

Article

Discerning Reality through Haze: An Image Dehazing Network Based on Multi-Feature Fusion

Shengchun Wang^{1,2}, Sihong Wang¹ , Yue Jiang³ and Huijie Zhu^{4,5,*}

¹ College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China; scwang@hunnu.edu.cn (S.W.); 202120293802@hunnu.edu.cn (S.W.)

² Hunan Provincial Meteorological Bureau, Changsha 410021, China

³ Hunan Air Traffic Management Sub-Bureau of Civil Aviation Administration of China, Changsha 410137, China; silverbreathing@gmail.com

⁴ Science and Technology on Near-Surface Detection Laboratory, Wuxi 214035, China

⁵ High Impact Weather Key Laboratory of CMA, Changsha 410007, China

* Correspondence: zhuhuijiehao@163.com

Abstract: Numerous single-image dehazing algorithms have been developed, employing a spectrum of techniques ranging from intricate physical computations to state-of-the-art deep-learning methodologies. However, conventional deep-learning approaches, particularly those based on standard convolutional neural networks (CNNs), often result in the persistence of residual fog patches when applied to images featuring high fog concentration or heterogeneous fog distribution. In response to this challenge, we propose an innovative solution known as the multi-feature fusion image dehazing network (MFID-Net). This approach employs an end-to-end methodology to directly capture the mapping relationship between hazy and fog-free images. Central to our approach is the introduction of a novel multi-feature fusion (MF) module, strategically designed to address channel and pixel characteristics in regions with uneven or high fog concentrations. Notably, this module achieves effective haze reduction while minimizing computational resources, thereby mitigating the issue of residual fog patches. Experimental results underscore the superior performance of our algorithm compared to similar dehazing methods, as evidenced by higher scores in structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and computational velocity. Moreover, MFID-Net exhibits significant advancements in restoring details within expansive monochromatic areas, such as skies and white walls.



Citation: Wang, S.; Wang, S.; Jiang, Y.; Zhu, H. Discerning Reality through Haze: An Image Dehazing Network Based on Multi-Feature Fusion. *Appl. Sci.* **2024**, *14*, 3243. <https://doi.org/10.3390/app14083243>

Academic Editor: Thomas Lindner

Received: 25 February 2024

Revised: 29 March 2024

Accepted: 7 April 2024

Published: 12 April 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/>).

Keywords: deep learning; dehazing; attention mechanism

1. Introduction

Atmospheric conditions such as fog and haze induce light refraction or scattering, impeding the direct propagation of light. The intersection of distorted rays with reflected light from observed objects leads to visual degradation in images captured by imaging devices, including diminished clarity, loss of detail, and color distortion [1,2]. These challenges not only distort the representation of the real environment but also present complications for subsequent application processing [3–6]. Therefore, the restoration of clear images from foggy ones is of paramount importance.

Single-image dehazing algorithms aim to employ specific technical methods to eliminate the influence of atmospheric environmental factors on images captured by imaging devices. These algorithms seek to maximize color and detail restoration while enhancing overall clarity. Currently, such algorithms fall into two broad categories: prior-knowledge-based algorithms [7,8] and deep-learning-based algorithms [9–12].

Prior-knowledge-based image dehazing algorithms leverage mathematical statistics to infer the common characteristics of hazy images and estimate haze-free images using either the atmospheric scattering model (ASM) [1] or its variants [13]. For instance, Schechner

et al. [14] utilized the polarization properties of light to capture polarized components in two different directions from hazy images, calculating their difference to eliminate scattered light and haze. However, this approach is limited to scenes with polarizing properties. The median filter method is applied for transmittance estimation in ASM [15], but it may obliterate details smaller than the filter's aperture size, resulting in diminished resolution and loss of detail in restored images. The dark channel prior (DCP) dehazing algorithm introduced by He et al. [16] relies on the discovery that outdoor haze-free images have at least one RGB channel with the lowest pixel values. Nevertheless, DCP may cause color distortion in large monochromatic areas such as the sky or white walls. Color prior attenuation (CPA) [17] applies color prior values to each pixel's color value in hazy images, attenuating fog accordingly. However, these algorithms heavily depend on specific prior knowledge that may not be universally applicable, leading to suboptimal dehazing results in certain scenarios (e.g., nighttime scenes, indoor environments, sky regions).

Deep-learning-based image dehazing algorithms adaptively adjust parameters according to input images, avoiding the complexities of modeling and parameter estimation. These algorithms are generally divided into two types: parameter estimation-based algorithms and direct resolution algorithms. In the former, networks estimate parameters such as transmittance and atmospheric light, followed by inverse computation using ASM to obtain haze-free images. For instance, Dehaze-Net [18] employs CNNs to learn the mapping relationship between hazy images and transmittance. AOD-Net [13] reconstructs the ASM by combining atmospheric light with transmittance into a single variable to reduce the loss of image features. DCPDN-Net [19], employing two subnetworks for the sequential learning of ambient light and transmittance, gradually optimizes different components of the network. However, DCPDN-Net may introduce image artifacts. The latter type of algorithms directly trains networks with foggy images to seek optimal dehazing parameters. For example, Yang et al. [20] proposed the concept of perceptual dehazing, which utilizes a physical model to decouple and reconstruct the hazy image into hidden feature values. Subsequently, a multiscale adversarial network is employed to restore the haze-free image. Cycle-Dehaze [21] provides unpaired hazy and clear images, which are fed into a cycle adversarial network for training. Additionally, it introduces a cycle perceptual consistency loss function to address network optimization and evaluation. GFN-Net [22] divides its architecture into feature extraction modules and restoration modules. GCANet [23] was pioneering in employing smooth dilated convolution to tackle grid artifacts in the image dehazing process. EDN-GTB [24] takes both the hazy image and dark channel transmission map as inputs for its U-Net-based model with spatial pyramid pooling modules during the training phase. AECR-NET [25] incorporates contrast regularization by treating haze-free images as positive samples while considering hazy images as negative samples, respectively. AECR-NET approximates restored images towards positive samples while moving away from negative ones, thereby enhancing the transformation capability of models accordingly.

The complexity of the image dehazing task lies in the often uneven distribution of haze within images. Faced with the uneven concentration of haze, it necessitates algorithms to pay more attention to areas with higher haze density. Deep-learning algorithms, which rely solely on traditional convolutional neural networks for dehazing, primarily focus on the local features of hazy images and neglect the overall haze density across the image [26]. This bias towards local processing inevitably leads to issues in identifying and dealing with areas of high haze density within the image. Therefore, effectively recognizing and addressing areas with an uneven distribution of haze in hazy images poses a significant challenge for dehazing algorithms [27,28]. Although the introduction of attention mechanisms can to some extent optimize the algorithm's capture of global features and afford more processing to key features, this strategy also has clear limitations [28,29]. Attention mechanisms that extract features based on fully connected (FC) layers often lead to a significant increase in the number of model parameters due to the high connectivity of FC layers. This not only exacerbates the algorithm's time complexity but also its space complexity. Moreover, an excessive number of parameters increases the risk of overfitting, leading to insufficient

generalizability of the algorithm. In the face of these challenges, dehazing algorithms must balance dehazing effectiveness with reduced demands on computational resources.

Considering that FFA-Net still fails to completely eliminate haze from images, particularly in areas of large, similarly colored blocks where the processed image pixel values are not sufficiently smooth, and further, that the channel attention mechanism employed by FFA-Net involves a large number of parameters and significant computational load, we propose an enhanced dehazing algorithm based on multi-feature fusion (MFID-Net) built upon the framework of FFA-Net to address these limitations. MFID-Net integrates multi-scale features and employs a novel lightweight attention mechanism, which not only reduces the demand for computational resources but also effectively enhances the recognition and capture of global image features. Furthermore, MFID-Net demonstrates exceptional detail restoration capabilities in scenarios with large areas of similar color blocks, such as white walls, skies, and roads. The pixel values in these areas are smoother after restoration, reducing the occurrence of black blocks and ghosting. Figure 1 clearly demonstrates the comparative effectiveness of sky scene processing, showcasing the distinct advantage of MFID-Net in handling such specialized scenarios.

