$\rho = \sum_{i=1}^{N} \omega_i \frac{||h_{relu}^a - h_{relu}^p||_1}{||h_{relu}^a - h_{relu}^n||_1 + \epsilon} \tag{19}$$

For the ablation mode, when ablation equals True, ρ is defined as

$$\rho = \sum_{i=1}^{N} \omega_i ||h_{relu}^a - h_{relu}^p||_1 \tag{20}$$

In this context, ω_i represents the weight of each layer, which is defined as follows:

$$\omega_i = [\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1] \tag{21}$$

$||\ldots||_1$ denotes the L_1 loss, and ϵ is a small constant used for numerical stability.

4. Experiments

This section showcases both the quantitative and qualitative experimental results, demonstrating the effectiveness of our innovative methodology. A comprehensive experimental campaign is undertaken, comparing the performance of our method against the cutting-edge techniques on benchmark datasets such as O-HAZE [40] and RESIDE [41]. The O-HAZE [40] dataset includes 40 sets of blurred and corresponding ground truth images, while 5 such sets are reserved for testing purposes, enabling a rigorous evaluation. Furthermore, we perform an exhaustive analysis of the individual components' contributions, meticulously examining the contribution of each component within our network architecture, thereby demonstrating their individual and collective effectiveness.

4.1. Experiment Setup

Implementation details. The proposed network can be trained in a completely end-to-end manner, eliminating the need for the pre-training of individual sub-modules. We implement our experiments utilizing the PyTorch framework. For training, we employ the Adam optimizer, configured with exponential decay rates β_1 and β_2 set to 0.9 and 0.999, respectively. Each model undergoes a total of 1000 epochs of training, with an initial learning rate of 0.0001, which undergoes a step-wise reduction by a factor of three-quarters every 100 epochs. The hyperparameters of our hybrid loss function are optimized at (α_{rec}, β_{ssim}, λ_{cr}) = (1, 0.55, 0.001), ensuring effectiveness.

Datasets. To validate the proposed method, we evaluate it on both synthetic and real-world datasets. RESIDE [41] and O-HAZE [40] are chosen as our synthetic datasets for testing. On the O-HAZE [40] dataset, we perform a qualitative and quantitative comparison of our method with state-of-the-art approaches, including NLD [3], IDE [4], FFA [10], MSCNN [9], MSBDN [42], and FAD-U-Net [13]. RESIDE [41] is a comprehensive synthetic dataset. We compared with the state-of-the-art methods, which include NLD [3], FFA [10], DehazeNet [22], MSCNN [9], and MSBDN [42].

Evaluation metric and competitors. For quantitative assessment, we utilized the Peak Signal-to-Noise Ratio (PSNR) [11], Structural Similarity Index (SSIM) [12], and Reduced-Reference Partial Discrepancy (RRPD) [43] metrics. These metrics are standard for evaluating image quality in dehazing tasks, with higher PSNR [11] and SSIM [12] indicating superior image recovery and lower RRPD signifying optimized gray value distribution for enhanced image clarity. Our comparisons encompass both prior-based methods, such as NLD [3] and IDE [4], and image translation-based approaches, including FFA [10] and FAD-U-Net [13], to validate the efficacy of our novel approach.

4.2. Comparison with State-of-the-Art Methods

Results on Synthetic Dataset. We selected SOTS-OUT, a data subset of RESIDE [41] containing outdoor scenes, as the baseline synthetic dataset for the network to train and test. The quantitative results of our method and other state-of-the-art methods on synthesis datasets are shown in Table 1, and the qualitative results are exhibited in Figures 6 and 7. As depicted in Table 1, our method achieved a PSNR [11] of 29.55 dB and an SSIM [12] of 0.96. In the outdoor testing scenario, our PSNR [11] score surpasses the second-best method by a notable margin of 1.67 dB, solidifying the effectiveness of our proposed approach.

Table 1. Quantitative comparisons with the SOTA methods on the SOTS-OUT dehazing datasets.

