

Article

Multi-Dimensional and Multi-Scale Physical Dehazing Network for Remote Sensing Images

Hao Zhou ^{1,2} , Le Wang ¹ , Qiao Li ¹, Xin Guan ³ and Tao Tao ^{1,4,*} 

- ¹ School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243002, China; haozhou@ahut.edu.cn (H.Z.); wangle@ahut.edu.cn (L.W.); qiaoli@ahut.edu.cn (Q.L.)
- ² School of Big Data and Intelligent Engineering, Chongqing College of International Business and Economics, Chongqing 401520, China
- ³ College of Computer and Information Science, Southwest University, Chongqing 400715, China; guanxin2020@email.swu.edu.cn
- ⁴ College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
- * Correspondence: taotao@ahut.edu.cn

Abstract: Haze obscures remote sensing images, making it difficult to extract valuable information. To address this problem, we propose a fine detail extraction network that aims to restore image details and improve image quality. Specifically, to capture fine details, we design multi-scale and multi-dimensional extraction blocks and then fuse them to optimize feature extraction. The multi-scale extraction block adopts multi-scale pixel attention and channel attention to extract and combine global and local information from the image. Meanwhile, the multi-dimensional extraction block uses depthwise separable convolutional layers to capture additional dimensional information. Additionally, we integrate an atmospheric scattering model unit into the network to enhance both the dehazing effectiveness and stability. Our experiments on the SateHaze1k and HRSD datasets demonstrate that the proposed method efficiently handles remote sensing images with varying levels of haze, successfully recovers fine details, and achieves superior results compared to existing state-of-the-art dehazing techniques.



Citation: Zhou, H.; Wang, L.; Li, Q.; Guan, X.; Tao, T. Multi-Dimensional and Multi-Scale Physical Dehazing Network for Remote Sensing Images. *Remote Sens.* **2024**, *16*, 4780. <https://doi.org/10.3390/rs16244780>

Academic Editors: Sidike Paheding and Ashraf Saleem

Received: 23 October 2024

Revised: 18 December 2024

Accepted: 20 December 2024

Published: 22 December 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: remote sensing image; haze removal; image dehazing; computer vision

1. Introduction

Remote sensing plays a pivotal role in observing and analyzing Earth's surface, providing critical data for various applications such as environmental monitoring [1], agriculture [2], urban planning [3], and disaster management [4]. By utilizing advanced imaging technologies, remote sensing allows us to gather information over vast areas, facilitating real-time assessments and informed decision-making. However, the quality of remote sensing images is often compromised due to atmospheric conditions, particularly haze, which significantly degrades the visual clarity and accuracy of the captured data. Haze is primarily caused by the scattering of light due to atmospheric particles such as dust, smoke, and water droplets. This scattering effect leads to reduced contrast, color distortion, and a loss of detail in the images, making it challenging to extract meaningful information.

To address this issue, image dehazing techniques are categorized into two main types: prior-based and learning-based methods. Prior-based dehazing methods use predefined assumptions to estimate the transmission map and atmospheric light and then apply the atmospheric scattering model (ASM) [5,6] to restore a clear image. Prior-based methods, which rely on this physical model, have proven to be effective in many cases. However, they often struggle when applied to more complex scenes with heterogeneous haze distributions or non-uniform lighting conditions, as the assumptions behind these models may not hold

in real-world environments. These methods are also limited in their ability to generalize across different types of haze and atmospheric conditions.

In contrast, learning-based methods, particularly those utilizing deep learning techniques [7–12], have demonstrated remarkable adaptability and performance improvements over prior-based approaches. These methods leverage large-scale datasets to learn the intricate relationships between hazy and clear images without relying on explicit physical models. This allows them to perform well even in challenging and varied haze conditions, as they can capture more nuanced features and patterns from the data. Early deep learning-based dehazing approaches [13,14] were built around the atmospheric scattering model but struggled with real-world haze scenes. The primary limitation was that the physical scattering model does not always accurately reflect the complexity of actual atmospheric conditions, leading to suboptimal results in practical applications.

To overcome these shortcomings, more recent methods have shifted toward end-to-end learning frameworks, which aim to bypass the constraints of the physical model altogether. Examples of such approaches include multi-scale convolutional neural networks (CNNs) [15,16], which directly map hazy images to their dehazed counterparts. These methods have been successful largely due to their ability to automatically extract powerful, discriminative features through multiple layers of convolution operations. However, one of the inherent limitations to convolution is that it is a local operation. This means that, while convolutional layers are excellent at capturing fine, localized details, they struggle to model long-range dependencies and global structural information, which are often critical for accurately restoring hazy images.

To address this limitation, researchers have introduced multi-scale pooling modules and attention mechanisms to incorporate global context into the dehazing process. FFA-Net [17] further improved on this by introducing channel attention mechanisms to simulate non-local relationships between different regions of the image, effectively injecting global information into the dehazing process. This attention mechanism allows the model to prioritize important features and regions of the image, thereby enhancing overall dehazing performance.

More recently, advanced approaches such as transformer-based architectures [18–20], Mamba frameworks [21], and diffusion models [22,23] have gained prominence in image dehazing. Transformers excel at capturing long-range dependencies and global structures through self-attention mechanisms, enabling accurate recovery in challenging haze conditions. Mamba-based methods emphasize efficient design, balancing performance and computational complexity for practical applications. Diffusion models, leveraging iterative denoising and probabilistic generative modeling, effectively handle haze-induced uncertainty and restore fine-grained details, achieving high fidelity even under dense haze scenarios. These methods collectively push the boundaries of image restoration performance.

Despite the success of learning-based methods, many researchers recognize that physical priors still play a crucial role in the dehazing process. By incorporating knowledge from the atmospheric scattering model or other physical phenomena, dehazing methods can be made more robust and interpretable. Hybrid approaches, which combine the strengths of both learning-based techniques and physical priors, have emerged as a promising direction in the field. One such example is SwinTD-Net [24], which integrates physical priors with transformer-based architectures to address common issues in dehazed images, such as reduced contrast and color distortion. SwinTD-Net utilizes physical priors related to haze formation to improve the model's capacity for accurately restoring the scene, especially in terms of maintaining color accuracy and preserving details, resulting in more visually precise dehazing outcomes.

For accurate image dehazing, both local context features and global structural features are needed. As mentioned earlier, convolution operations capture local context features within a local region to preserve fine details, while multi-scale modules can effectively capture long-range dependencies to model global features. To extract more comprehensive features from hazy images, it is necessary to explore useful features across different dimensions by increasing dimensionality. However, deep learning methods alone cannot fully compensate for the loss of image information caused by atmospheric particles.

In this paper, we introduce a dehazing network for remote sensing images, called the Multi-Dimensional and Multi-Scale Physical Dehazing Network (MMPD-Net). Our proposed network is built upon a U-net architecture [25] and features a detailed multi-scale, multi-dimensional extraction block (MMB). This block integrates both a multi-scale extraction block (MSB) and a multi-dimensional extraction block (MDB) to enhance the extraction of fine details. MSB is inspired by the pixel attention (PA) and channel attention (CA) proposed in FFA-Net [17]. Before the PA block, we use dilated convolutions with dilation rates of 1, 2, 4, and 8 to extract multi-scale detail features. MDB uses depthwise separable convolutional layers to extract features with more dimensions. In addition, we design an atmospheric scattering model-based dehazing unit (ASMU), which extracts A through fully connected layers and t through convolutional layers and finally obtains the restored image, J . This architecture helps extract fine details between hazy and haze-free images, and it can effectively remove haze. Finally, we optimized SK-Fusion and proposed MSK-Fusion.

The main contributions of this paper are as follows:

- A novel multi-dimensional and multi-scale detail extraction module is proposed to mine hidden, multi-scale, and detailed features. Parallel dilated convolutions enhance the receptive field, enabling better distant feature modeling and wider context gathering, thereby improving image detail capture.
- A physical module based on the atmospheric scattering model adaptively extracts A and t , handling uneven haze distribution and improving remote sensing image dehazing.
- The MSK-Fusion module integrates global and salient features, dynamically adjusting their weights with learnable parameters to enhance the network's generalization and robustness.
- A new dehazing network, integrating ASMU and MSK-Fusion into MMPD-Net, is designed for remote sensing images. Tested on the challenging SateHaze1k [26] and HRSD [27] datasets, it outperforms current state-of-the-art methods, providing more efficient haze removal.