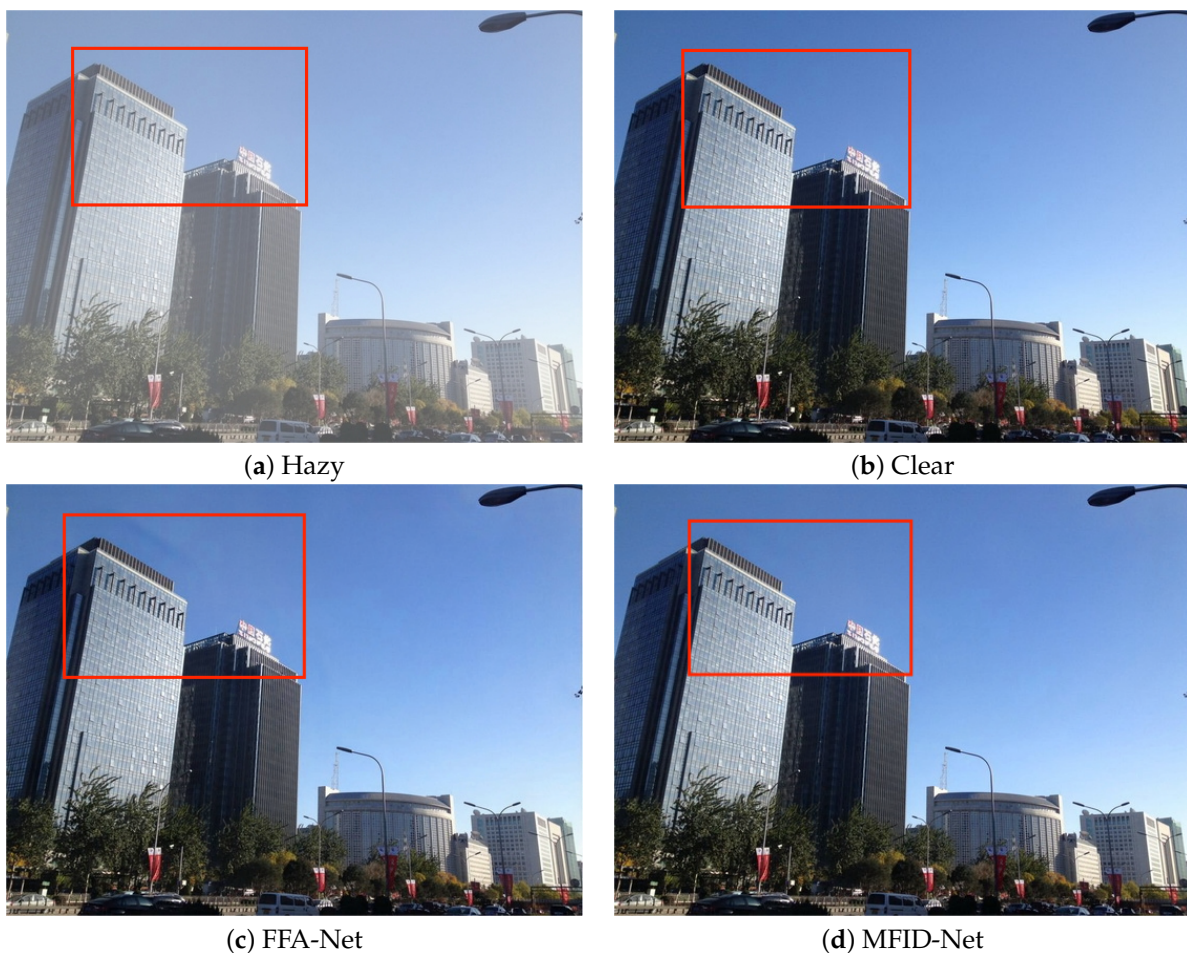


Figure 1. An example of image dehazing when processing sky scenes.

To effectively focus on the global contextual information of hazy images while reducing the algorithm's computational resource requirements, we introduce a new lightweight gated channel attention (LGCA) mechanism. LGCA utilizes regularization components and learnable parameters instead of traditional FC layers to simulate the interaction between channels, enabling better capture of global contextual information and extraction of image features in areas with high haze density. Experimental results demonstrate that LGCA

achieves significant improvements in dehazing performance for hazy images, while also reducing dependence on computational resources.

A gating mechanism plays a pivotal role in neural networks by effectively regulating the flow and selection of information between models, thereby significantly enhancing algorithmic accuracy and generalization capability [30]. The effectiveness of a gating mechanism hinges on its precise control of input–output ratios and its capacity to filter out irrelevant information while reinforcing important information. This capability enables the gating mechanism to finely adjust information, empowering the algorithm to adaptively select and integrate useful features. With this consideration, we propose a novel gating mechanism $1 + \tanh(x)$, designed to control the feature flow between channels and pixels, adaptively filtering and integrating valuable features to enhance the dehazing effect of the algorithm. Through this approach, MFID-Net exhibits improved performance and robustness.

The key contributions of this work are listed below.

- Building upon FFA-Net, we proposed MFID-Net to enhance the algorithm’s capability in detail recovery, particularly when addressing complex scenes containing extensive structural color blocks, such as white walls and sky. Table 1 presents a comparative analysis based on the FFA-Net architecture, which includes modifications such as replacing its pixel attention (PA) mechanism with the GPA, substituting the conventional channel attention (CA) mechanism with the LGCA, and integrating these improvements to develop the MFID-Net. The results indicate that MFID-Net not only achieves significant improvements in dehazing effects but also effectively reduces the algorithm’s parameter count.

Table 1. The quantitative comparison results of the impact of the key technologies introduced on FFA-Net. The arrows signify the value by which the evaluation metrics increase or decrease based on the FFA-Net baseline.

Algorithms	Indoor Dataset (IST)		Outdoor Dataset (OST)	
	PSNR	SSIM	PSNR	SSIM
FFA	36.13	0.9842	33.38	0.9839
FFA + GPA	36.58 (0.45↑)	0.9845 (0.0003↑)	33.46 (0.08↑)	0.9839
FFA + LGCA	37.21 (1.08↑)	0.9880 (0.0038↑)	33.70 (0.32↑)	0.9841 (0.0003↑)
MFID-Net	37.42 (1.29↑)	0.9890 (0.0048↑)	34.21 (0.83↑)	0.9844 (0.0005↑)

- An innovative multi-feature fusion (MF) module is proposed, focusing on the efficient integration of channel and pixel features within images. The MF module ingeniously combines three distinct components to work collaboratively, thereby significantly enhancing the algorithm’s expressive capability in image processing.
- Proposed LGCA, utilizing ℓ_1 -Norm and ℓ_2 -Norm methods in lieu of FC layers to reduce model parameters and alleviate computational resource requirements. Furthermore, LGCA allows the algorithm to concurrently focus on both local and global information, enhancing feature extraction capabilities.
- A novel gated pixel attention (GPA) module has been introduced, aimed at enhancing the cooperation and competition among pixel features, thereby significantly improving the flexibility and adaptability of dehazing algorithms. This innovative GPA module demonstrated superior dehazing performance in processing foggy images.

The remaining sections of this work are structured as follows. Section 2 provides an overview of the existing literature in the field of image dehazing. Section 3 elaborates on the implementation methods and details of the proposed model. Section 4 presents experimental results and facilitates discussion. Finally, Section 5 offers a concise summary encompassing this work.

2. Related Work

2.1. Atmospheric Scattering Model

The atmospheric scattering model (ASM) is used to explain the imaging process of devices capturing images under atmospheric environments such as fog and haze. This model simulates the propagation of light through the atmosphere, calculating the extent to which the brightness value of each pixel in the image is affected by atmospheric scattering, thus providing a robust theoretical basis for the task of image dehazing. The model was originally proposed by McCartney et al. [1], and it has been continuously improved in subsequent research [2,31]. It can be expressed as follows:

$$I(z) = J(z)t(z) + A(1 - t(z)) \quad (1)$$

where z is the pixel coordinate values of the image, $I(z)$ is the single foggy image, $J(z)$ is the haze-free image, A is the global atmosphere light, and $t(z)$ is the transmission rate, which can be described as follows when light propagates through a homogeneous medium:

$$t(z) = e^{-\beta d(z)} \quad (2)$$

where β is the atmospheric scattering coefficient and $d(z)$ is the distance between the object and the imaging equipment. In the task of image dehazing, $t(z)$ and A are estimated from the foggy image using prior knowledge or other methods; by substituting it into Equation (1), the dehazed image can be restored.

$$J(z) = \frac{A(1 - e^{-\beta d(z)}) - I(z)}{e^{-\beta d(z)}} \quad (3)$$

2.2. Attention Mechanism

The attention mechanism, a computational model that imitates the pattern of human attention, primarily addresses the issue of selective attention in information processing. It allows neural networks to concentrate on essential parts of the information being processed while neglecting insignificant or irrelevant parts. To date, the attention mechanism has been widely applied in multiple deep-learning domains such as natural language processing, computer vision, and speech recognition, including classic structures like SE-Net [28], ECA-Net [32], CBAM [33], and SAM [34]. Xu et al. [35] were the first to employ the attention mechanism for the resolution of computer vision problems in 2015, laying a solid foundation for the mechanism's development within computer vision. Since then, the application field of the attention mechanism has continually expanded, making it an indispensable technology in deep-learning tasks.

In the task of single-image dehazing, the attention mechanism is utilized to compute the weights of different regions of the hazy image, enabling the network to focus on the areas most severely affected by haze. Moreover, the attention mechanism assists the network in striking a balance between global and local considerations. Global attention helps the network focus on overall features to restore the overall clarity of the image. In contrast, local attention emphasizes the recovery of details in local areas, enhancing the image's detailed effects, taking into consideration that both global and local attention can yield superior dehazing results. GridDehazeNet [36], inspired by SE-Net, introduced a multi-scale estimation method based on channel attention. This method effectively mitigated the bottleneck problem traditional multi-scale networks encounter due to information flow restrictions in the hierarchical structure, providing the network with more flexible information exchange and aggregation abilities. Xu et al. [27] proposed a feature fusion attention network (FFA-Net) in 2020, addressing the issue of uneven haze density in different areas of hazy images by constructing a feature attention module combining channel and pixel attention mechanisms. However, FFA-Net does not perform well when processing images with large areas of color blocks, causing blurry image details and potential double-image effects. Moreover, the repeated invocation of the feature attention module leads to excessive

computational resource consumption. Tu et al. [37] were the first to propose a generic low-level visual U-Net backbone network based on MLP, combined with residual channel attention for image dehazing. While the attention mechanism played a pivotal role in enhancing dehazing effects, a balance must still be struck between computational resource consumption and improvements in dehazing results. Therefore, finding a more efficient attention mechanism to reduce computational resource demand and improve algorithm generalizability remains a significant challenge.