Methods		NLD	IDE	DCPDN	DehazeNet	MSCNN	GCA	FAD-U-Net	Ours
SOTS-OUT	**SSIM**	0.85	0.78	0.66	0.87	0.86	0.86	0.93	**0.96**
	PSNR	20.66	16.40	19.39	22.00	19.75	22.47	**27.88**	**29.55**

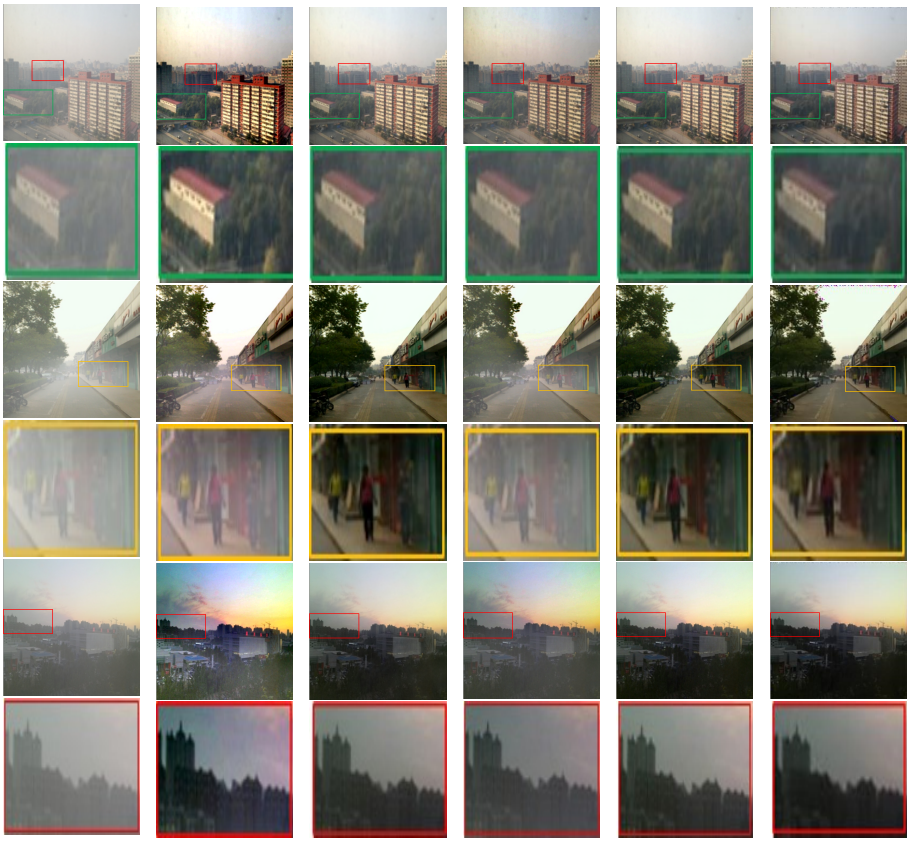

Figure 6. The dehazing results of the proposed method and the state-of-the-art methods on SOTS [41] datasets. The image is zoomed-in for a good view.

Figure 6 exhibits the enlarged restored image block on the SOTS-OUT datasets. Specificakky, the texture details of the image recovered by IDE [4] are fuzzy, and the boundaries between objects in the image recovered by FFA [10], MSCNN [9], and MSBDN [42] are blurry. Our proposed method is able to recover better the edge regions compared to other methods. In Figure 7, we compare our network with compared methods in terms of the quality of the restored images on SOTS datasets. In particular, the colors of images recovered by NLD [3] are severely distorted. FFA [10] and DehazeNet [22] leave hazy residues in the sky area of the recovered images. Similarly, although the visual performance of images recovered by MSBDN [42] is excellent, the texture between floors is still not clearly restored. The images recovered by MSCNN [9] often have higher color saturation compared to the

ground truth. In comparison, our method is able to achieve high evaluation metrics while maintaining superior visual performance.