2. Relate Works

2.1. Prior-Based Dehazing Methods

Image dehazing techniques based on priors estimate the transmission map and atmospheric light by exploiting the characteristics of clean images, relying extensively on atmospheric scattering models and carefully designed priors. These approaches are grounded in certain assumptions about how haze affects image formation and attempt to reverse these effects using various priors. One of the earliest methods, proposed by Tan et al. [28], was based on the observation that haze-free images generally have higher contrast than hazy ones. By maximizing local contrast, Tan's method aimed to enhance visibility and recover haze-free images. Following this, He et al. [29] introduced the widely used dark channel prior (DCP), which is based on the observation that, in most outdoor haze-free images, there are often low-intensity pixels in at least one color channel. This insight allowed for the estimation of the transmission map, leading to effective haze removal. Fattal [30] improved on previous work by introducing a refined image formation model that accounted for surface shading and scene transmission, providing a more accurate dehazing method by separating the effects of light scattering and scene structure. Tang et al. [31], on the other hand, combined various haze-related priors such as contrast, saturation, and edges, and

employed a random forest regressor to automatically predict the transmission map, thus leveraging machine learning to improve the prior-based approach. Zhu et al. [32] presented the color attenuation prior (CAP), which estimates scene depth and recovers hazy images by utilizing the contrast between pixel brightness and saturation. However, this method encounters difficulties when addressing varying atmospheric conditions. Berman et al. [33] observed that haze tends to form linear clusters of pixels, referred to as “haze lines”, and developed a non-local prior based on this phenomenon to better model the relationship between clean and hazy images.

Despite the strengths of these methods, such as simplicity and the ability to incorporate domain-specific knowledge, they rely on idealized assumptions about haze distribution and image properties. This limits their adaptability in complex scenes with diverse haze conditions, where these assumptions often break down. Furthermore, these methods are typically local in nature and fail to capture global contextual information, leading to reduced robustness when dealing with varying haze densities or heterogeneous environments. Consequently, their performance can degrade in scenarios with non-uniform haze or intricate scene structures, where the simplifying assumptions of the priors are less applicable.

2.2. Learning-Based Dehazing Methods

Learning-based approaches to image dehazing predominantly rely on deep neural networks to either estimate transmission maps and atmospheric light, based on physical scattering models, or directly transform hazy images into clear ones without an explicit reliance on such models. The former approach leverages the atmospheric scattering model to guide the learning process, while the latter learns end-to-end mapping from hazy inputs to dehazed outputs, bypassing traditional physical models. Cai et al. [13] were pioneers in applying convolutional neural networks (CNNs) to this problem with the introduction of DehazeNet, an end-to-end network that inputs hazy images and predicts transmission maps. This approach utilized multi-scale convolutional layers to capture haze-specific features at different resolutions, allowing for the better handling of varying haze densities. Building on this, Li et al. [34] proposed AOD-Net, which advanced the field by introducing a more efficient approach that combined atmospheric light and transmittance into a single parameter. This fusion improved both computational efficiency and restoration quality, as it reduced the reliance on separate estimations for different components of the haze model. Liu et al. [16] took a different route with GridDehaze-Net, incorporating attention mechanisms that allowed the model to focus on the most crucial areas of an image affected by haze. This approach generated diverse feature representations and selectively enhanced the more relevant parts of the image, leading to improved dehazing results. Attention-based networks continued to play a significant role in subsequent advancements. Qin et al. [17] introduced FFA-Net, which used a feature attention module to amplify relevant features and suppress irrelevant ones, further enhancing the visibility of dehazed images. Mei et al. [35] employed a U-Net architecture, which progressively fused features at multiple stages of the network, allowing for a more detailed reconstruction of haze-free images. This U-Net structure, with its skip connections, helped retain spatial details and achieve more accurate dehazing. Recent approaches have explored more sophisticated network designs. Lu et al. [36] introduced MixDehazeNet, a network that employs a mix structure block to effectively capture both local and global features. By combining these feature scales, the network was able to achieve better performance in challenging scenarios with complex haze distributions. Similarly, Cui et al. [37] proposed an omni-kernel network, which adapts flexibly to various types of image degradation, making it suitable for a wide range of restoration tasks beyond dehazing. This network’s unified approach allowed it to generalize across different degradation patterns, further enhancing its versatility. Transformer-based models have also gained popularity for dehazing. Song et al. [19] developed DehazeFormer, a model based on the Swin Transformer architecture, which improved dehazing by capturing long-range dependencies and modeling global context

more effectively than CNNs. In a related approach, Sun et al. [38] tackled color distortion and texture inconsistencies in dehazed images by utilizing a partially shared multi-scale dual encoding–decoding network. This method ensured the preservation of both fine local details and the broader global structures. Song et al. [39] further refined transformer-based architectures with RSDFormer, which captures both local and global dependencies, improving the network’s ability to recover fine details and enhance image clarity under various haze conditions. He et al. [40] introduced a lightweight convolutional neural network based on axial depthwise convolution and hybrid attention, aimed to enhance dehazing performance for remote sensing images while reducing model complexity.

While these learning-based methods have demonstrated impressive results in terms of dehazing performance, they come with certain limitations. Unlike prior-based approaches, which rely on explicit physical models, these methods are often criticized for their lack of physical interpretability. The neural networks in these models learn mappings that may not always align with the underlying physics of haze formation and removal, leading to inconsistencies when applied to different types of haze. Furthermore, their performance can be unpredictable in diverse or unstructured environments, where the learned patterns may not generalize well. These challenges highlight the need for more robust and interpretable solutions that can adapt to various haze conditions while maintaining strong performance.

2.3. Combining Learning-Based and Prior-Based Dehazing Methods

By combining learning-based methods with knowledge from atmospheric scattering models or other physical priors, dehazing methods can become more robust and interpretable. Zhang et al. [41] introduced an end-to-end deep learning framework that combines a physical model-based dark channel prior (DCP) with a densely connected pyramid architecture to estimate transmission maps and atmospheric light more accurately, achieving high-quality image dehazing. Liu et al. [42] introduced a method that combines traditional haze-removal priors, such as a dark-channel prior and atmospheric light, with deep convolutional neural networks (CNNs), enabling the iterative refinement of these priors for more accurate image dehazing. Zhao et al. [43] introduced a weakly supervised learning framework called RefinedNet, which refines haze-removal results by incorporating both physical priors and weakly labeled data to enhance the dehazing performance without requiring fully annotated training data. Zheng et al. [44] proposed a physics-aware dehazing framework that incorporates curricular contrastive regularization to improve the model’s robustness and generalization, effectively learning better representations of hazy images while respecting physical principles.

3. Method

Figure 1 illustrates our dehazing network architecture. The overall structure of MMPD-Net is a 5-stage U-net, with each stage consisting of our designed MMB and ASMU modules. The MMB consists of MSB and MDB, which are employed to dynamically capture multi-scale high-frequency components, edge information, and various dimensional details. ASMU is designed based on Equation (5), using convolutional blocks to extract the A and t components from the equation. Additionally, previous research [19] introduced SK Fusion, which achieves superior feature fusion capabilities. We incorporated and appropriately modified SK Fusion to develop MSK Fusion, which merges the skip connections and main branch for image reconstruction, thereby enhancing the dehazing performance of MMPD-Net.

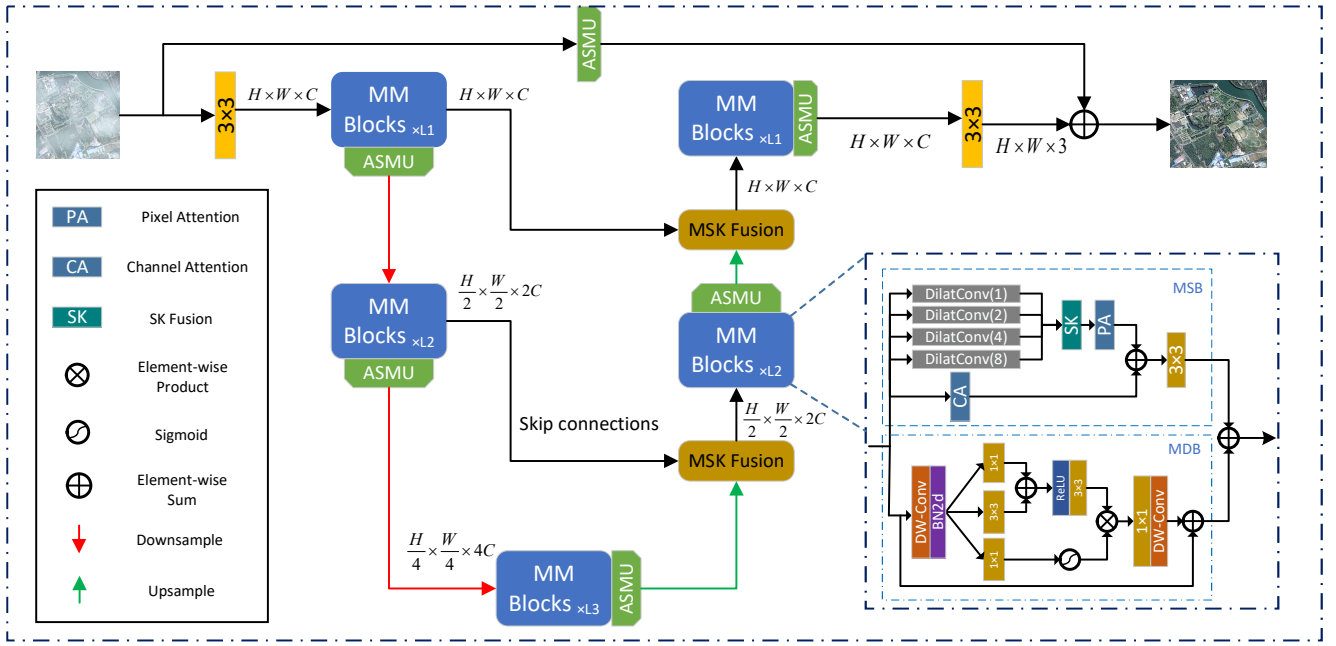


Figure 1. Multi-Dimensional and Multi-Scale Physical Dehazing Network.