3. Multi-Feature Fusion Image Dehazing Network

Deep-learning models are essentially comprised of a complex set of nonlinear functions, with the core objective being the gradual optimization of model parameters to minimize the value of the loss function. The task of image dehazing is formalized as an optimization problem of a nonlinear function F . Through the analysis of a vast array of data samples, a set of optimal parameters is sought. These parameters are capable of effectively transforming hazy images $X \in \mathbb{R}^{H \times W \times C}$ into their corresponding clear images $Y \in \mathbb{R}^{H \times W \times C}$, where C is the number of channels in the image and H and W denote the height and width of the image, respectively. This problem is modeled in the following form in this paper:

$$\theta^* = \operatorname{argmin}_{\theta} L(F(X; \theta), Y) \quad (4)$$

where θ^* is the set of parameters that minimize the loss function, $F(X; \theta)$ is the predicted output of the haze-free image by the function F when given the parameter set θ , and L is the loss function, utilized to quantify the difference between the predicted haze-free image and the actual haze-free image.

Based on the FFA-Net architecture, we propose the multi-feature fusion image dehazing network (MFID-Net) as the prediction function F to approximate the clear image Y with high fidelity; the differences between FFA-Net and MFID-Net are illustrates in Figure 2. In Figure 3, MFID-Net takes the $X \in \mathbb{R}^{H \times W \times C}$ as input and extracts shallow features through a convolution operation layer. Subsequently, these features are fed into three consecutive Group Blocks for further feature extraction. Depth fusion across different levels is achieved through residual connections, resulting in the feature map $g(X)$, a process detailed in Equation (5). Specifically, each Group Block is composed of 15 multi-feature fusion (MF) modules and a convolutional layer, with residual connections also employed within the Group Block to enhance the transmission and fusion of features. Next, lightweight gated channel attention (LGCA) and gated pixel attention (GPA) are utilized to re-calibrate the features at the channel and pixel levels to enhance the dehazing effect. Finally, the predicted haze-free image $F(X) \in \mathbb{R}^{H \times W \times C}$ is generated through two convolution operations, a process that can be represented by Equation (6).

$$g(X) = \operatorname{Concat}(g_1(\operatorname{Conv}(X)), g_2(g_1(\operatorname{Conv}(X))), g_3(g_2(g_1(\operatorname{Conv}(X))))) \quad (5)$$

$$F(X) = \operatorname{Conv}(\operatorname{Conv}(\operatorname{GPA}(\operatorname{LGCA}(g(X))))) + X \quad (6)$$

where Conv is a convolution operation that is characterized by a kernel size of 3 and a stride of 1. g_i is the Group Block that maintains a constant channel size during computation, $i = 1, 2, 3$. LGCA and GPA represent the lightweight gated channel attention mechanism and gated pixel attention mechanism proposed in MFID-Net, respectively, which are components of the MF module.

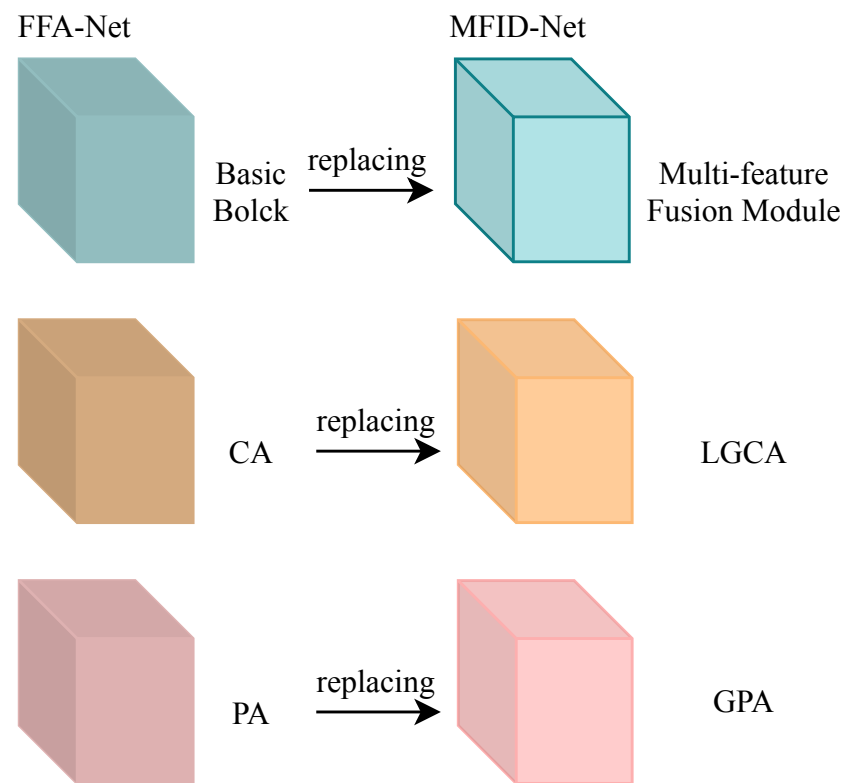


Figure 2. The differences between FFA-Net and MFID-Net.

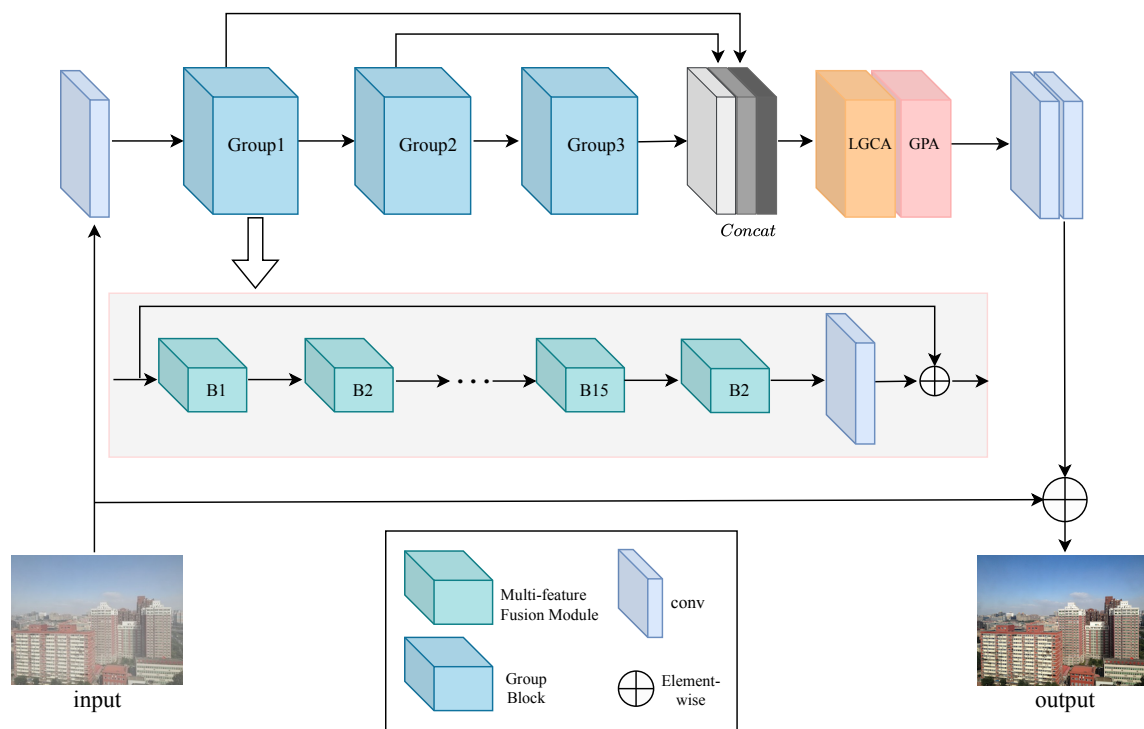


Figure 3. Structure of multi-feature fusion image dehazing network.

3.1. Multi-Feature Fusion Module

We propose the multi-feature fusion (MF) module, designed for efficient extraction and fusion of image channel and pixel features, resolving the problem of residual fog patches in cases of high or uneven fog concentration. As shown in Figure 4, MF comprises three

parts: a shallow feature extraction layer, LGCA, and GPA. The shallow feature extraction layer is responsible for learning the basic features of the image, such as edges, textures, etc. LGCA identifies and enhances channel features of significance. GPA, learning from the relationships among pixels and based on the importance of each pixel, performs fine-grained image adjustments, avoiding excessive smoothing or enhancement of local features. These three components work in conjunction, collectively improving the module's feature extraction and fusion capabilities. Experimental results show that the algorithm performs better when convolutions are placed after the attention mechanisms.