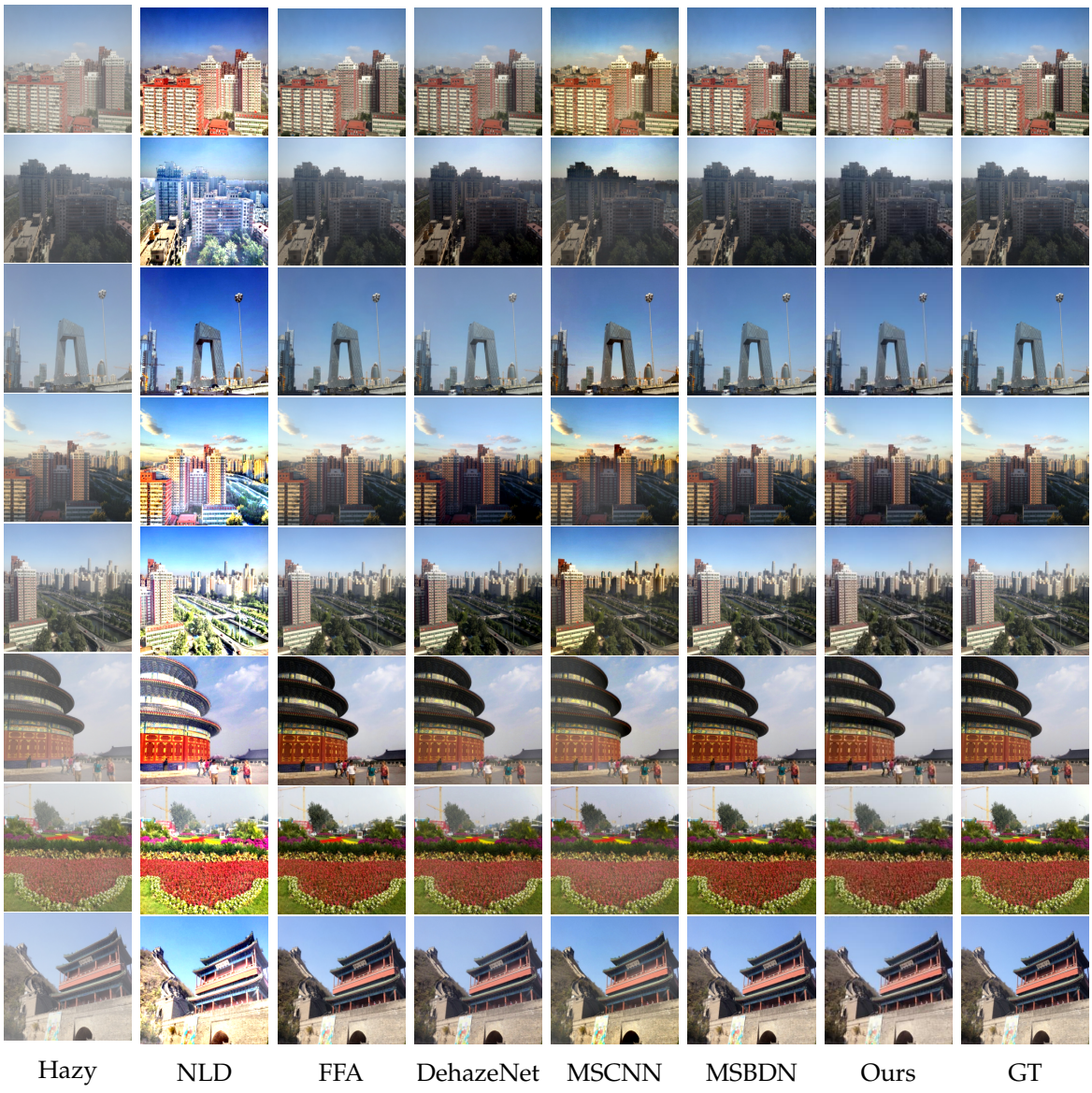

Figure 7. Visual comparison of SOTS-OUT datasets.

