remote sensing

MDPI

Article

DWTMA-Net: Discrete Wavelet Transform and Multi-Dimensional Attention Network for Remote Sensing Image Dehazing

Xin Guan [1], Runxu He [1], Le Wang [2], Hao Zhou [2], Yun Liu [3] and Hailing Xiong [4,*]

[1] College of Computer and Information Science, Southwest University, Chongqing 400715, China; guanxin2020@email.swu.edu.cn (X.G.); hrx221226@email.swu.edu.cn (R.H.)
[2] School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China; wangle@ahut.edu.cn (L.W.); haozhou@ahut.edu.cn (H.Z.)
[3] College of Artificial Intelligence, Southwest University, Chongqing 400715, China; yunliu@swu.edu.cn
[4] College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
* Correspondence: xionghl@swu.edu.cn

Abstract: Haze caused by atmospheric scattering often leads to color distortion, reduced contrast, and diminished clarity, which significantly degrade the quality of remote sensing images. To address these issues, we propose a novel network called DWTMA-Net that integrates discrete wavelet transform with multi-dimensional attention, aiming to restore image information in both the frequency and spatial domains to enhance overall image quality. Specifically, we design a wavelet transform-based downsampling module that effectively fuses frequency and spatial features. The input first passes through a discrete wavelet block to extract frequency-domain information. These features are then fed into a multi-dimensional attention block, which incorporates pixel attention, Fourier frequency-domain attention, and channel attention. This combination allows the network to capture both global and local characteristics while enhancing deep feature representations through dimensional expansion, thereby improving spatial-domain feature extraction. Experimental results on the SateHaze1k, HRSD, and HazyDet datasets demonstrate the effectiveness of the proposed method in handling remote sensing images with varying haze levels and drone-view scenarios. By recovering both frequency and spatial details, our model achieves significant improvements in dehazing performance compared to existing state-of-the-art approaches.

Keywords: remote sensing; image dehazing; wavelet transformation; attention

check for updates

Academic Editor: Konstantinos Topouzelis

Received: 25 April 2025
Revised: 7 June 2025
Accepted: 9 June 2025
Published: 12 June 2025

Citation: Guan, X.; He, R.; Wang, L.; Zhou, H.; Liu, Y.; Xiong, H. DWTMA-Net: Discrete Wavelet Transform and Multi-Dimensional Attention Network for Remote Sensing Image Dehazing. *Remote Sens.* **2025**, *17*, 2033. https://doi.org/10.3390/rs17122033

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The quality of remote sensing imagery is often compromised by complex atmospheric interferences, including haze and semi-transparent clouds. Not only do these detrimental factors weaken the signal quality, but they also distort visual information and obscure important details, thereby limiting large-scale data collection and real-time monitoring. Therefore, it becomes challenging to restore the information of remote sensing images. High-quality and clear images can provide richer and more accurate visual information, which is essential to improve the performance and reliability of various downstream tasks [1–4] (in areas such as temporal variation analysis, hazard assessment, ecological evaluation, and defense-related observations).

Existing image dehazing solutions can be grouped into two fundamental paradigms: prior-informed algorithms and data-driven learning approaches. In the early stages of

research, prior-based methods were proposed to reduce the impact of haze on images. These methods typically rely on the atmospheric scattering model (ASM) [5] to reconstruct clear images. Nevertheless, these methods frequently lack robustness and fail to effectively adapt to the highly variable haze characteristics present in diverse imaging scenarios, thereby limiting their real-world applicability.

Deep learning-powered dehazing techniques [6–17] have emerged as highly effective alternatives, exhibiting superior generalization and enhanced performance when compared to prior-guided methods. Through the use of comprehensive datasets, such techniques can autonomously map hazy to clear imagery, bypassing reliance on predefined physical models. This allows them to excel even in variable atmospheric conditions, as they effectively capture intricate features and trends from the data. Although early deep learning-based dehazing methods [18,19] were designed based on the atmospheric scattering model, their practical application in real-world haze scenes remains challenging. This is mainly because the physical scattering model cannot fully capture the complexity and diversity of real atmospheric conditions, limiting the effectiveness of these methods in complex environments.

To bypass the limitations of traditional learning-based dehazing techniques, recent approaches have adopted end-to-end learning architectures that aim to completely eliminate the dependence on physical modeling. Among these approaches, multi-scale convolutional neural networks (CNNs) [20,21] have gained significant attention, as they can directly learn the mapping from hazy images to their clear counterparts. The effectiveness of these methods largely stems from their ability to automatically extract rich and discriminative features through stacked convolution layers. However, a fundamental limitation of convolution operations is their inherently local nature, which restricts the model's ability to capture long-range dependencies and global contextual information. In response to this issue, researchers have employed hierarchical feature extraction techniques, along with adaptive focus modules, to enhance the model's ability to capture scene-wide information during haze removal. A notable example is FFA-Net [22], which employs a channel-wise attention mechanism to model non-local dependencies across different image regions, thereby effectively enhancing the network's ability to incorporate global information and improve dehazing performance.