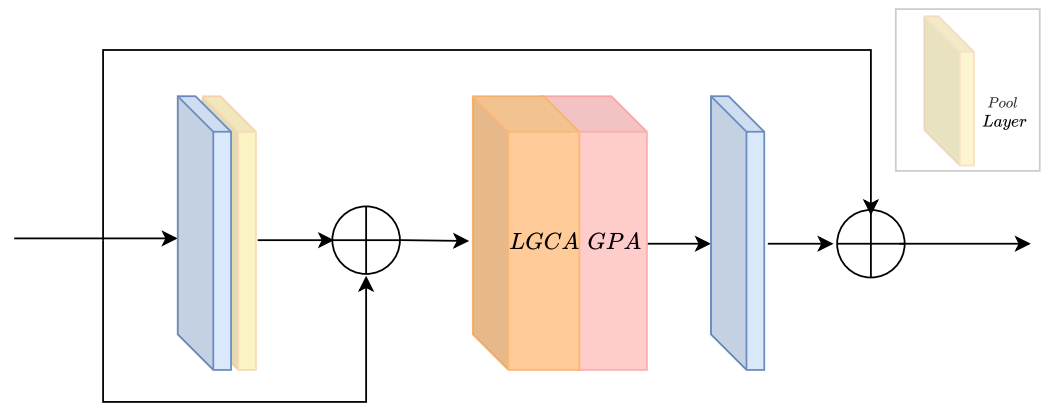


Figure 4. Structure of multi-feature fusion module.

3.2. Lightweight Gated Channel Attention

The FFA-Net [27] utilizes SE-Net [28] as the channel attention mechanism to capture the interrelationship among feature channels. However, this design introduces two issues. First, SE-Net relies on two FC layers to handle channel embeddings, which imposes significant demands on computational resources. Secondly, due to the complexity of the parameters in the FC layers, understanding and analyzing the inter channel interactions among different layers becomes exceedingly challenging. FC layers implicitly learn the channel relationships, leading to neuron output behaviors that are difficult to interpret. To address these problems, we propose a lightweight, interpretable, and efficient attention mechanism LGCA.

Figure 5 illustrates the detailed structure of LGCA, which comprises three components: a global context embedding layer, a channel normalization layer, and a gated adaptive layer. Initially, a parameterless $\ell_1 + \ell_2$ -norm is employed by the global context embedding layer to aggregate global context information. To adjust the proportions of ℓ_1 -norm and ℓ_2 -norm regularization and to endow the LGCA with learning capabilities, a global context embedding parameter α is introduced, which is capable of embedding global context information and adjusting the weights of individual channels before channel normalization. Subsequently, the stability and generalizability of the algorithm are ensured by the channel normalization layer. Finally, based on the output of the channel normalization layer, the gated adaptive layer adjusts the input channel features using gated weights γ and bias β , and competition and cooperation among features are promoted through a gated “ \tanh ” activation function. It has been demonstrated by experiments that the image dehazing effect is enhanced by LGCA while utilizing a reduced number of parameters.

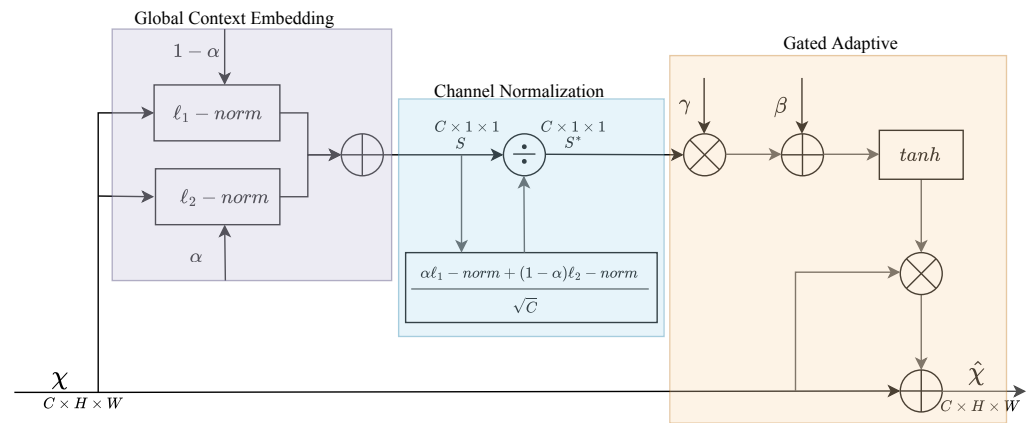


Figure 5. Structure of lightweight gated channel attention mechanism.

$$\hat{\chi} = \hat{F}(\chi|\alpha, \gamma, \beta), \alpha, \gamma, \beta \in \mathbb{N}^C \quad (7)$$

where $\chi \in \mathbb{N}^{C \times H \times W}$ is a feature map, where H and W are the height and width of the feature map and C is the channel of the map. Let $\chi = [\chi_1, \chi_2, \dots, \chi_C]$, and χ_c is a corresponding feature for each channel in χ , where $\chi_c = [X_c^{i,j}]_{H \times W} \in \mathbb{N}^{H \times W}$ and $c = \{1, 2, \dots, C\}$. The final output result $\hat{\chi}$ is determined collectively by the trainable parameters α , γ , and β , where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_C]$, $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_C]$, and $\beta = [\beta_1, \beta_2, \dots, \beta_C]$. It is worth noting that the parameter complexity of LGCA is $O(C)$, significantly less than the $O(C^2)$ of SE-Net.

Global Context Embedding Layer: The performance of an algorithm is often constrained by the quality and quantity of the dataset. The aggregation of global context information can provide a broader and more comprehensive data background, thereby improving the precision of understanding input data. Importantly, the integration of global context information can help mitigate the influence of local features on algorithm performance, thereby enhancing the model's robustness and generalizability. Alex et al. pointed out that a regularization function can establish connections among neurons [38,39]. In this regard, ℓ_1 -norm is capable of reducing irrelevant feature parameters that have minimal or no impact on the task to zero, exploiting the sparsity of the network to achieve rapid feature selection. ℓ_2 -norm, conversely, assigns higher weights to critical features and lower weights to secondary features, thereby preventing model overfitting. Based on these principles, LGCA has designed a global information embedding layer, the implementation of which is detailed in Equations (8)–(10).

$$\ell_1 = |\sum_{i=1}^H \sum_{j=1}^W \chi_c^{i,j}| \quad (8)$$

$$\ell_2 = \left\{ \sum_{i=1}^H \sum_{j=1}^W (\chi_c^{i,j})^2 + \epsilon \right\}^{\frac{1}{2}} \quad (9)$$

$$S_c = \alpha_c \ell_1 + (1 - \alpha_c) \ell_2 \quad (10)$$

where ℓ_1 is the absolute value of the channel features of each pixel coordinate in the image and ℓ_2 is the square root of the sum of squares of the channel values of the image pixel coordinates. Moreover, a constant ϵ is set to prevent the derivative from being zero.

Channel Normalization Layer: Normalization layers are employed to accentuate channels with high feedback values and suppress those with lower feedback values, thus achieving a dynamic adjustment of features [40–42]. LGCA utilizes $\ell_1 + \ell_2$ -norm formulations to normalize the channels, optimizing the weight distribution among them. The formula for channel normalization is as follows:

$$S_c^* = \frac{\sqrt{C} S_c}{\|S_c\|_2} = \frac{\sqrt{C} S_c}{[(\sum_{c=1}^C S_c^2 + \epsilon)]^{\frac{1}{2}}} \quad (11)$$

where $S = [S_1, S_2, \dots, S_c]$; to avoid excessively small values of S_c when c is too large, we introduce a constant \sqrt{C} , and a small constant ε is introduced to prevent abnormal algorithm performance when the denominator is zero. Notably, the channel normalization method proposed in this study ($O(C)$) involves much less computation than FC ($O(C^2)$).

Gated Adaptive Layer: The gating mechanism is capable of enhancing the overall performance of the algorithm by inducing competition and cooperation among channel features [43]. Concurrently, the gating mechanism is instrumental in preventing the potential loss of features within deep networks, augmenting the algorithm's flexibility and robustness and enabling adaptability to various task requirements. The LGCA augments each original channel feature $\hat{\chi}$ by integrating a gating mechanism, thereby not only increasing the algorithm's flexibility but also ensuring the integrity of the original feature $\hat{\chi}$. The specific definition of the gated adaptive layer is as follows:

$$\hat{\chi}_c = \hat{\chi}[1 + \tanh(\gamma_c S_c^* + \beta_c)] \quad (12)$$

The operation of the gating mechanism allows the network to dynamically adjust. When a channel's gating weight is positively activated, the channel is encouraged to compete with other channels. Conversely, if the gating weight is negatively activated, it encourages the channel to cooperate with other channels. That is, when both weights γ and biases β are zero, the original features are passed to the next layer without modification.

3.3. Gated Pixel Attention

The pixel attention mechanism adjusts the weight of each pixel in the image, thereby directing the algorithm's focus towards areas with higher fog density. MFID-Net incorporates a gated pixel attention mechanism (GPA) to foster cooperation and competition among pixel features, enhancing the efficiency of defogging. The specific structure is depicted in Figure 6. Initially, the GPA sequentially conducts convolution, pooling, and convolution operations on the feature map $\hat{\chi}$ to preliminarily extract features. Subsequently, the degree of activation for the gating mechanism is controlled utilizing weight parameters $\delta = [\delta_1, \delta_2, \dots, \delta_C]$ and biases $\mu = [\mu_1, \mu_2, \dots, \mu_C]$. During the activation process, the GPA employs the \tanh function to modulate pixel-level features. Ultimately, the features processed by the activation function are subjected to multiplication operations and residual connections with the original feature map $\hat{\chi}$, resulting in the feature map X^* , as manifested in Equation (13).

$$X_c^* = \hat{\chi}_c(1 + \tanh(\delta_c(\text{Conv}(\text{Pool}(\text{Conv}(\hat{\chi}_c)))) + \mu_c)) \quad (13)$$

where in the shallow layers of the algorithm, δ_c is set to be less than zero, indicating that they can promote collaboration between pixel features, which aids in more comprehensively capturing and utilizing image information. Conversely, in the deeper layers of the algorithm, δ_c is set to be greater than zero, leading to competition between pixel positions, which helps to highlight key features and suppress unnecessary information. Empirical evidence demonstrates that GPA optimizes the defogging effect of MFID-Net.