Results on Realistic Dataset. We have chosen the outdoor image dataset O-HAZE [40] as the realistic benchmark dataset to train and test our method. The quantitative results on this dataset are shown in Table 2; our method achieved 21.24 dB PSNR [11], 0.76 dB SSIM [12], and 12.85 RRPD [43], which surpasses the second-place 3.12 dB on the PSNR [11] in O-HAZE [40] testing datasets.

Table 2. Quantitative comparisons with SOTA methods on the O-HAZE datasets.

Methods		NLD	IDE	FFA	DehazeNet	MSBDN	MSCNN	FAD-U-Net	MFAF	Ours
O-HAZE	**SSIM**	0.61	0.52	0.42	0.67	0.74	**0.76**	0.61	0.73	**0.76**
	PSNR	13.38	11.54	16.00	16.21	**24.36**	19.07	18.53	18.26	**21.24**
	RRPD	41.46	21.73	9.69	5.08	31.27	37.31	39.39	**5.01**	**12.85**

As shown in Figure 8, our network exhibits remarkable performance in haze removal while preserving vital information like colors and edges. Notably, other popular dehazing

methods exhibit limitations. For instance, NLD [3] suffers from significant color distortion, leading to the loss of important visual details. Similarly, IDE [4] struggles to accurately estimate color information. On the other hand, FFA [10] and MSCNN [9] leave noticeable hazy areas and produce images with low overall contrast. In contrast, our network effectively mitigates these issues. The images recovered by MSBDN [42] exhibit good visual effects, but when compared to the ground truth, the overall tone appears darker. Although the image recovered by the MFAF [44] method has better color recovery, it has obvious haze residue. FAD-U-Net [13] generates blurry edges, and noise remains. Although our method is based on FAD-U-Net, the visual effects of the recovered images far surpass those of FAD-U-Net. As illustrated in Figure 6, our method effectively preserves the vibrancy of overall colors while recovering intricate details.

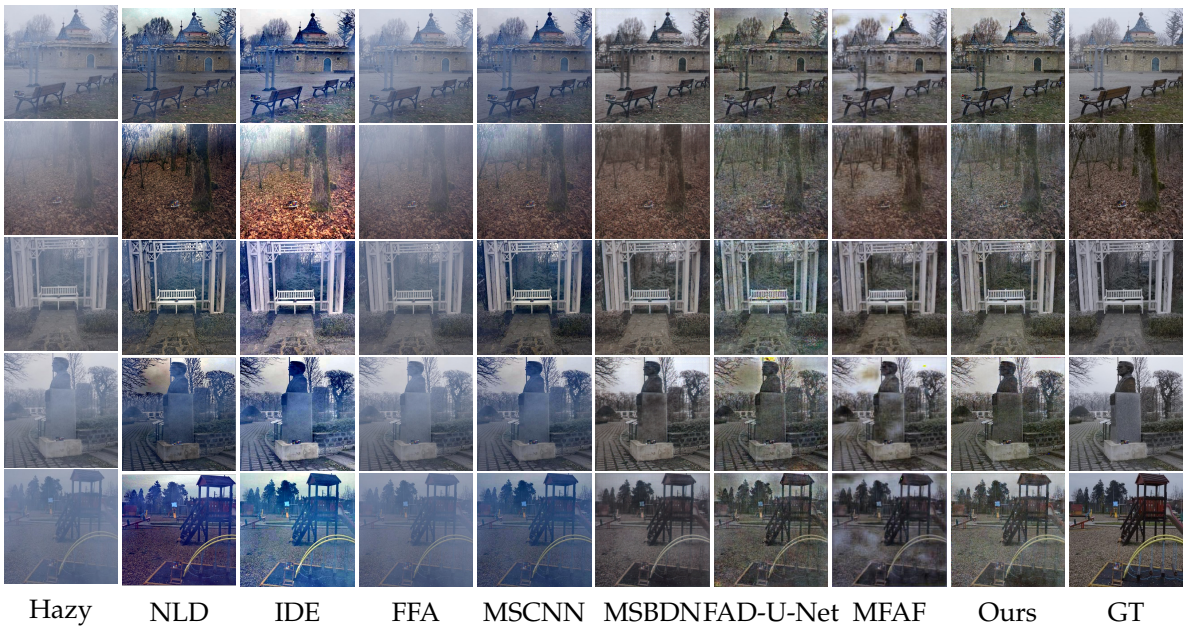

Figure 8. Visual comparison on O-HAZE datasets.