3.1. Multi-Dimensional and Multi-Scale Block

Our MMB aims to extract multi-scale and multi-dimensional features. MSB extracts global and local features, and MDB discovers hidden features of other dimensions. In essence, MMB is an additive fusion of MSB and MDB, as shown in Figure 1

$$y = \text{MSB}(\hat{x}) + \text{MDB}(\hat{x}). \quad (1)$$

where $\hat{x}$ refers to the input of the previous moment, and $\text{MSB}(\cdot)$ and $\text{MDB}(\cdot)$ respectively represent the multi-scale and multi-dimensional processing of the input of the previous moment.

In the multi-scale block (MSB), we first apply four parallel dilated convolutions (Dilat-Conv) with different dilation rates based on the HDN proposed by Chalavadi et al. [45]. This allows for the extraction of features at multiple scales, effectively capturing both fine details and broader contextual information. To further enhance the feature representation, we use the selective kernel (SK) module to fuse these multi-scale features, which helps in retaining the important information from each scale. The combination of these operations ensures that the local features and multi-scale contextual features of the feature map are well preserved. After the feature map is processed in this manner, the learned features are passed through a pixel attention (PA) mechanism, as proposed in FFA-Net [17]. This step is crucial for extracting key scale-specific details by emphasizing the most relevant pixel-level information. The pixel attention mechanism focuses on identifying and enhancing significant areas of the image that are crucial for tasks like image dehazing, where localized clarity is important.

In addition to pixel attention, we introduce a channel attention (CA) mechanism that runs in parallel. While pixel attention is focused on specific spatial regions, channel attention concentrates on global information and adaptively highlights haze-related features across the image. This dual attention approach combining pixel and channel attention ensures that the network captures both localized fine details and broader, scene-wide features. The features extracted via the two attention mechanisms are then fused by summing them, creating a more comprehensive and rich feature set. To further refine the fused features, a final 3×3 convolution is applied. This step helps smooth and refine the features, making them more useful for downstream tasks like image restoration or dehazing. The use of different dilation rates in the dilated convolutions is also significant. Small dilated

convolutions are particularly effective at recovering detailed texture information, while larger dilations provide a wider receptive field, allowing the model to capture long-distance dependencies and broader structural information in the image.

$$\begin{aligned} x_{out} &= \text{PA}(\text{SK}(\text{DilatConv}_i(\hat{x}))), \\ \text{MSB}_{out} &= \text{Conv3}(\text{CA}(\hat{x}) + x_{out}). \end{aligned} \quad (2)$$

In the multi-dimensional block (MDB) architecture, we start by applying depthwise convolution (DWConv) to maintain feature representation while ensuring computational efficiency. This is followed by normalization to stabilize the feature distribution, improving convergence during training. Next, we utilize a combination of 1×1 convolution and 3×3 convolution layers to expand the number of channels, enhancing the richness of the feature information. 1×1 convolution is a convolution operation where the kernel size is 1×1 . Unlike larger kernels, a 1×1 kernel does not capture spatial information directly because it only operates on a single pixel at a time. Instead, it is primarily used for channel-wise transformations and dimensionality reduction or expansion. The 1×1 convolution increases the feature map's depth, while the 3×3 convolution captures local patterns. The normalized features extracted from these convolutions are then fused, allowing for a more comprehensive representation of the input data. After fusion, we apply another 3×3 convolution to refine the extracted features further. Following the approach proposed by Ma et al. [46], we multiply the refined features with another branch processed through a Sigmoid function, effectively rescaling the feature values and enhancing the dimensional representation.

We then reduce the dimensionality with a 1×1 convolution to condense the useful information and prevent overfitting. Finally, we pass the reduced feature map through another DWConv layer and incorporate a residual connection to preserve the original feature information. This results in a final output that effectively combines enhanced and preserved features, ready for subsequent processing tasks.

$$\begin{aligned} x_0 &= \text{BN}(\text{DWConv}(\hat{x})), \\ x_1 &= \text{Conv3}(\text{Relu}(\text{Conv1}(x_0) + \text{Conv3}(x_0))), \\ x_2 &= \sigma(\text{Conv1}(x_0)), \\ \text{MDB}_{out} &= \hat{x} + \text{DWConv}(\text{Conv1}(x_1 * x_2)). \end{aligned} \quad (3)$$

In the above formula, $\sigma(\cdot)$ is the Sigmoid function, $*$ represents the element-wise product, and the Sigmoid function is defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

3.2. Atmospheric Scattering Model Unit

The formula for the atmospheric scattering model (ASM) is as follows:

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (5)$$

where I is the observed hazy image, represented as a 2D matrix, with x is the 2D spatial location in an image, J is the haze-free scene radiance, also represented as a 2D matrix, t is the transmission map, a 2D matrix that indicates the proportion of light that reaches the camera after passing through the atmosphere, with values in the range $[0, 1]$, and A is the global atmospheric light, representing the haze's color component, which can be either a scalar (grayscale images) or a 3D vector (RGB images), applied consistently across the image.

Inspired by PDU [44], we propose an atmospheric scattering model unit (ASMU), which is derived from the physical model of Equation (5), as shown in Figure 2. First, we represent the clear image, J , according to the physical model as follows:

$$\begin{aligned}\tilde{t}(x) &= \frac{1}{t(x)} \\ J(x) &= I(x) \frac{1}{t(x)} + A(1 - \frac{1}{t(x)}) \\ &= I(x)\tilde{t}(x) + \tilde{A}(1 - \tilde{t}(x)).\end{aligned}\quad (6)$$

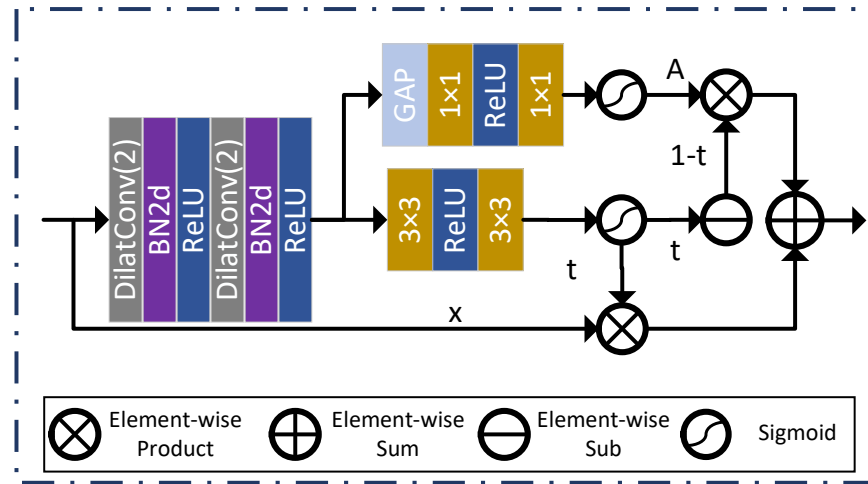


Figure 2. Structure of ASMU.

In order to better extract the feature information of A and t , we introduce two DilatConvBlocks before extracting features. Each DilatConvBlock consists of a DilatConv layer, a BatchNormal layer, and a Relu layer. Dilated convolution can avoid information loss while expanding the receptive field, especially when processing details such as image edges and textures, and it can better retain these key features. In the dehazing process, it can capture contextual information in a wider range, which helps with better understanding the haze distribution and global structure in the image.

$$\tilde{x} = \text{Relu}(\text{BN}(\text{DilatConv}(x))) * 2. \quad (7)$$

Given that atmospheric light is typically assumed to be uniform, we employ global average pooling (GAP) to remove redundant information in the feature space. Global average pooling (GAP) is used to calculate the average value of the feature map in its spatial dimension. After global average pooling, the feature map becomes a one-dimensional feature vector. The feature vector conforms to the characteristics of the atmospheric light value. Then, it undergoes 1×1 convolution to perform feature transformation, and finally, it performs sigmoid processing to obtain the atmospheric light. $\tilde{A}$ is expressed as follows:

$$\tilde{A} = \sigma(\text{Conv1}(\text{Relu}(\text{Conv1}(\text{GAP}(\tilde{x}))))). \quad (8)$$

However, due to the loss of information, we cannot apply GAP to t since the transmission map is non-homogeneous. Instead, we choose to use a series of 3×3 convolutional layers to extract features, as described below:

$$\tilde{t} = \sigma(\text{Conv3}(\text{Relu}(\text{Conv3}(\tilde{x}))))). \quad (9)$$

In the above formula, $\sigma(\cdot)$ is the sigmoid function, $\text{Conv1}(\cdot)$ denotes 1×1 convolution, $\text{Conv3}(\cdot)$ is 3×3 convolution, and $\text{BN}(\cdot)$ denotes BatchNorm2d. Conv1 refers to a 1×1 convolution, primarily used for channel mixing or dimensionality reduction/expansion without capturing spatial relationships. Conv generally denotes a standard convolution operation with an unspecified kernel size, often 3×3 , used for general feature extraction. Conv3 specifically indicates a 3×3 convolution, which balances computational cost and spatial feature capture, making it common for extracting local patterns such as edges or textures.