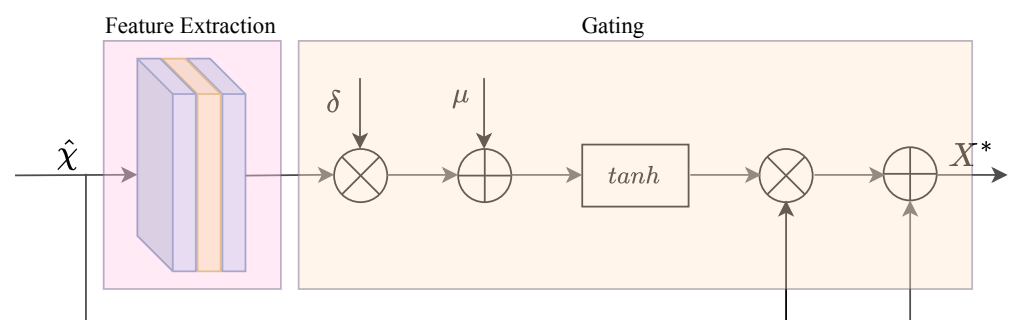


Figure 6. Structure of gated pixel attention mechanism.

3.4. Loss Function

In the field of image dehazing, commonly used loss functions primarily include L1, L2, and structural similarity index (SSIM) loss functions. Zhao et al. [44] posited that different loss functions might engender distinct issues in varying application contexts. While relying on the L2 loss function can achieve certain dehazing effects, it may lead to rasterization distortion in the restored images, negatively impacting the overall image quality. Employing the L1 loss function may result in relatively higher peak signal-to-noise ratio (PSNR) values for images, but this does not necessarily indicate that it is the optimal approach. Although using the L1 loss function enhances the PSNR values of images, it does not imply that it is the best choice. This is because the L1 loss function does not take into account the variances in human visual perception of images, which may lead to significant visual discrepancies between the restored images and the original authentic images. In contrast, the SSIM loss function, which integrates structural similarity and luminance information, better reflects human visual perception of image quality.

To achieve the dual objectives of optimizing visual quality and accuracy in image dehazing, this chapter adopts a hybrid loss function—MS-SSIM + L1 loss [44]—as specified in Equation (14). This mixed loss function ingeniously combines multi-scale structural similarity (MS-SSIM) with L1 loss, reducing prediction error while maintaining image structural details. The MS-SSIM component is employed to assess and optimize the structural similarity of images across different scales, while the L1 component focuses on minimizing the absolute error at the pixel level, thereby enhancing the accuracy of the restored images.

$$L_{MS-SSIM+L1} = \rho \cdot \mathcal{L}^{MS-SSIM} + (1 - \rho) \cdot G_{\sigma_G^M} \cdot \mathcal{L}^{\ell_1} \quad (14)$$

where ρ is a constant employed to balance the proportions of two loss functions. The experimental results of Zhao et al. [44] indicate that the optimal dehazing effect can be achieved when $\rho = 0.84$, and G represents the parameters of the Gaussian distribution. $\mathcal{L}^{MS-SSIM}$ and \mathcal{L}^{ℓ_1} represent the MS-SSIM and L1 loss functions, respectively, with their specific formulas provided in [44]. By training with this loss function, MFID-Net is able to produce more realistic and clearer dehazed images.

4. Experiment and Result Analysis

4.1. Experimental Environment

In order to comprehensively evaluate the performance and effects of the algorithm, it is necessary to use datasets for training, testing, and validation. The commonly used image dehazing datasets include NYU2 [45], D-HAZY [46], and RESIDE [47]. We chose RESIDE for training, testing, comparison, and ablation experiments. This dataset consists of five subsets, including multiple synthetic and real foggy images. We selected 13,990 indoor and 13,990 outdoor images from it to ensure that the algorithm could handle images under various scenarios. To further validate the effect of the algorithm, we also selected 500 indoor and 500 outdoor images from the SOTS subset of the RESIDE dataset as the validation set. To evaluate the performance of the algorithm in real applications, we chose the real dataset from RESIDE for validation.

To ensure the fairness of the experimental results, all algorithms were trained in the PYTORCH framework and on two GTX 1080ti graphics cards. During the training process, we used the Adam optimizer for 100 epochs of training, with 5000 samples trained per epoch. The initial learning rate was set to 0.0001, and the learning rate was adjusted using the cosine annealing strategy until the learning rate dropped to zero. Finally, we chose the best-performing result during training as the final result of the experiment.

4.2. Evaluation Metrics

The evaluation of dehazing algorithms is commonly divided into subjective and objective assessments. Subjective assessment is primarily based on human visual perception to judge factors such as image clarity, color fidelity, and contrast. This method of evaluation

is intuitive and easily understood, yet it possesses a certain degree of subjectivity due to the influence of personal preferences and observation conditions. Objective assessment, on the other hand, is conducted through quantifiable metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), which quantify the dehazing effects and provide a replicable, standardized evaluation outcome.

Currently, PSNR and SSIM are widely accepted and employed as benchmarks for evaluation, allowing for the convenient comparison of the effectiveness between different dehazing algorithms. Consequently, we select PSNR and SSIM as the evaluation metrics for dehazing effectiveness. Specifically, SSIM measures the structural similarity between images, taking into account variations in brightness, contrast, and structure. By focusing on visual perception quality, SSIM is able to reflect the subjective experience of image quality more accurately, in Equation (15). PSNR, being a pixel-level metric based on error, is used to measure the similarity between the original and the processed images, proving highly effective in assessing image distortion, in Equation (16).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (15)$$

where x and y are the two images being compared, μ_x and μ_y are the mean grayscale values of the images, and c_1 and c_2 are constants used to prevent division by zero.

$$PSNR = 10 * \log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (16)$$

MAX is the maximum color value of the image and MSE is mean squared error, which is defined as follows:

$$MSE = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m ||X(i, j) - Y(i, j)||^2 \quad (17)$$

where $m * n$ is the pixel values of the image.

Concurrently, in order to comprehensively assess the performance of the LGCA algorithm, not only was the effectiveness of the algorithm itself considered, but also two critical indicators; namely, the number of parameters (Params) and computational cost (FLOPs) were specifically introduced for an in-depth analysis.

4.3. Results on RESIDE Dataset

In this section, our proposed algorithm MFID-Net is compared with classical single-image dehazing algorithms (DCP, AOD-Net, GCA-Net, GridDehazeNet, FFA-Net) to evaluate their dehazing effects on indoor datasets (IST) and outdoor datasets (OST). Table 2 presents a quantitative comparison of SSIM, PSNR, and rate of improvement-time on IST and OST. It is evident that MFID-Net achieved the best results in terms of both SSIM and PSNR indicators. Moreover, compared to GridDehazeNet and FFA-Net, which are based on multiple stacked attention mechanisms, MFID-Net also demonstrates superiority in computational speed. Figure 7 demonstrates the convergence properties of MFID-Net on the IST and OST datasets, validating the algorithm's stability throughout the training process and unequivocally illustrating its ability to successfully and robustly converge to the anticipated optimal solution.

Table 2. Quantitative comparisons for different methods. The Rate of improvement-time indicates the average speed improvement in processing an image of the same size compared to DCP.