4.3. Ablation Study

To demonstrate the effectiveness of each component of the proposed model, we performed an ablation study on the O-HAZE datasets [40].

We first exploited FAD-U-Net as our baseline dehazing network. Specifically, we augmented the base network with various modules as fikkiws: (1) **base + AMLFF:** Add the attention multi-layer feature fusion module into the baseline. (2) **base + PWA:** Add the point-wise weighted attention module into baseline. (3) **base + PWA + AMLFF:** Add both point-wise weighted attention module and attention multi-layer feature fusion module into the baseline. (4) **base + SK + PWA + AMLFF:** Add the single-kernel, point-wise weighted attention module, and attention multi-layer feature fusion modules into the baseline. (5) **Ours:** Add the multi-kernel, point-wise weighted attention module, and attention multi-layer feature fusion module into the baseline.

Table 3 presents the averaged SSIM [12] and PSNR [11] results for dehazed images obtained from various model configurations. As shown in Figure 9, the PSNR [11] values of different models under different epochs are shown. Based on Table 3, we put forward the following conclusions. (1) The performance of the attention network with global information and local information is stronger than that of only the global attention network. As can be seen from the results, fully integrating both local and global information effectively enhances the performance of the network. (2) Whether the location and attention map of the sigmoid activation function is multiplied or added to the feature map is a question worth verifying. From the PSNR [11] and SSIM [12] values in the table, it can be found that the obtained attention map that is multiplied by the feature map and then passed

through the sigmoid function is the best. (3) It can be found from the table that the performance of our proposed PWA is higher than that of MS-CAM and has higher values in PSNR [11] and SSIM [12], respectively, so our proposed PWA can improve the feature fusion strategy. (4) The effect of our proposed AMLFF with three inputs is due to the two-input feature fusion strategy AFF, and the fusion of multi-scale features is beneficial to the removal of haze. It also shows that our proposed multi-scale feature fusion strategy AMLFF has better performance. (5) As evident from the table, the feature fusion through multi-layer convolutions facilitates better convergence of the network, further enhancing its performance.

Table 3. Comparisons in terms of average SSIM and PSNR on O-HAZE dataset for different configurations of AMFAN.

AMFAN Options						
Base	√	√	√	√	√	√
+PWA		√			√	√
+MS-CAM			√	√		
+AMLFF		√	√		√	√
+AFF				√		
+SK					√	
+MK						√
SSIM	0.6118	0.7569	0.7412	0.7123	0.7515	**0.7591**
PSNR	18.53	20.30	19.35	19.31	20.94	**21.24**

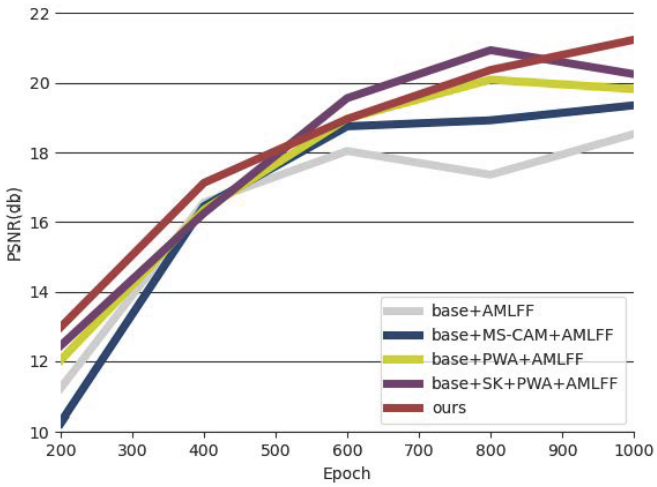

Figure 9. The average PSNRs of base + AMLFF, base + MS-CAM + AMLFF, base + PWA + AMLFF, base + SK + PWA + AMLFF and our network at the 200th, 400th, 600th, 800th, and 1000th epoch.