The field of image dehazing has seen rapid progress in recent years, including Transformer-based networks [23–25], the Mamba paradigm [26], and diffusion-based frameworks [27–29], all of which have contributed to notable improvements through novel architectural designs. By utilizing self-attention, Transformer networks are capable of modeling distant spatial dependencies, addressing the limitations inherent in convolutional structures with restricted receptive scopes. Modeling global context allows the dehazing network to capture the holistic structure of the scene, thereby producing outputs with higher fidelity and visual consistency. The Mamba framework further enhances dehazing accuracy and detail recovery by incorporating multi-scale learning and efficient feature fusion strategies. This framework leverages a multi-scale feature extraction network, allowing it to effectively handle haze at varying intensities while maintaining robustness in complex scenes. Diffusion models, as generative models, have also shown great potential for image dehazing by simulating a gradual denoising process. These models learn to reverse the noise process, helping to recover clear images while preserving finer details and structural information, making them well-suited for addressing complex haze and visual impairments. Together, Transformer-based methods, the Mamba framework, and diffusion models provide more precise and flexible solutions for image dehazing.

This paper proposes a dehazing network called DWTMA-Net, which is designed to restore both frequency- and spatial-domain information in remote sensing images. Built on a

U-shaped architecture, the model consists of three key modules: the Discrete Wavelet Block (DWB), the Multi-dimensional Attention Block (MAB), and the Wavelet Downsampling Module (WDM). The DWB uses the Haar Discrete Wavelet Transform (DWT) to decompose features into four frequency components, where low-frequency features are processed by a small AOD network for feature extraction, and high-frequency features are refined using dilated residual blocks. The inverse wavelet transform is then applied to reconstruct spatial information. The MAB employs depthwise separable convolutions for deep feature extraction, followed by convolutions with various kernel sizes to enhance feature diversity. It further applies channel attention, pixel attention, and Fourier frequency attention, integrating them into a multi-dimensional attention mechanism to capture both global and local features. Meanwhile, the WDM leverages the Haar DWT for downsampling, combining frequency information from the wavelet transform with spatial information from convolutional downsampling for improved feature representation.

Main Contributions of This Paper

- A novel model is proposed that combines frequency-domain information from the discrete wavelet transform (DWT) with spatial-domain features from convolution. Validation using the complex SateHaze1k [30], HRSD [31], and HazyDet [32] datasets confirms its effectiveness in enhancing detail and visual quality.
- To enhance spatial-domain feature information, a novel multi-dimensional attention module is proposed, applying different attention mechanisms to various features extracted through different convolutions.
- To achieve frequency-domain processing, a novel frequency processing module is proposed, which extracts and refines features from four distinct frequency components generated by the Haar discrete wavelet transform (DWT).
- To capture both frequency- and spatial-domain features, a novel downsampling method is proposed, combining Haar wavelet transform and convolution for effective downsampling.

2. Related Works

2.1. Prior-Guided Image Dehazing Methods

Traditional image dehazing methods based on prior knowledge typically utilize statistical assumptions and physical constraints derived from haze characteristics to infer the transmission map and estimate atmospheric illumination. Among the earliest contributions, Tan et al. [33] exploited the contrast difference between hazy and haze-free images to enhance visibility in degraded scenes. The technique focuses on amplifying local contrast to boost image visibility and reconstruct haze-free content. He et al. [34] used the dark channel prior to estimate the transmission map, assuming that haze-free images contain at least one color channel with low intensity in most non-sky regions. It then applies the atmospheric scattering model to recover clear images. Fattal [35] utilized the observation that color lines in the RGB space remain invariant in hazy images and applied this property to estimate the transmission map and recover clear images. Tang et al. [36] advanced prior-driven dehazing by fusing diverse priors related to haze appearance, including edge sharpness, color richness, and contrast, and utilized a random forest regressor to learn the transmission map estimation process. Zhu et al. [37] relied on the observation that haze causes a color attenuation effect, particularly in the blue channel, and utilized this prior to estimate the transmission map and restore clear images efficiently. According to Berman et al. [38], haze induces the transformation of pixel clusters in clear images into haze-like structures. They leveraged this phenomenon to propose a non-local prior aimed at representing clean image characteristics.

By analyzing the haze formation process and simulating its physical characteristics, the physical prior-based method reconstructs clear images from hazy ones. The foundation of this approach is the atmospheric scattering model detailed below:

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

In this model, $I(x)$ denotes the hazy image captured by the camera, while $J(x)$ represents the corresponding clear image. The transmission map ($t(x)$) characterizes how much light from the scene directly reaches the camera, and A stands for the global atmospheric illumination. Accurate estimation of $t(x)$ and A allows for the recovery of the original scene radiance ($J(x)$), thereby producing a haze-free image.