DateSet	Evaluation	DCP	AOD-Net	GCA-Net	GridDehazeNet	FFA-Net	MFID-Net
IST	PSNR	16.86	19.26	30.23	32.16	36.13	37.42
	SSIM	0.8601	0.8414	0.9501	0.9836	0.9842	0.9890
OST	PSNR	19.37	20.17	29.76	30.86	33.38	34.21
	SSIM	0.8431	0.8769	0.9498	0.9752	0.9839	0.9844
Rate of improvement-time		1	97.31%	97.39%	91.03%	94.94%	96.68%

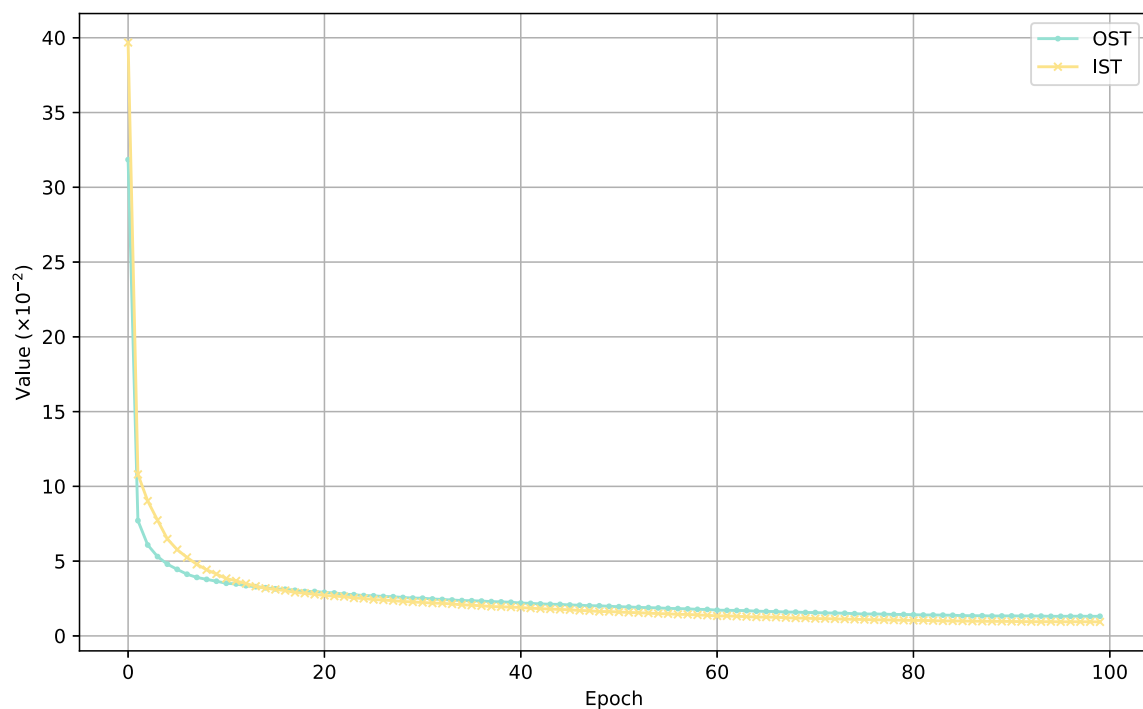


Figure 7. The convergence characteristics of MFID-Net on the IST and OST datasets.

Figures 8 and 9 provide a visual comparison of MFID-Net with other techniques on IST and OST. Although DCP removed most of the haze, the color saturation of its restored images was too high, leading to noticeable distortions. AOD-Net only eliminated superficial fog; the amount of remaining fog residue increased with rising haze concentration. GCA-Net still left a small amount of fog residue in areas with high fog concentration. Despite significant improvements in dehazing by GridDehazeNet and FFA-Net, they could still be enhanced in terms of image detail restoration, and problems like color block loss were common. In contrast to these algorithms, MFID-Net showed excellent performance in removing image haze and restoring image details and color. The images restored by MFID-Net were closer to the original images.

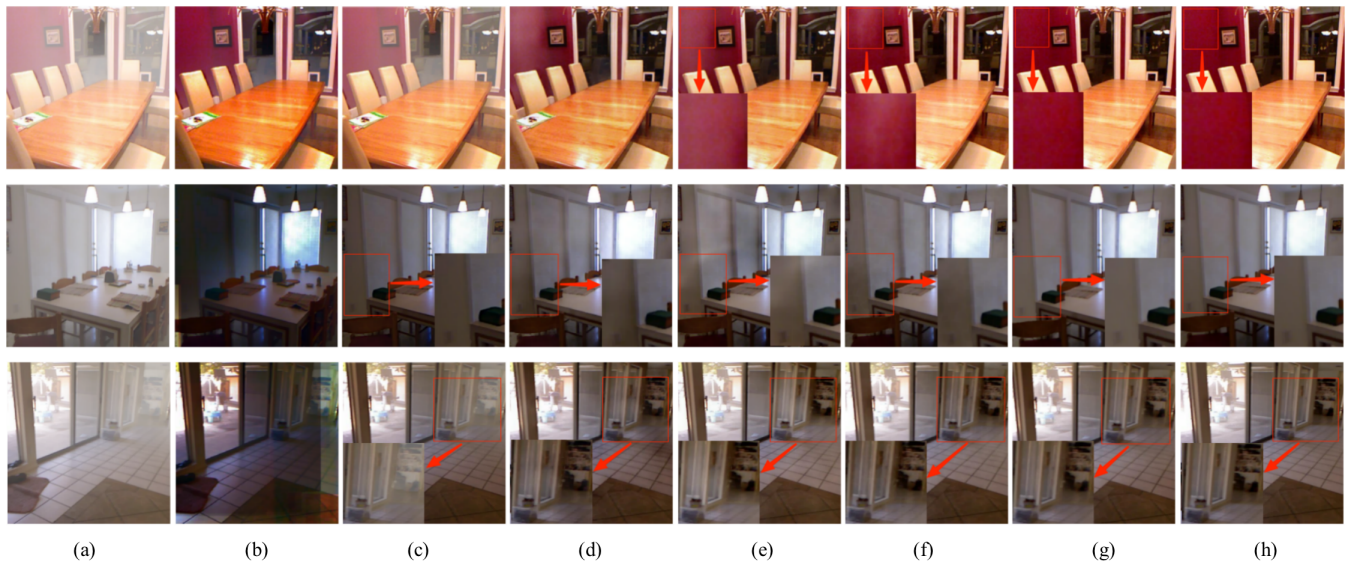


Figure 8. Indoor results. (a) Hazy, (b) DCP, (c) AOD-Net, (d) GCA-Net, (e) GridDehazeNet, (f) FFA-Net, (g) MFID-Net, (h) Clear.

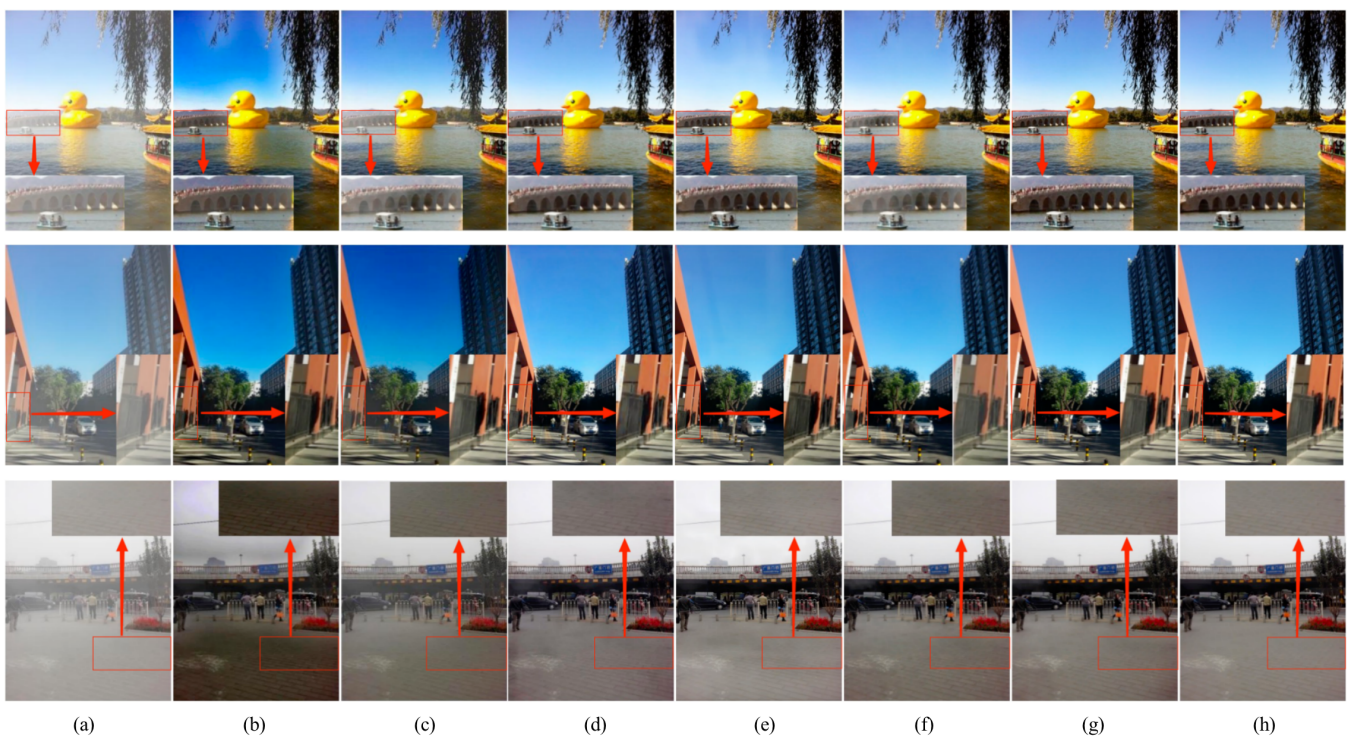


Figure 9. Outdoor results. (a) Hazy, (b) DCP, (c) AOD-Net, (d) GCA-Net, (e) GridDehazeNet, (f) FFA-Net, (g) MFID-Net, (h) Clear.

To further validate the superiority of MFID-Net in dehazing large areas of color blocks such as roads and sky, we selected parts of the outdoor dataset for comparison. The dehazing comparison is shown in Figure 10. It is clearly seen that DCP, when handling areas like the sky and road, tends to generate high-saturation images, and sometimes produces black blocks when dealing with large white areas. The images produced by AOD-Net have low contrast in the background. When GCA, GridDehazeNet, and FFA-Net deal with large color block areas such as sky and roads, their restoration of image details is not ideal, often resulting in problems like black blocks and ghosting. In contrast, MFID-Net, when handling large color block areas, is superior to the aforementioned methods in terms

of restoring image color and preserving image details. More importantly, it also reduces the occurrence of undesirable phenomena such as black blocks and ghosting.