3.3. MSK Fusion

The SK fusion layer, based on the selective kernel (SK) module [47], enhances image feature selection by adapting receptive fields. Similar concepts are found in MIRNet [48] and DehazeFormer. To improve the original SK fusion, which only uses global average pooling ($\text{GAP}(\cdot)$) and may miss key features, the MSK fusion layer introduces global max pooling ($\text{GMP}(\cdot)$) in parallel with $\text{GAP}(\cdot)$ to better retain important details like edges and textures. The structure of MSK is shown in Figure 3.

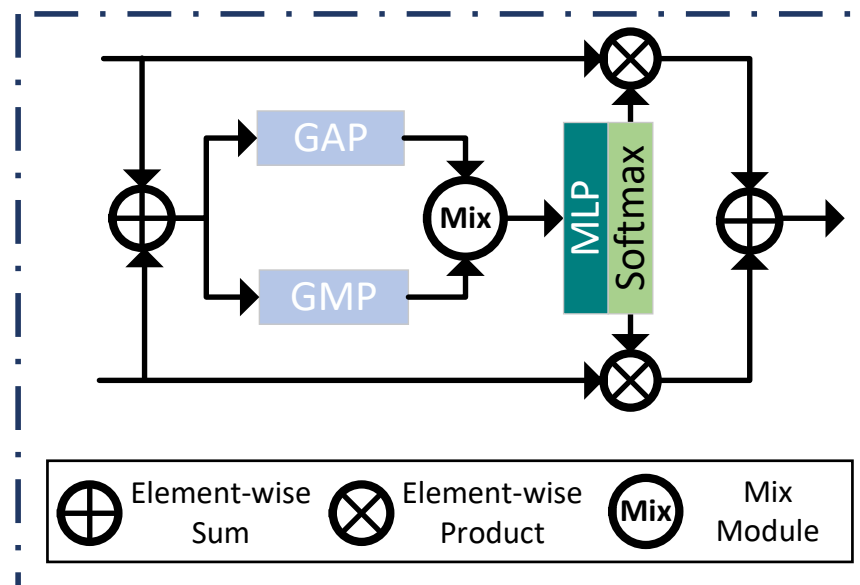


Figure 3. Structure of MSK.

To balance the contributions of $\text{GAP}(\cdot)$ and $\text{GMP}(\cdot)$, a mix module is introduced with a learnable parameter, m . This parameter is converted into a weight, λ , using a Sigmoid function. This weight dynamically adjusts during training, allowing the model to adaptively prioritize global context or salient details when needed, as follows:

$$\tilde{x} = \lambda \cdot \text{GAP}(x) + (1 - \lambda) \cdot \text{GMP}(x). \quad (10)$$

The combination of these pooling methods ensures better feature fusion, improving performance in tasks like image dehazing by capturing both global and local features effectively. In the setting of the initial parameter m , m is set to 0.4 when the second layer downsampling and the first layer upsampling are fused, and m is set to 0.8 when the first layer downsampling and the second layer upsampling are fused.

3.4. Loss Function

The mean square error, or $L2$ loss, is commonly used as the primary loss function in image dehazing tasks. However, many image recovery tasks show improved performance in PSNR and SSIM metrics when trained with an $L1$ loss instead of an $L2$ loss. Following this trend, we chose to use the simple $L1$ loss as our default option. Although perceptual loss is frequently employed in various dehazing algorithms, we prioritize optimizing the $L1$ loss.

$$L1 = \|J - GT\|_1 \quad (11)$$

where J represents the dehazed remote sensing image produced via MMPD-Net, while GT refers to the corresponding ground-truth remote sensing image.

4. Experiment

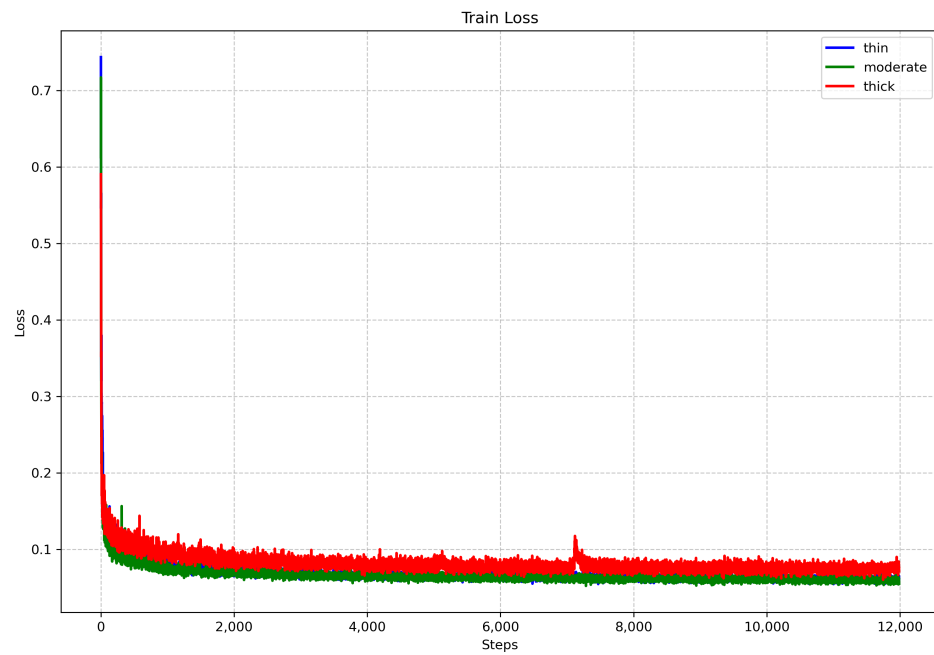
4.1. Datasets

We evaluate the proposed MMPD-Net using two synthetic remote sensing (RS) haze datasets: SateHaze1k [26] and HRSD [27]. SateHaze1k includes three subsets: thin, moderate, and thick haze, with training sets of 320 images and test sets of 45 images each. Thin-haze images utilize haze masks derived from real clouds, while moderate samples combine characteristics of mist and medium haze. Thick-haze images are generated using transmittance maps from dense haze. The HRSD dataset is divided into two subsets: LHID and DHID. LHID comprises a total of 30,517 training images and 500 test images, all of which are generated using the atmospheric scattering model. This subset simulates various haze conditions to train the network on handling different levels of haze in remote sensing images. On the other hand, DHID is composed of 14,990 images in total, and it was created using real-haze maps, providing a more realistic representation of haze. Of these, 14,490 images are used for training, while 500 are set aside for testing. The combination of these two subsets ensures a comprehensive evaluation of the dehazing method across both synthetic and more authentic haze scenarios. To demonstrate the dehazing performance of our network in real-world scenarios, we tested it using a real haze remote sensing dataset provided in the study [49]. This dataset contains high-quality real haze images, and such tests effectively evaluate the robustness and dehazing ability of our method in real environments.

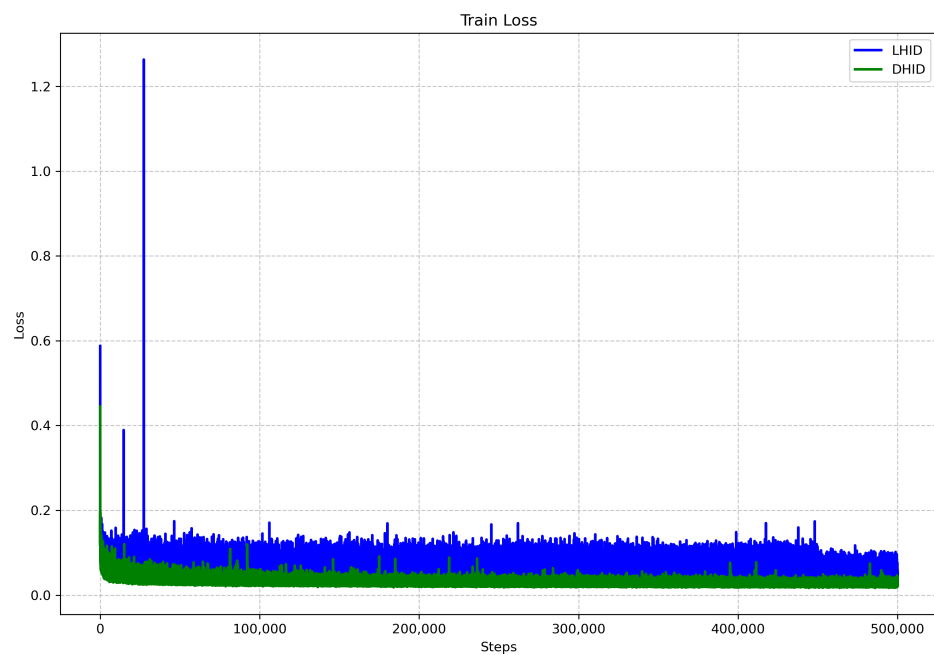
4.2. Implementation Details

We trained and evaluated our MMPD-Net using the PyTorch framework on a system equipped with four NVIDIA GeForce GTX 1080 Ti GPUs. To enhance the training dataset, we performed random rotations of 90, 180, and 270 degrees, along with horizontal flips. The input for MMPD-Net consisted of RGB remote sensing images cropped to a size of 256×256 pixels. For the MMB, we set the configuration of $[N1, N2, N3, N4, N5]$ to $[2, 2, 4, 2, 2]$, with corresponding embedding channels of $[24, 48, 96, 48, 24]$. We employed the Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$), using a batch size of 4 for training on each sub-dataset. The starting learning rate was set to 2.0×10^{-4} , and we employed a cosine annealing strategy to progressively reduce it to 0. The corresponding training loss for each dataset is shown in Figure 4.