2.2. Data-Driven Approaches for Image Dehazing

These data-driven techniques typically utilize deep learning architectures to either predict transmission and atmospheric components informed by physical scattering mechanisms or transform hazy images into their dehazed versions through end-to-end learning, bypassing the need for handcrafted physical assumptions. The first strategy incorporates the principles of atmospheric scattering to guide the training procedure, whereas the second directly establishes a haze-to-clear transformation pipeline, eliminating the reliance on traditional physical formulations. In one of the earliest works using CNNs for haze removal, Cai et al. [18] developed DehazeNet, which processes hazy images through a learnable architecture to estimate the corresponding transmission maps. Li et al. [39] proposed an all-in-one dehazing network with the aim of simultaneously estimating both the atmospheric light and the transmission map, directly producing haze-free images. It uses an adaptive network structure to improve the accuracy and efficiency of haze removal in a unified framework. Liu et al. [20] employed an attention-based multi-scale network for image dehazing, incorporating grid-based attention mechanisms to focus on haze-affected regions. This approach enables the model to effectively capture both global and local features, improving haze removal across varying densities. Qin et al. [22] utilized a feature fusion attention network that combines multi-scale feature fusion with attention mechanisms to enhance haze removal while preserving important image details. Lu et al. [40] employed a mixed-structure block that integrates multiple network components to capture both global and local features, improving the performance of image dehazing. Similarly, Cui et al. [41] used an omni-kernel convolutional approach that combines multiple kernel sizes in a unified framework to capture diverse image features. This method allows the network to effectively handle various image restoration tasks, including dehazing, by adapting to different spatial structures. Sui et al. [42] proposed a U-shaped dual attention network based on the Vision Mamba architecture, which utilizes multi-scale feature extraction and attention mechanisms to effectively remove haze from satellite remote sensing images.

3. Method

As shown in Figure 1, the overall structure of DWTMA-Net adopts a U-shaped design and is divided into five levels. Each level embeds a Frequency Feature Extraction Block (FFEB), which integrates the proposed DWB and MAB modules in series. The input and output dimensions of the DWB and MAB modules are $B \times C \times H \times W$ for the first and fifth levels, $B \times 2C \times \frac{H}{2} \times \frac{W}{2}$ for the second and fourth levels, and $B \times 4C \times \frac{H}{4} \times \frac{W}{4}$ for the third level. The SK fusion module serves as an adaptive mechanism for fusing multi-scale or cross-level features, aiming to enhance detail and structural information in dehazed images. The WDM module employs the Haar discrete wavelet transform (DWT) to reduce feature resolution by half and combines it with convolutional downsampling to capture both frequency- and spatial-domain information. The DWB module transforms

features into the frequency domain and applies customized processing strategies to different frequency components. The MAB module leverages multiple attention mechanisms to process spatial information and dynamically capture edge details, global context, and multi-dimensional features.

This network achieves multi-scale feature fusion through an encoder–decoder architecture with skip connections, balancing the need for global haze distribution modeling and local detail preservation. The core innovation lies in the dual-branch wavelet downsampling module: the Haar wavelet branch explicitly decomposes frequency-domain sub-bands (LL/LH/HL/HH), effectively preventing the loss of high-frequency information typically caused by traditional downsampling; meanwhile, the parallel convolutional branch extracts spatial-domain features, enhancing adaptability to the spatial distribution of haze. The two branches are fused via element-wise addition to achieve complementary modeling between the frequency and spatial domains. Combined with the wavelet processing block (for frequency-domain feature enhancement) and the multi-dimensional attention block (for dynamic feature recalibration), this forms a collaboratively optimized feature representation mechanism. The inverse wavelet transform further ensures lossless reconstruction during the decoding phase.