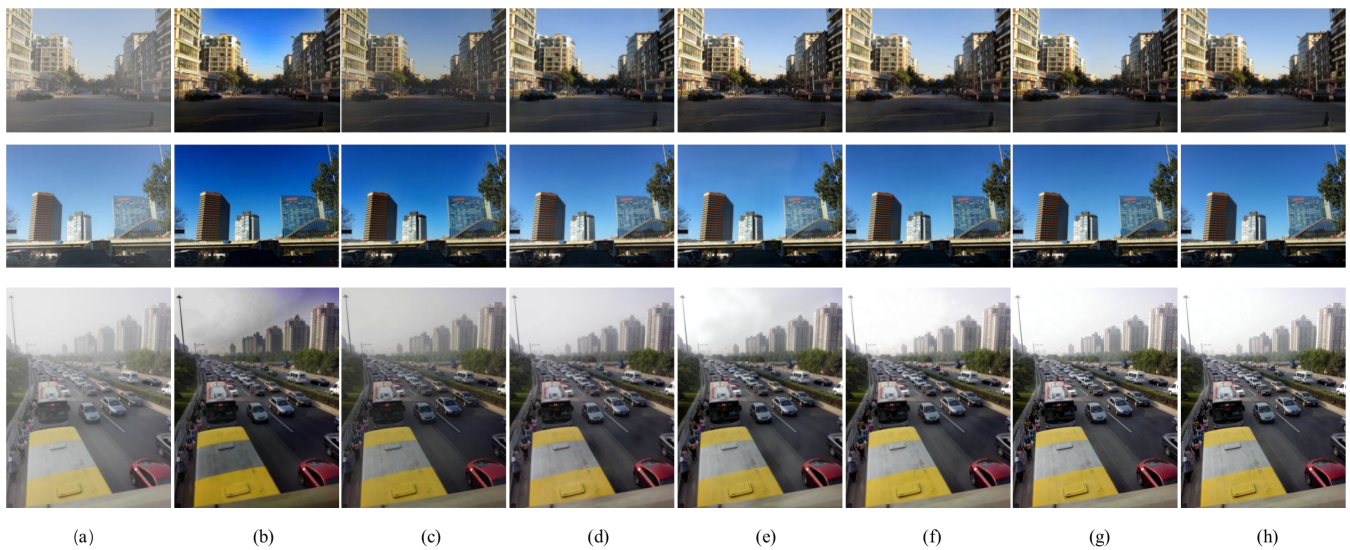


Figure 10. Large area color block results. (a) Hazy, (b) DCP, (c) AOD-Net, (d) GCA-Net, (e) GridDehazeNet, (f) FFA-Net, (g) MFID-Net, (h) Clear.

In the real world, foggy images usually have an uneven distribution of fog density, which adds complexity to the dehazing task of algorithms in real environments. Traditional single-image dehazing algorithms, when dealing with this issue, often lead to problems such as color deviation, loss of detail, and poor edge handling in the recovered images. Moreover, some deep-learning-based dehazing algorithms demonstrate insufficient generalization capability when dealing with real foggy images. Figure 11 shows the visual comparison of dehazing results on real outdoor foggy images between classical image dehazing algorithms and MFID-Net. It can be clearly seen that the results of the DCP algorithm are overly saturated in color overall, which easily leads to information loss in the image. AOD's performance is not satisfactory when dealing with dehazing tasks in distant areas. As for GCA-Net, GridDehazeNet, and FFA-Net, their dehazing effects are not pronounced, only capable of removing surface fog, with FFA-Net even possibly leading to loss of image detail information. In contrast, the algorithm proposed in this paper, when dealing with these real foggy images, can not only effectively remove most of the fog but can also maintain a good preservation of image details and color authenticity.



Figure 11. Real hazy image results. (a) Hazy, (b) DCP, (c) AOD-Net, (d) GCA-Net, (e) GridDehazeNet, (f) FFA-Net, (g) MFID-Net.

Finally, Table 3 presents a comparison between LGCA and other attention mechanisms in terms of the number of parameters and computational cost. Through this comparison, it can be demonstrated that LGCA possesses characteristics of being lightweight.

Table 3. The comparison results of LGCA with other attention mechanisms in terms of parameter quantity and computational effort.

Evaluation	SENet	PA	CA	GPA	LGCA	CBAM	SENet + PA	SENet + GPA	LGCA + PA	LGCA + GPA
Params	512	657	776	657	192	610	1169	1169	849	849
FLOPs	1024	1314	1552	1314	384	1220	3250	3506	1442	1442

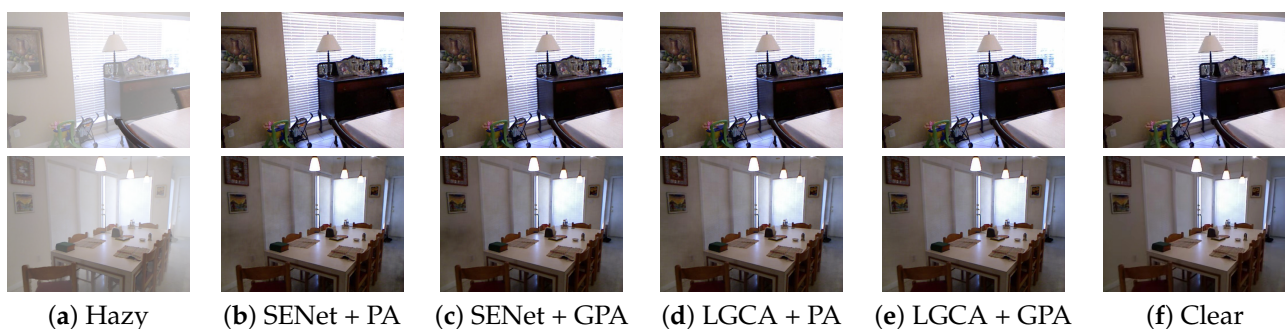
4.4. Ablation Analysis

To further demonstrate the advantages of MFID-Net, a detailed ablation study was conducted in this section, where the various major components of the network were separately analyzed. These components include LGCA and GPA. To ensure the fairness of the experimental results, only one network component was added or modified in each ablation experiment.

To verify the effectiveness of LGCT in image dehazing tasks, we replaced the SE-Net channel attention mechanism used in FFA-Net with LGCT. In Table 4 and Figure 12, the results show that the model using LGCT performs better in image dehazing tasks, not only improving the values of SSIM and PSNR evaluation indicators but also reducing the number of network parameters. In Figure 12, it is evident that LGCT has successfully reduced the phenomena of halo and dark block generation when recovering large-area color blocks in images.

Table 4. Comparisons on SOTS indoor testset for different configurations.

SENet	✓		✓	
LGCA		✓		✓
PA	✓	✓		
GPA			✓	✓
SSIM	36.13	37.21	36.58	37.42
PSNR	0.9842	0.9852	0.9845	0.9890
Parameter(MB)	14.20	14.03	14.20	14.07

**Figure 12.** Comparison chart of results from ablation experiments on different module combinations conducted on IST.

Furthermore, an ablation study was conducted on the global information embedding layer in LGCT, examining the effects under three different scenarios: using ℓ_1 -Norm, ℓ_2 -Norm, and $\ell_1 + \ell_2$ -Norm, respectively. The experimental results indicate that when $\ell_1 + \ell_2$ -Norm is employed to aggregate global contextual information, the model performs best in image dehazing tasks. The related experimental results are shown in Table 5 and Figure 13.

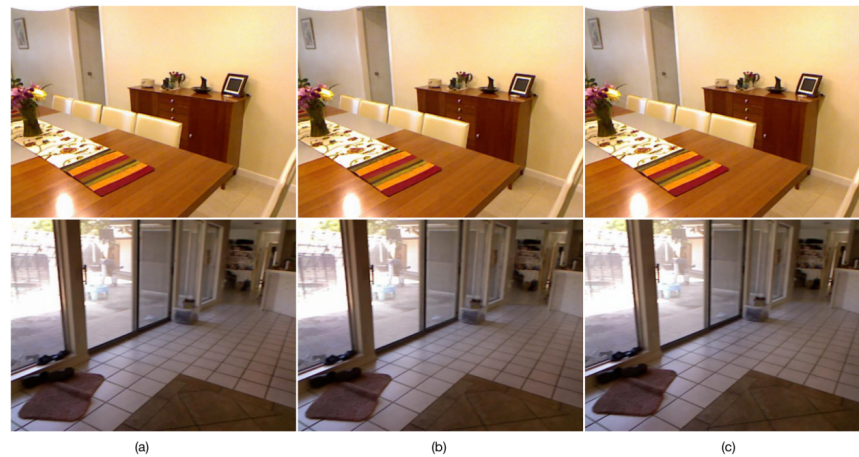


Figure 13. Comparison of dehazing effects from ablation experiments on LGCT conducted on IST. (a) $\ell_1 + \ell_2$ -Norm, (b) ℓ_2 -Norm, (c) ℓ_1 -Norm.