We compared MMPD-Net against seven advanced dehazing methods: DCP [29], AOD-Net [34], FCTF-Net [50], GridDehaze-Net [16], FFA-Net [17], MixDehaze-Net [36], VmambaIR [21], FCDM [22], and OK-Net [37]. To ensure an equitable comparison, we utilized the official implementations of these deep learning models, as provided by their respective authors, during the training process.



(a) SateHaze1k datasets loss



(b) HRSD datasets loss

Figure 4. “Thin ” represents the Haze 1k-thin dataset, “moderate” represents the Haze1k-moderate dataset, and “thick” represents the Haze1k-thick dataset.

4.3. Quantitative Evaluations

Tables 1 and 2 present the quantitative evaluation results for the HRSD and SateHaze1k datasets. The metrics used for the performance assessment are PSNR, SSIM, and their average values across varying levels of haze density.

Table 1. SateHaze1k results. For the compared methods, we use **bold** and underline to mark the best and second-best methods.

Methods	Thin Haze		Moderate Haze		Thick Haze		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP [29]	20.15	0.8645	20.51	0.8932	15.77	0.7117	18.81	0.8241
AOD-Net [34]	15.97	0.8169	15.39	0.7442	14.44	0.7013	15.27	0.7541
FCTF-Net [50]	19.13	0.8532	22.32	0.9107	17.78	0.7617	19.74	0.8419
GridDehaze-Net [16]	19.81	0.8556	22.75	0.9085	17.94	0.7551	20.17	0.8397
FFA-Net [17]	<u>24.04</u>	0.9130	<u>25.62</u>	0.9336	<u>21.70</u>	<u>0.8422</u>	<u>23.79</u>	<u>0.8963</u>
MixDehaze-Net [36]	22.12	0.8822	23.92	0.9040	19.96	0.7950	22.00	0.8604
OK-Net [37]	20.68	<u>0.8860</u>	25.39	<u>0.9406</u>	20.21	0.8186	22.09	0.8817
VmambaIR [21]	20.81	0.8753	24.34	0.9132	20.04	0.8045	21.73	0.8643
FCDM [22]	18.94	0.8486	17.36	0.8753	16.97	0.7530	17.76	0.8256
Ours	25.16	0.9227	27.30	0.9454	22.85	0.8571	25.10	0.9084

Table 2. HRSD results. For the compared methods, we use **bold** and underline to mark the best and second-best methods.

Methods	LHID		DHID		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP [29]	21.34	0.7976	19.15	0.8195	20.25	0.8086
AOD-Net [34]	21.91	0.8144	16.03	0.7291	18.97	0.7718
FCTF-Net [50]	28.55	0.8727	22.43	0.8482	25.49	0.8605
GridDehaze-Net [16]	25.80	0.8584	26.77	0.8851	26.29	0.8718
FFA-Net [17]	29.33	0.8755	24.62	0.8657	26.98	0.8706
MixDehaze-Net [36]	<u>29.47</u>	0.8631	27.36	0.8864	<u>28.42</u>	0.8748
OK-Net [37]	29.03	<u>0.8766</u>	<u>27.80</u>	<u>0.8973</u>	<u>28.42</u>	<u>0.8870</u>
FCDM [22]	15.16	0.6459	17.13	0.6978	16.15	0.6719
Ours	29.76	0.8771	28.23	0.8977	29.00	0.8874

In Table 1, our MMPD-Net provides better dehazing results with the light-haze, moderate-haze, and dense-haze datasets. However, the method exhibits a minor decrease in effectiveness when dealing with thick haze, as the severe degradation from heavy haze poses greater challenges for image restoration in remotely sensed data. DCP, FCDM, and AOD-Net exhibit poor performance, whereas FFA-Net and OK-Net demonstrate solid results. Nevertheless, they are still lacking compared with our method. For instance, on the dense-haze dataset, our approach surpasses the second-best FFA-Net by 1.15 in PSNR and 0.0149 in SSIM. These findings underscore the strong performance and competitiveness of MMPD-Net in remote sensing image dehazing.

In Table 2, MMPD-Net delivers the top performance on both synthetic datasets, LHID and DHID. Specifically, compared to the second-best approach, MMPD-Net shows an improvement of 0.29 dB in PSNR and 0.0005 in SSIM on the LHID dataset. On the DHID dataset, it achieves a 0.43 dB increase in PSNR and a 0.0004 improvement in SSIM. OK-Net is the second-best method in most indicators, close to our method in indicators, which is related to the wider receptive field provided via large kernel convolution. These results generally reflect the excellent dehazing ability of MMPD-Net in remote sensing dehazing.

To comprehensively evaluate the quality of restored images, we add MaPSNR, FQeIQA, and LOSSIM for comparison. These additional metrics complement PSNR and SSIM by providing insights into perceptual quality, feature fidelity, and structural accuracy. The results are summarized in Tables 3 and 4, demonstrating the robustness and versatility of our proposed model across multiple evaluation dimensions.

Table 3. SateHaze1k other results. For the compared methods, we use **bold** and underline to mark the best and second-best methods.

Column 1	Thin Haze			Moderate Haze			Thick Haze		
	MaPSNR	FQeIQA	L0SSIM	MaPSNR	FQeIQA	L0SSIM	MaPSNR	FQeIQA	L0SSIM
DCP [29]	28.69	0.9808	0.8795	28.80	0.9762	0.9009	28.24	0.9496	0.7299
AOD-Net [34]	28.25	0.9690	0.8208	27.93	0.9690	0.7650	28.14	0.9552	0.7227
FCTF-Net [50]	28.35	0.9799	0.8673	28.80	0.9856	0.9209	28.16	0.9674	0.7799
GridDehaze-Net [16]	28.54	0.9794	0.8703	29.03	0.9834	0.9199	28.33	0.9649	0.7780
FFA-Net [17]	<u>29.56</u>	<u>0.9908</u>	<u>0.9242</u>	<u>30.03</u>	0.9912	0.9407	<u>28.95</u>	<u>0.9807</u>	<u>0.8597</u>
MixDehaze-Net [36]	28.98	0.9863	0.8986	29.37	0.9860	0.9176	28.50	0.9719	0.8184
OK-Net [37]	28.60	0.9862	0.8994	29.85	<u>0.9913</u>	<u>0.9460</u>	28.65	0.9753	0.8384
VmambaIR [21]	28.71	0.9821	0.8893	29.64	0.9794	0.9195	28.53	0.9722	0.8260
FCDM [22]	27.90	0.8603	0.8543	27.90	0.8684	0.8495	27.90	0.8637	0.7862
Ours	29.96	0.9921	0.9334	30.76	0.9930	0.9521	29.29	0.9823	0.8741

Table 4. HRSD other results. For the compared methods, we use **bold** and underline to mark the best and second-best methods.

Column 1	LHID			DHID		
	MaPSNR	FQeIQA	L0SSIM	MaPSNR	FQeIQA	L0SSIM
DCP [29]	28.58	0.9362	0.8124	28.35	0.9645	0.8341
AOD-Net [34]	28.64	0.9704	0.8263	27.97	0.9382	0.7281
FCTF-Net [50]	31.24	0.9800	0.8845	29.31	0.9686	0.8576
GridDehaze-Net [16]	29.79	0.9785	0.8699	30.33	0.9756	0.8947
FFA-Net [17]	<u>31.90</u>	0.9800	0.8868	30.25	0.9725	0.8750
MixDehaze-Net [36]	<u>31.90</u>	0.9801	0.8764	<u>31.25</u>	<u>0.9785</u>	0.8980
OK-Net [37]	31.57	0.9812	<u>0.8881</u>	30.84	0.9776	<u>0.9046</u>
FCDM [22]	27.88	0.6658	0.7492	20.68	0.8421	0.7042
Ours	32.25	<u>0.9805</u>	0.8884	31.54	0.9790	0.9091

We evaluate the computational efficiency and memory requirements of the proposed model using two key metrics: FLOPs (floating-point operations) and parameters. FLOPs measure the number of floating-point operations required for a single forward pass, providing insight into the computational cost and runtime efficiency. Lower FLOPs indicate a more efficient model suitable for resource-constrained or real-time applications. Parameters quantify the total number of trainable weights in the model, reflecting its memory footprint and storage requirements. A compact model with fewer parameters is advantageous for deployment on devices with limited resources, as shown in Table 5.