4.4. Comparisons of FLOPs and Parameters

To quantify the computational complexity and storage requirements of the proposed method and to demonstrate its superiority in these aspects, we have selected Floating Point Operations (FLOPs) and parameters as the indicators to compute the time and complexity of the model. FLOPs reflect the amount of computation required for the model to perform a single forward pass. The higher the FLOPs, the more computational resources the model demands, resulting in longer runtime. The parameter count represents the total number of trainable parameters in the model. The more parameters, the larger the storage space required by the model, which in turn necessitates more training data and computational resources.

The FLOPs and parameters for the state-of-the-art methods are shown in Table 4. We chose images with a resolution of 512 × 512 as the test benchmark for FLOPs. As shown in Table 4, although MSBDN [42], FFA [10], and SCANet [32] use very few parameters, their

FLOPs far exceed FAD-U-Net [13] and the proposed method. In comparison, although the proposed method is an improvement based on FAD-U-Net [13], its FLOPs value is 1.02 G lower than that of FAD-U-Net [13], indicating that the improvement in this paper can effectively enhance the computation speed for image dehazing tasks. By integrating the PSNR [11] and SSIM [12] metrics from MSBDN [42], FFA [10], and FAD-U-Net [13], it becomes evident that the method presented in this paper is capable of maintaining superior restoration performance while also achieving a lower memory footprint and higher computational efficiency. This demonstrates the effectiveness of the proposed approach in balancing restoration quality with resource utilization, making it an attractive solution for image dehazing tasks.

Table 4. The Comparisons of FLOPs and Parameters among the state-of-the-art methods.

Computation	MSBDN	FFA	SCANet	FAD-U-Net	Ours
FLOPS	166.02 G	215.07 G	258.63 G	6.32 G	5.3 G
Params	31.4 M	0.8 M	2.39 M	10.52 M	12.15 M

5. Discussion

In this paper, we propose an adaptive multi-feature attention dehazing network. To address the matter of insufficient information interaction, we designed an attention multi-layer feature fusion (AMLFF) module with three inputs in the decoder section. This module is capable of redistributing the weights of feature maps at different scales, thereby achieving the full fusion of multi-scale features. For the limited, uneven distribution of features spanning various scales, we append the point-wise weights attention module to enhance the feature fusion property of this module. Experimental outcomes, both quantitative and qualitative, on the O-Haze and SOTS datasets demonstrate the superiority of the restoration performance achieved by our proposed method compared to the majority of contemporary approaches. However, the proposed method still suffers from issues such as blurred details at the edges and a relatively large number of parameters. On this basis, we will further refine the AMLFF module to better preserve details and sharpen edges. Furthermore, our future endeavors will encompass investigating lighter attention mechanisms and crafting more potent loss functions, aimed at minimizing the parameter count and FLOPs while simultaneously enhancing the model's performance.

Author Contributions: Software, S.W.; Formal analysis, A.C.; Resources, J.C. and C.Z.; Data curation, H.J.; Writing—review & editing, H.J.; Supervision, H.J.; Project administration, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is mainly supported by National Natural Science Foundation of China (NSFC) under Grant 62473055 and the Beijing Nova Program under Grant 20230484477, and the Beijing Natural Science Foundation Program under Grant L223022, and it is partly supported by Beijing Municipal Education Commission Research Foundation under Grant KM202111417008 and the Beijing Union University under Grant ZK20202401.