Table 5. Comparisons on SOTS indoor testset for LGCT.

ℓ_1 -norm	✓		
ℓ_2 -norm		✓	
$\ell_1 + \ell_2$ -norm			✓
SSIM	36.95	37.11	37.42
PSNR	0.9862	0.9830	0.9890

These experimental results further substantiate that MFID-Net is capable of achieving an SSIM of 37.42 and a PSNR of 0.9890, significantly higher than other methods. Simultaneously, it proves that each component involved in MFID-Net has played an essential role.

5. Conclusions

To address the limitations of traditional convolutional neural networks (CNNs) and attention mechanisms based on fully connected layers in feature extraction, this paper introduces a novel attention-based multi-feature fusion dehazing algorithm named MFID-Net. The core of the MFID-Net algorithm lies in proposing an innovative multi-feature fusion module (FM), specifically designed to handle areas in images with high haze concentration and significantly enhance the restoration capability for large areas of similar color blocks. The FM module integrates a lightweight gated channel attention mechanism (LGCA) and a gated pixel attention mechanism (GPA), enhancing the overall performance of the algorithm. LGCA replaces the traditional fully connected layers by combining L1- and L2-norms, not only effectively capturing global contextual information but also substantially reducing the demand for computational resources. Meanwhile, GPA optimizes the flow of feature information, reinforcing crucial information while suppressing irrelevant information, further improving the dehazing effects. This innovative FM module is not only structurally concise and highly adaptable but also easily integrated into various image dehazing frameworks. Validation through comparative experiments and ablation studies shows the significant performance improvements of MFID-Net: on the RESIDE indoor dataset, the SSIM reached 37.42 and the PSNR increased to 0.9890; on the RESIDE outdoor dataset, the SSIM reached 34.21 and the PSNR improved to 0.9844.

Despite the superior performance of MFID-Net in comparison to classical dehazing methods as evaluated by SSIM and PSNR, several limitations require careful consideration. Notably, the stability of the algorithm necessitates further enhancement, the training speed warrants acceleration, and the optimization of dehazing performance on real images continues to be a primary area of focus. Consequently, our forthcoming research endeavors

will be dedicated to a comprehensive optimization of this model, aiming to expand the breadth of research within the domain of image dehazing tasks.

Author Contributions: Conceptualization, S.W. (Shengchun Wang); methodology, S.W. (Shengchun Wang) and H.Z.; software, S.W. (Sihong Wang); validation, S.W. (Sihong Wang); formal analysis, S.W. (Shengchun Wang); investigation, S.W. (Shengchun Wang) and S.W. (Sihong Wang); resources, S.W. (Shengchun Wang) and H.Z.; data curation, Y.J.; writing—original draft preparation, S.W. (Shengchun Wang), H.Z. and S.W. (Sihong Wang); writing—review and editing, S.W. (Shengchun Wang) and S.W. (Sihong Wang); visualization, S.W. (Sihong Wang); supervision, S.W. (Shengchun Wang) and H.Z.; project administration, S.W. (Shengchun Wang) and H.Z.; funding acquisition, S.W. (Shengchun Wang), Y.J. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Foundation of Hunan Province (Grant No. 2021JC0009), the Key open laboratory of high-impact weather of China Meteorological Administration, the Science and Technology on Near-Surface Detection Laboratory (Grant No. 6142414221302), the Natural Science Foundation of Jiangsu Province (Grant No. BK20220226), and the Major Program of Xiangjiang Laboratory (Grant No. 23XJ01005).

Data Availability Statement: Data available on request due to restrictions (privacy and ethical).

Conflicts of Interest: Author Yue Jiang was employed by the company Hunan Civil Aviation Administration Air Traffic Management Branch. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. McCartney, E.J. *Optics of the Atmosphere: Scattering by Molecules and Particles*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 1976; Volume 1, p. 421.
2. Nayar, S.K.; Narasimhan, S.G. Vision in bad weather. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 820–827.
3. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. End-to-end united video dehazing and detection. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 1, p. 32.
4. Li, Y.; You, S.; Brown, M.S.; Tan, R.T. Haze visibility enhancement: A survey and quantitative benchmarking. *Comput. Vis. Image Underst.* **2017**, *165*, 1–16. [\[CrossRef\]](#)
5. Zhao, D.; Xu, L.; Yan, Y.; Chen, J.; Duan, L.-Y. Multi-scale optimal fusion model for single image dehazing. *Signal Process. Image Commun.* **2019**, *74*, 253–265. [\[CrossRef\]](#)
6. Sakaridis, C.; Dai, D.; Van Gool, L. Semantic foggy scene understanding with synthetic data. *Int. J. Comput. Vis.* **2018**, *126*, 973–992. [\[CrossRef\]](#)
7. Hua, Z.; Ding, Y.; Li, J. Image dehazing using near-infrared information based on dark channel prior. *Procedia Comput. Sci.* **2021**, *187*, 18–23. [\[CrossRef\]](#)
8. Bi, P.; Wang, D.; Chen, W.; Yang, L.; Liang, J.; Li, G.; Zhang, F.; Wang, Z.; Zhang, X. Image dehazing based on polarization information and deep prior learning. *Optik* **2022**, *267*, 169–746. [\[CrossRef\]](#)
9. Yin, S.; Wang, Y.; Yang, Y.-H. A novel image-dehazing network with a parallel attention block. *Pattern Recognit.* **2020**, *102*, 107255. [\[CrossRef\]](#)
10. Zhang, X.; Wang, T.; Wang, J.; Tang, G.; Zhao, L. Pyramid channel-based feature attention network for image dehazing. *Comput. Vis. Image Underst.* **2020**, *197*, 103003. [\[CrossRef\]](#)
11. Guo, F.; Yang, J.; Liu, Z.; Tang, J. Haze removal for single image: A comprehensive review. *Neurocomputing* **2023**, *537*, 85–109. [\[CrossRef\]](#)
12. Guo, X.; Yang, Y.; Wang, C.; Ma, J. Image dehazing via enhancement, restoration, and fusion: A survey. *Inf. Fusion* **2022**, *86*, 146–170. [\[CrossRef\]](#)
13. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. Aod-net: All-in-one dehazing network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4770–4778.
14. Schechner, Y.; Narasimhan, S.; Nayar, S. Instant dehazing of images using polarization. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, HI, USA, 8–14 December 2001; pp. 325–332.
15. Tarel, J.-P.; Hautiere, N.; Caraffa, L.; Cord, A.; Halmaoui, H.; Gruyer, D. Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intell. Transp. Syst. Mag.* **2012**, *4*, 6–20. [\[CrossRef\]](#)
16. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [\[PubMed\]](#)
17. Zhu, Q.; Mai, J.; Shao, L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **2015**, *24*, 3522–3533. [\[PubMed\]](#)

18. 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. [\[CrossRef\]](#)
19. Zhang, H.; Patel, V.M. Densely connected pyramid dehazing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–28 June 2018; pp. 194–3203.
20. Yang, X.; Xu, Z.; Luo, J. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
21. Engin, D.; Genc, A.; Kemal Ekenel, H. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–28 June 2018; pp. 825–833.
22. Ren, W.; Ma, L.; Zhang, J.; Pan, J.; Cao, X.; Liu, W.; Yang, M.-H. Gated fusion network for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–28 June 2018; pp. 3253–3261.
23. Das, S.; Islam, M.; Amin, M. GCA-Net: Utilizing gated context attention for improving image forgery localization and detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 81–90.
24. Tran, L.-A.; Moon, S.; Park, D.-C. A novel encoder-decoder network with guided transmission map for single image dehazing. *Procedia Comput. Sci.* **2022**, *204*, 682–689. [\[CrossRef\]](#)
25. Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; Ma, L. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10551–10560.
26. Li Z.; Liu F.; Yang W.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [\[CrossRef\]](#) [\[PubMed\]](#)
27. 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, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–28 June 2018; pp. 7132–7141.
29. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [\[CrossRef\]](#)
30. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Ju, M.; Zhang, D.; Wang, X. Single image dehazing via an improved atmospheric scattering model. *Vis. Comput.* **2017**, *33*, 1613–1625. [\[CrossRef\]](#)
32. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 14–19 June 2020; pp. 11534–11542.
33. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
34. Shu, Y.; Yu, B.; Xu, H.; Liu, L. Improving fine-grained visual recognition in low data regimes via self-boosting attention mechanism. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 449–465.
35. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.C.; Salakhutdinov, R.; Zemel, R.S.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2048–2057.
36. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. Griddehazenet: Attention-based multi-scale network for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7314–7323.
37. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. Maxim: Multi-axis mlp for image processing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5769–5780.
38. Hinton, G.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors. *Comput. Sci.* **2012**, *3*, 212–223.
39. Srivastava, N.; Hinton, G.E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
40. Torralba, A. Contextual priming for object detection. *Int. J. Comput. Vis.* **2003**, *53*, 169–191. [\[CrossRef\]](#)
41. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
42. Wu, Y.; He, K. Group normalization. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
43. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long shortterm memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [\[CrossRef\]](#)
44. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2017**, *3*, 47–57. [\[CrossRef\]](#)
45. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor segmentation and support inference from rgb-d images. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 746–760.

46. Ancuti, C.; Ancuti, C.O.; Vleeschouwer, C. De D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2226–2230.
47. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2019**, *28*, 492–505. [[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.