Table 5. Comparison of FLOPs and Parameters across models.

	DCP	AOD-Net	FCTF-Net	GridDehaze-Net	FFA-Net	MixDehaze-Net	OK-Net	VmambaIR	FCDM	Ours
FLOPs (G)	-	114.4	10.0	21.4	302.1	28.6	39.6	20.5	47.2	81.2
Parameters (M)	-	0.002	0.163	0.956	4.68	3.17	4.43	10.50	155.36	8.66

4.4. Qualitative Evaluations

In this section, we present a qualitative comparison between MMPD-Net and seven top dehazing techniques, using the SateHaze1k and HRSD datasets as reference benchmarks for remote sensing haze images.

Figure 5 shows the visual results of various methods applied to the thin haze test set. From these figures, it is evident that DCP, AOD-Net, and FCTF-Net leave noticeable residual haze in the dehazed images. While GridDehaze-Net and MixDehaze-Net produce

a certain level of dehazing, their results still retain some haze when compared to the ground-truth images. FFA-Net and OK-Net deliver strong dehazing performance, but the grass and road colors in their restored images appear lighter than those in the real images. The differences between the methods can be more clearly understood from the indicators shown below the images. In general, our method produces dehazed images that more closely match the ground truth, with an improved restoration of fine details.

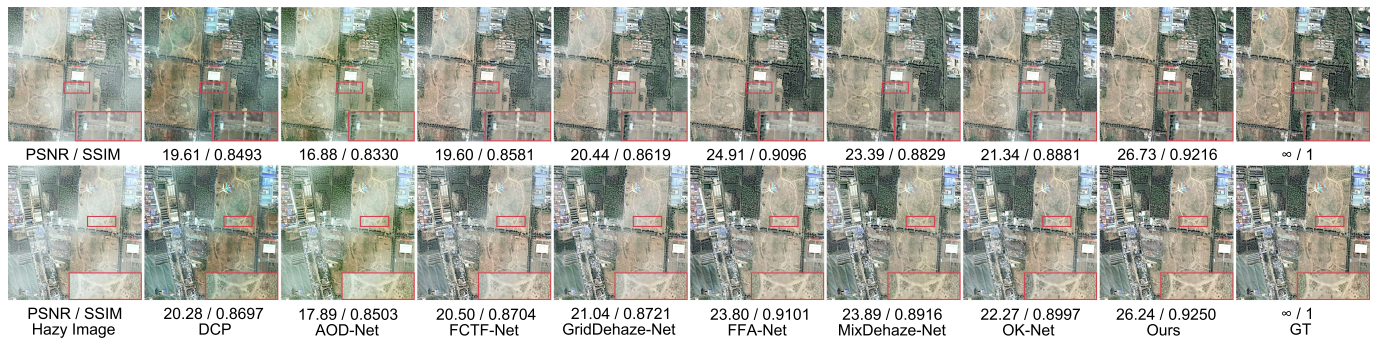


Figure 5. Visual comparison of two images within the Haze1k-thin dataset.

Figure 6 illustrates a visual comparison of various dehazing methods applied to the moderate haze test set. In these examples, moderate haze obscures critical features in the remote sensing (RS) images. DCP demonstrates limited dehazing, leaving significant haze in the processed results, while AOD-Net manages some dehazing but introduces darkening and noticeable color distortion. Methods such as FCTF-Net, GridDehaze-Net, FFA-Net, MixDehaze-Net, and OK-Net deliver generally strong dehazing performance. However, there are still visible discrepancies compared to the original images. For instance, the roof in the lower right corner of the first image restored via FCTF-Net looks washed out, and the plants in the second image appear unnaturally green. In contrast, images restored via GridDehaze-Net, FFA-Net, MixDehaze-Net, and OK-Net exhibit a redder roof in the first image and a white roof in the second image. Based on the performance indicators shown below each image, our method demonstrates the best dehazing capability, with FFA-Net coming in second. Overall, our approach demonstrates improved dehazing performance, delivering a more precise color representation and effectively maintaining texture details that are closely aligned with the ground-truth images.

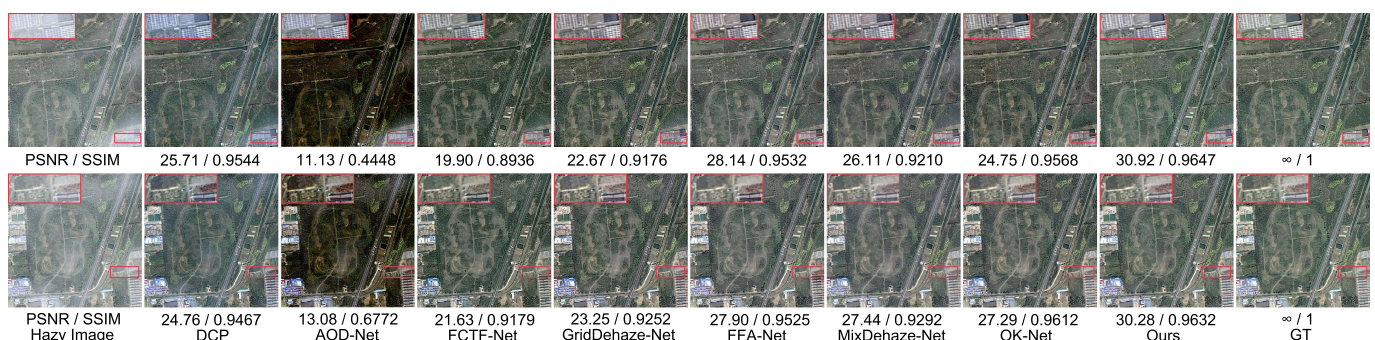


Figure 6. Visual comparison of two images within the Haze1k-moderate dataset.

Figure 7 displays the qualitative outcomes of various methods on the thick-haze test set. This dataset is largely made up of remote sensing images that are significantly affected by dense haze, resulting in considerable loss of texture details. The images restored via DCP and AOD-Net still exhibit considerable haze residue. GridDehaze-Net introduces a noticeable color shift in the restored images; for instance, the red roof in both restored images nearly loses all its color when juxtaposed with the reference images. Additionally, FCTF-Net, FFA-Net, MixDehaze-Net, and OK-Net also suffer from varying levels of color

information loss. In particular, their restored images show a pale or light red roof, deviating from the ground truth. In comparison, our approach exhibits robust performance on all dehazing metrics for both images, aligning closely with the ground-truth images regarding overall structure, intricate details, and color fidelity.

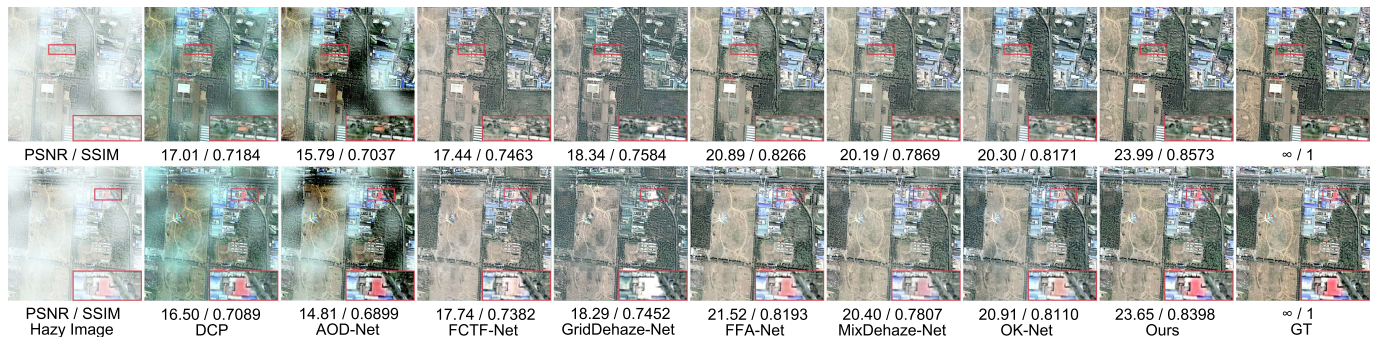


Figure 7. Visual comparison of two images within the Haze1k-thick dataset.