Data Availability Statement: The data presented in this study are openly available in O-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Outdoor Images at https://doi.org/10.1109/CVPRW.2018.00119, reference number [40]. [O-HAZE: A Dehazing Benchmark with Real 509 Hazy and Haze-Free Outdoor Images] [https://doi.org/10.1109/CVPRW.2018.00119] [40].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Narasimhan, S.G.; Nayar, S.K. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 713–724. [CrossRef]
2. Makkar, D.; Malhotra, M. Single Image Haze Removal Using Dark Channel Prior. *Int. J. Adv. Trends Comput. Sci. Eng.* **2016**, *33*, 2341–2353. [CrossRef]

3. Berman, D.; Treibitz, T.; Avidan, S. Non-local Image Dehazing. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
4. Ju, M.; Ding, C.; Ren, W.; Yang, Y.; Guo, Y.J. IDE: Image Dehazing and Exposure Using an Enhanced Atmospheric Scattering Model. *IEEE Trans. Image Process.* **2021**, *30*, 2180–2192. [CrossRef] [PubMed]
5. Ju, M.; Ding, C.; Guo, Y.J.; Zhang, D. IDGCP: Image Dehazing Based on Gamma Correction Prior. *IEEE Trans. Image Process.* **2019**, *29*, 3104–3118. [CrossRef] [PubMed]
6. Huang, Z.; Jing, H.; Chen, A.; Hong, C.; Shang, X. Efficient image dehazing algorithm using multiple priors constraints. *J. Vis. Commun. Image Represent.* **2022**, *90*, 103694. [CrossRef]
7. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
8. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Dan, F. AOD-Net: All-in-One Dehazing Network. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
9. Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; Yang, M.H. single-image dehazing via Multi-scale Convolutional Neural Networks. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9906, pp. 154–169. [CrossRef]
10. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence 2020, New York, NY, USA, 7–12 February 2020.
11. Kamya, S. PSNR, Peak Signal to Noise Ratio. Available online: https://ww2.mathworks.cn/matlabcentral/fileexchange/41329-psnr-peak-signal-to-noise-ratio (accessed on 5 September 2024).
12. Peng, J.; Shi, C.; Laugeman, E.; Hu, W.; Zhang, Z.; Mutic, S.; Cai, B. Implementation of the Structural SIMilarity (SSIM) Index as a Quantitative Evaluation Tool for Dose Distribution Error Detection. *Med. Phys.* **2020**, *47*, 1907–1919. [CrossRef]
13. Jing, H.; Zha, Q.; Fu, Y.; Lv, H.; Chen, A. A feature attention dehazing network based on U-net and dense connection. In Proceedings of the Thirteenth International Conference on Graphics and Image Processing (ICGIP 2021), Kunming, China, 18–20 August 2021; Xiao, L., Xu, D., Eds.; SPIE: Bellingham, WA, USA, 2022; Volume 12083, p. 1208300. [CrossRef]
14. Bianco, S.; Celona, L.; Piccoli, F.; Schettini, R. High-Resolution single-image dehazing Using encoder–decoder Architecture. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019.
15. Feng, T.; Wang, C.; Chen, X.; Fan, H.; Li, Z. URNet: A U-Net based residual network for image dehazing. *Appl. Soft Comput.* **2020**, *102*, 106884. [CrossRef]
16. Lu, L.; Xiong, Q.; Chu, D.; Xu, B. MixDehazeNet: Mix Structure Block For Image Dehazing Network. *arXiv* **2023**, arXiv:2305.17654.
17. Shen, H.; Zhao, Z.Q.; Zhang, Y.; Zhang, Z. Mutual Information-driven Triple Interaction Network for Efficient Image Dehazing. In Proceedings of the 31st ACM International Conference on Multimedia, New York, NY, USA, 29 October–3 November 2023; pp. 7–16. [CrossRef]
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 4–9 December 2017; pp. 6000–6010.
19. Shi, J.; Wang, Y.; Yu, Z.; Li, G.; Hong, X.; Wang, F.; Gong, Y. Exploiting Multi-Scale Parallel Self-Attention and Local Variation via Dual-Branch Transformer-CNN Structure for Face Super-Resolution. *IEEE Trans. Multimed.* **2024**, *26*, 2608–2620. [CrossRef]
20. Meng, G.; Wang, Y.; Duan, J.; Xiang, S.; Pan, C. Efficient Image Dehazing with Boundary Constraint and Contextual Regularization. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013.
21. Fattal, R. Single image dehazing. *ACM Trans. Graph.* **2008**, *27*, 1–9. [CrossRef]
22. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [CrossRef] [PubMed]
23. Chen, Z.; Wang, Y.; Yang, Y.; Liu, D. PSD: Principled Synthetic-to-Real Dehazing Guided by Physical Priors. In Proceedings of the Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
24. Liang, Q.; Zhu, B.; Ngo, C.W. Pyramid Fusion Dark Channel Prior for Single Image Dehazing. *arXiv* **2021**, arXiv:2105.10192.
25. Zhao, D.; Xu, L.; Ma, L.; Li, J.; Yan, Y. Pyramid Global Context Network for Image Dehazing. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 3037–3050. [CrossRef]
26. Zhang, S.; Zhang, J.; He, F.; Hou, N. DRDDN: Dense residual and dilated dehazing network. *Vis. Comput.* **2022**, *39*, 953–969. [CrossRef]
27. Luo, P.; Xiao, G.; Gao, X.; Wu, S. LKD-Net: Large Kernel Convolution Network for Single Image Dehazing. In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 10–14 July 2023; pp. 1601–1606. [CrossRef]
28. Fu, M.; Liu, H.; Yu, Y.; Chen, J.; Wang, K. DW-GAN: A Discrete Wavelet Transform GAN for NonHomogeneous Dehazing. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 20–25 June 2021; pp. 203–212. [CrossRef]
29. Zhao, S.; Zhang, L.; Shen, Y.; Zhou, Y. RefineDNet: A Weakly Supervised Refinement Framework for Single Image Dehazing. *IEEE Trans. Image Process.* **2021**, *30*, 3391–3404. [CrossRef]