Figure 8 illustrates the qualitative results of different methods applied to the LHID test set. The restoration performed via DCP shows evidence of over-dehazing, leading to the loss of original texture details. On the other hand, the output from AOD-Net appears considerably darker. While FCTF-Net, GridDehaze-Net, FFA-Net, MixDehaze-Net, and OK-Net deliver satisfactory dehazing results in relation to the actual images, significant differences are evident in the PSNR/SSIM metrics presented beneath the images. Our proposed MMPD-Net effectively tackles the issue of texture recovery in remote sensing images affected by blurriness, enhancing image features and elevating the overall visual quality.

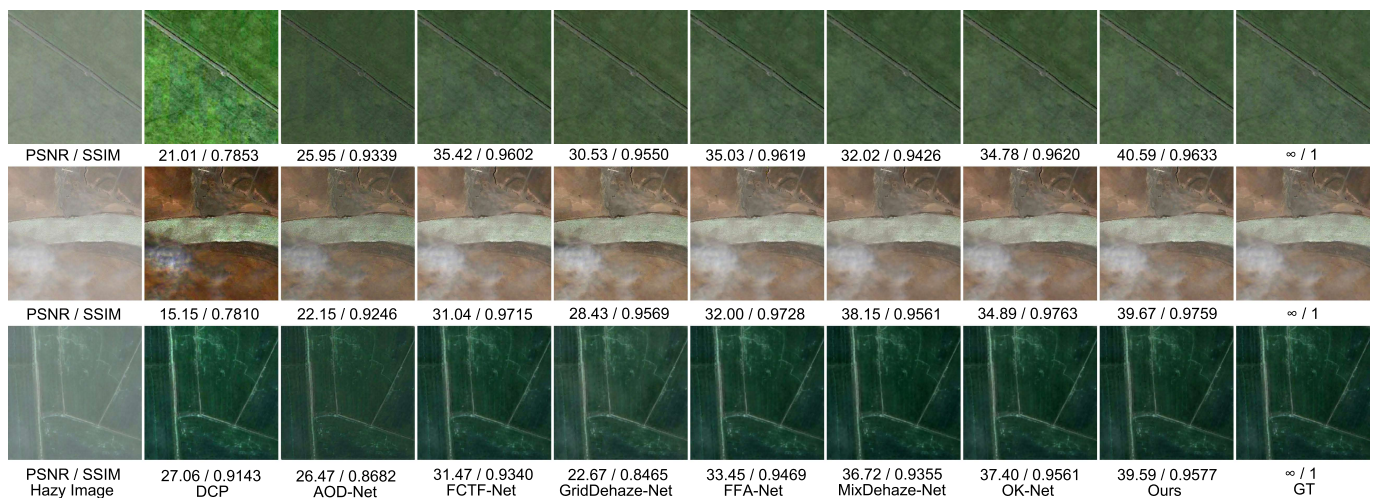


Figure 8. Visual comparison of three images within the LHID data.

Figure 9 presents the qualitative outcomes of different methods evaluated on the DHID test set, which includes remote sensing images affected by dense and uniform haze. The images processed via DCP retain a blurry appearance, hindering the visibility of finer details. Both DCP and AOD-Net yield results that are notably dark and feature black shadows. Although FCTF-Net, GridDehaze-Net, FFA-Net, MixDehaze-Net, and OK-Net succeed in effectively removing the dense haze, their outputs still display some residual black shadows in comparison to the actual images. In contrast, our MMPD-Net excels at preserving color accuracy and maintaining texture details, especially in the colors of roads and lawns.

Our network is evaluated on a real haze remote sensing dataset, and Figure 10 demonstrates its ability to handle complex and diverse conditions typical of the real world. The

dataset contains high-quality images with varying haze intensities, highlighting the challenges of recovering fine details and object visibility. Qualitative results show that our method can effectively remove haze while preserving edge and texture details.



Figure 9. Visual comparison of three images within the DHID data.



Figure 10. Visual comparison of six images within the real dataset.

5. Ablation Study

To assess the effectiveness of our proposed method, we carried out an ablation study utilizing the thin-haze dataset to evaluate the performance of the MMB (multi-dimensional block and multi-scale block), MSK, and ASMU components we developed. The quantitative results are displayed in Table 3. To expedite the training process, we resized the input RGB images to 80×80 pixels while keeping all other configurations consistent with our implementation specifications.

The basic architecture uses the foundational blocks from FFA [17] and modules without dimensional expansion as the basic learning blocks, replacing MSK with SK and removing ASMU. As shown in Table 6, we conducted a detailed analysis of the contribution of each module, including MSB, MDB, MMB, MSK, and ASMU. Each proposed component positively impacted the overall performance. Figure 11 provides a visual representation of the qualitative results from the ablation experiments performed on different modules,

offering a more straightforward illustration of the effectiveness of each component in our network.

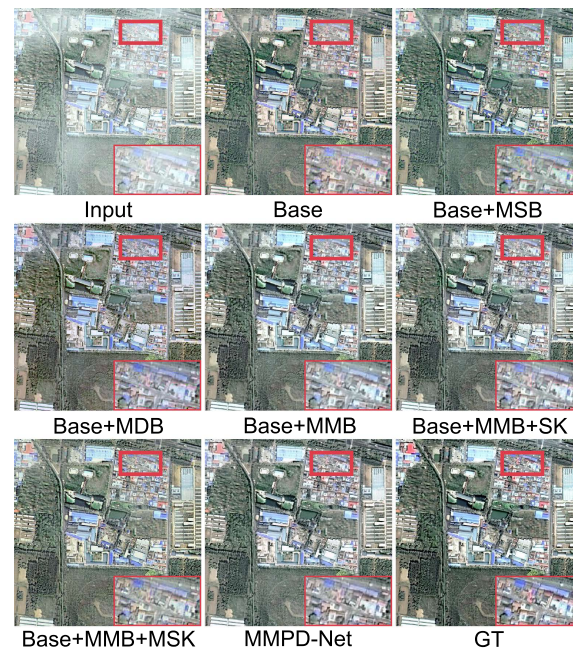


Figure 11. Visual comparison of results from the ablation experiments performed on the thin-haze dataset.

Table 6. Ablation study conducted on the thin-haze dataset. **Bold** values indicate the highest performance outcomes.

Methods	Thin Haze	
	PSNR	SSIM
Base	19.57	0.8249
Base + MSB	20.02	0.8297
Base + MDB	21.37	0.8621
Base + MMB	21.65	0.8704
Base + MMB + SK	21.57	0.8751
Base + MMB + MSK	21.68	0.8821
Base + MMB + MSK + ASMU	22.81	0.8962

In the Base state of the network, the PSNR and SSIM values are at their lowest, with severe distortion in the red-boxed regions, a reddish tone, and overall poor image quality. Adding MSB and MDB individually to the base network resulted in significant improvements, with PSNR increasing by 0.45 and 1.8 and SSIM improving by 0.0048 and 0.0372, respectively. Combining MSB and MDB into MMB led to further enhancements in metrics. While image colors were corrected, and overall quality significantly improved, slight blurriness remained, though MMB proved effective in enhancing fine details. To validate the improvement of MSK over SK, we conducted a comparative experiment in which integrating MSK into the network achieved further improvements in PSNR and SSIM, with noticeable enhancements in the red-boxed regions, demonstrating MSK's effectiveness in refining image details and enhancing overall quality. Finally, the integration of ASMU achieved additional improvements, with PSNR increasing by 1.13 and SSIM by 0.0141. In the red-boxed regions, the roofs and ground of different colors closely resembled the ground-truth image, confirming that ASMU effectively enhances fine details while maintaining high similarity to the reference image.

6. Conclusions

This paper has proposed a fine-detail extraction network to address the dehazing problem in remote sensing images. MMPD-Net includes blocks that fuse multi-scale and multi-dimensional features, consisting of a multi-scale extraction block and a multi-dimensional extraction block. The multi-scale extraction block uses an attention mechanism to extract detailed features. In order to extract pixel information at multiple scales, we apply dilated convolutions to capture details at different scales before pixel attention. The multi-dimensional extraction block extracts details through deep convolution layers, and then it expands the number of channels. After that, it performs dimensionality enhancement via the element-wise product, followed by deep convolution layers to extract the details of the enhanced dimensions. To enhance the model's dehazing ability, we introduced the atmospheric scattering model, using different convolutional kernels to extract A and t from the model. This approach improves the stability and generalization capability of the model's dehazing ability. We also improved SK, which lacked significant feature extraction, and proposed MSK, further improving the quality of remote sensing images. Experiments demonstrate that our method attains state-of-the-art dehazing results on the SateHaze-1k and HRSD benchmark datasets, effectively restoring intricate image details. In the thick-haze dataset, the dehazing network does not recover details sufficiently. We plan to further improve the dehazing network and develop a more robust, lightweight model tailored to complex scenarios for remote sensing images. Additionally, we aim to create a more comprehensive, large-scale public dataset focused on haze in remote sensing images in order to advance research in this field.

Author Contributions: Writing—original draft preparation, L.W.; writing—review and editing, H.Z.; validation, Q.L.; supervision, T.T.; resources, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (62401012), the Anhui University of Technology Young Teachers Research Fund Project (QZ202313), the University Synergy Innovation Program of Anhui Province (Grant No. GXXT-2023-021), and the Key Program of the Natural Science Foundation of the Educational Department of Anhui Province of China (Grant No. 2022AH050319).