30. Yang, D.; Sun, J. Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing. In Proceedings of the Computer Vision—ECCV 2018: 15th European Conference, Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 729–746. [CrossRef]
31. Li, T.; Li, L.; Wang, W.; Feng, Z. Dehazing-NeRF: Neural Radiance Fields from Hazy Images. *arXiv* **2023**, arXiv:2304.11448.
32. Guo, Y.; Gao, Y.; Liu, W.; Lu, Y.; Qu, J.; He, S.; Ren, W. SCANet: Self-Paced Semi-Curricular Attention Network for Non-Homogeneous Image Dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Vancouver, BC, Canada, 17–24 June June 2023; pp. 1885–1894.
33. Zhang, H.; Patel, V.M. Densely Connected Pyramid Dehazing Network. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
34. Lin, M.; Chen, Q.; Yan, S. Network In Network. *arXiv* **2013**, arXiv:1312.4400.
35. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
36. Li, H.; Li, J.; Zhao, D.; Xu, L. DehazeFlow: Multi-scale Conditional Flow Network for Single Image Dehazing. In Proceedings of the 29th ACM International Conference on Multimedia, Online, 20–24 October 2021.
37. Wang, Z. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
38. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Image Restoration With Neural Networks. *IEEE Trans. Comput. Imaging* **2017**, *3*, 47–57. [CrossRef]
39. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
40. Ancuti, C.O.; Ancuti, C.; Timofte, R.; Vleeschouwer, C.D. O-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Outdoor Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
41. Vashishth, S.; Joshi, R.; Prayaga, S.S.; Bhattacharyya, C.; Talukdar, P. RESIDE: Improving Distantly-Supervised Neural Relation Extraction using Side Information. *arXiv* **2018**, arXiv:1812.04361.
42. Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Yang, M.H. Multi-Scale Boosted Dehazing Network with Dense Feature Fusion. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020.
43. Zhou, W.; Zhang, R.; Li, L.; Yue, G.; Gong, J.; Chen, H.; Liu, H. Dehazed Image Quality Evaluation: From Partial Discrepancy to Blind Perception. *IEEE Trans. Intell. Veh.* **2024**, *9*, 3843–3858. [CrossRef]
44. Yi, W.; Dong, L.; Liu, M.; Hui, M.; Kong, L.; Zhao, Y. MFAF-Net: Image dehazing with multi-level features and adaptive fusion. *Vis. Comput.* **2024**, *40*, 2293–2307. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.