Data Availability Statement: The StateHaze1k dataset and HRSD dataset are made publicly available for research purposes. For more information, please refer to the following links: SateHaze1k <https://www.dropbox.com/s/k2i3p7puuw12g59/Haze1k.zip?dl=0> (accessed on 20 September 2023) and HRSD <https://github.com/Shan-rs/DCI-Net> (accessed on 20 September 2023).

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

References

1. Kulk, G.; Platt, T.; Dingle, J.; Jackson, T.; Jönsson, B.F.; Bouman, H.A.; Babin, M.; Brewin, R.J.; Doblin, M.; Estrada, M.; et al. Primary production, an index of climate change in the ocean: satellite-based estimates over two decades. *Remote Sens.* **2020**, *12*, 826. [\[CrossRef\]](#)
2. Kumar, P.; Singh, S.; Pandey, A.; Singh, R.K.; Srivastava, P.K.; Kumar, M.; Dubey, S.K.; Sah, U.; Nandan, R.; Singh, S.K.; et al. Multi-level impacts of the COVID-19 lockdown on agricultural systems in India: The case of Uttar Pradesh. *Agric. Syst.* **2021**, *187*, 103027. [\[CrossRef\]](#)
3. Amaro García, A. Relationship between blue economy, cruise tourism, and urban regeneration: Case study of Olbia, Sardinia. *J. Urban Plan. Dev.* **2021**, *147*, 05021029. [\[CrossRef\]](#)
4. Li, S.; Fang, H.; Zhang, Y. Determination of the leaf inclination angle (LIA) through field and remote sensing methods: Current status and future prospects. *Remote Sens.* **2023**, *15*, 946. [\[CrossRef\]](#)
5. McCartney, E. *Optics of the Atmosphere: Scattering by Molecules and Particles*; John Wiley and Sons, Inc.: New York, NY, USA, 1976.
6. 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; IEEE: New York, NY, USA, 1999; Volume 2, pp. 820–827.
7. Kulkarni, A.; Phutke, S.S.; Vipparthi, S.K.; Murala, S. C2AIR: Consolidated Compact Aerial Image Haze Removal. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2024; pp. 749–758.
8. Ali, A.; Sarkar, R.; Chaudhuri, S.S. Wavelet-based Auto-Encoder for simultaneous haze and rain removal from images. *Pattern Recognit.* **2024**, *150*, 110370. [\[CrossRef\]](#)

9. Wang, T.; Tao, G.; Lu, W.; Zhang, K.; Luo, W.; Zhang, X.; Lu, T. Restoring vision in hazy weather with hierarchical contrastive learning. *Pattern Recognit.* **2024**, *145*, 109956. [[CrossRef](#)]
10. Liu, Y.; Yan, Z.; Chen, S.; Ye, T.; Ren, W.; Chen, E. Nighthazeformer: Single nighttime haze removal using prior query transformer. In Proceedings of the Proceedings of the 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 4119–4128.
11. Yan, Q.; Gong, D.; Zhang, Y. Two-Stream Convolutional Networks for Blind Image Quality Assessment. *IEEE Trans. Image Process.* **2019**, *28*, 2200–2211. [[CrossRef](#)]
12. Zhang, Y.; Zhang, H.; Nasrabadi, N.M.; Huang, T.S. Multi-metric learning for multi-sensor fusion based classification. *Inf. Fusion* **2013**, *14*, 431–440. [[CrossRef](#)]
13. 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](#)]
14. Pang, Y.; Xie, J.; Li, X. Visual haze removal by a unified generative adversarial network. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 3211–3221. [[CrossRef](#)]
15. Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; Yang, M.H. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 11–15 June 2020; pp. 2157–2167.
16. 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, Montreal, BC, Canada, 11–17 October 2019; pp. 7314–7323.
17. 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.
18. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693.
19. Song, Y.; He, Z.; Qian, H.; Du, X. Vision transformers for single image dehazing. *IEEE Trans. Image Process.* **2023**, *32*, 1927–1941. [[CrossRef](#)] [[PubMed](#)]
20. Nie, J.; Xie, J.; Sun, H. Remote Sensing Image Dehazing via a Local Context-Enriched Transformer. *Remote Sens.* **2024**, *16*, 1422. [[CrossRef](#)]
21. Shi, Y.; Xia, B.; Jin, X.; Wang, X.; Zhao, T.; Xia, X.; Xiao, X.; Yang, W. Vmambair: Visual state space model for image restoration. *arXiv* **2024**, arXiv:2403.11423.
22. Wang, J.; Wu, S.; Yuan, Z.; Tong, Q.; Xu, K. Frequency compensated diffusion model for real-scene dehazing. *Neural Netw.* **2024**, *175*, 106281. [[CrossRef](#)]
23. Huang, Y.; Xiong, S. Remote sensing image dehazing using adaptive region-based diffusion models. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 8001805. [[CrossRef](#)]
24. Zhou, H.; Chen, Z.; Liu, Y.; Sheng, Y.; Ren, W.; Xiong, H. Physical-priors-guided DehazeFormer. *Knowl. Based Syst.* **2023**, *266*, 110410. [[CrossRef](#)]
25. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
26. Huang, B.; Zhi, L.; Yang, C.; Sun, F.; Song, Y. Single satellite optical imagery dehazing using SAR image prior based on conditional generative adversarial networks. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1806–1813.
27. Zhang, L.; Wang, S. Dense haze removal based on dynamic collaborative inference learning for remote sensing images. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–16. [[CrossRef](#)]
28. Tan, R.T. Visibility in bad weather from a single image. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; IEEE: New York, NY, USA, 2008; pp. 1–8.
29. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.
30. Fattal, R. Dehazing using color-lines. *Acm Trans. Graph. (Tog)* **2014**, *34*, 1–14. [[CrossRef](#)]
31. Tang, K.; Yang, J.; Wang, J. Investigating haze-relevant features in a learning framework for image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2995–3000.
32. 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.
33. Berman, D.; Avidan, S. Non-local image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1674–1682.
34. 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.
35. Mei, K.; Jiang, A.; Li, J.; Wang, M. Progressive feature fusion network for realistic image dehazing. In Proceedings of the Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; Revised Selected Papers, Part I 14; Springer: Berlin/Heidelberg, Germany, 2019; pp. 203–215.

36. Lu, L.; Xiong, Q.; Xu, B.; Chu, D. Mixdehazenet: Mix structure block for image dehazing network. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; IEEE: New York, NY, USA, 2024; pp. 1–10.
37. Cui, Y.; Ren, W.; Knoll, A. Omni-Kernel Network for Image Restoration. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 1426–1434.
38. Sun, H.; Luo, Z.; Ren, D.; Hu, W.; Du, B.; Yang, W.; Wan, J.; Zhang, L. Partial siamese with multiscale bi-codec networks for remote sensing image haze removal. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 4106516. [\[CrossRef\]](#)
39. Song, T.; Fan, S.; Li, P.; Jin, J.; Jin, G.; Fan, L. Learning an effective transformer for remote sensing satellite image dehazing. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 8002305. [\[CrossRef\]](#)
40. He, Y.; Li, C.; Li, X.; Bai, T. A Lightweight CNN Based on Axial Depthwise Convolution and Hybrid Attention for Remote Sensing Image Dehazing. *Remote Sens.* **2024**, *16*, 2822. [\[CrossRef\]](#)
41. 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, 18–23 June 2018; pp. 3194–3203.
42. Liu, Y.; Pan, J.; Ren, J.; Su, Z. Learning deep priors for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2492–2500.
43. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Zheng, Y.; Zhan, J.; He, S.; Dong, J.; Du, Y. Curricular contrastive regularization for physics-aware single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5785–5794.
45. Chalavadi, V.; Jeripothula, P.; Datla, R.; Ch, S.B.; C. Krishna, M. mSODANet: A network for multi-scale object detection in aerial images using hierarchical dilated convolutions. *Pattern Recognit.* **2022**, *126*, 108548. [\[CrossRef\]](#)
46. Ma, X.; Dai, X.; Bai, Y.; Wang, Y.; Fu, Y. Rewrite the Stars. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024, pp. 5694–5703.
47. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–19 June 2019; pp. 510–519.
48. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Learning enriched features for fast image restoration and enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1934–1948. [\[CrossRef\]](#)
49. Liu, B.; Chen, S.B.; Wang, J.X.; Tang, J.; Luo, B. An Oriented Object Detector for Hazy Remote Sensing Images. *IEEE Trans. Geosci. Remote. Sens.* **2024**, *62*, 1001711. [\[CrossRef\]](#)
50. Li, Y.; Chen, X. A coarse-to-fine two-stage attentive network for haze removal of remote sensing images. *IEEE Geosci. Remote. Sens. Lett.* **2020**, *18*, 1751–1755. [\[CrossRef\]](#)

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.