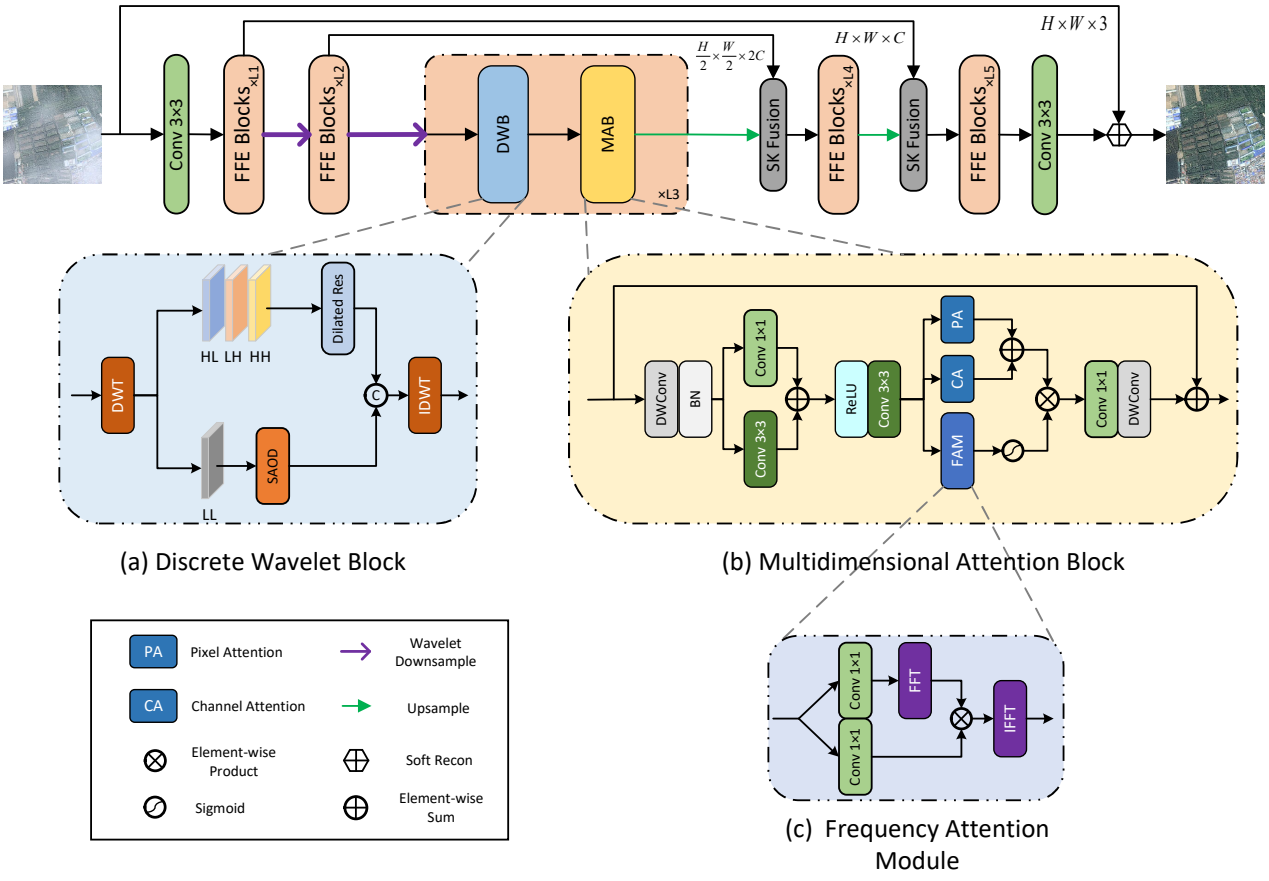


Figure 1. Structure of discrete wavelet transform and multi-dimensional attention.

3.1. Wavelet Downsampling Module

Our WDM extends traditional downsampling by incorporating discrete wavelet transform (DWT) to capture frequency information, in contrast to conventional methods that rely solely on convolutions for feature size reduction. By integrating the sampled spatial and frequency information, the WDM performs an additive fusion of convolutional downsampling and wavelet transform, as shown in Figure 2.

$$\hat{x} = \mathrm{Conv}(x) \oplus \mathrm{Conv}(\mathrm{DWT}(x)). \tag{2}$$

where $\hat{x}$ represents the output of downsampling; x represents the input of the previous stage; and $\mathrm{Conv}(\cdot)$ and $\mathrm{DWT}(\cdot)$ represent the convolution and wavelet transform processing of the input of the previous stage, respectively.

Our downsampling module is designed with a dual-branch architecture to effectively capture both frequency-domain and spatial-domain information, enhancing the representational capacity of the network. The first branch applies a Haar wavelet transform to the input feature map, decomposing it into four sub-bands: LL, LH, HL, and HH. The LL sub-band retains low-frequency components that represent the overall structure and contours of the image, while the LH, HL, and HH sub-bands extract directional high-frequency details such as edges and textures. This branch enables explicit modeling of multi-scale and multi-directional frequency features, which is particularly beneficial for capturing fine details and haze boundaries. The second branch employs a standard convolution with a stride of 2 to perform spatial downsampling, preserving local context and semantic structure in the spatial domain. The outputs of the two branches are fused via element-wise addition, allowing the network to integrate complementary features from both domains. This design improves the network's ability to perceive structural and textural details, leading to more effective dehazing in complex remote sensing scenarios.

In the network design, we innovatively improve the encoder's downsampling process by introducing a self-developed dual-branch wavelet downsampling module in the downsampling layers, replacing the traditional max pooling or strided convolution operations. This module integrates a multi-scale frequency-domain feature extraction mechanism, significantly enhancing the network's ability to represent detailed image features while maintaining efficient downsampling.

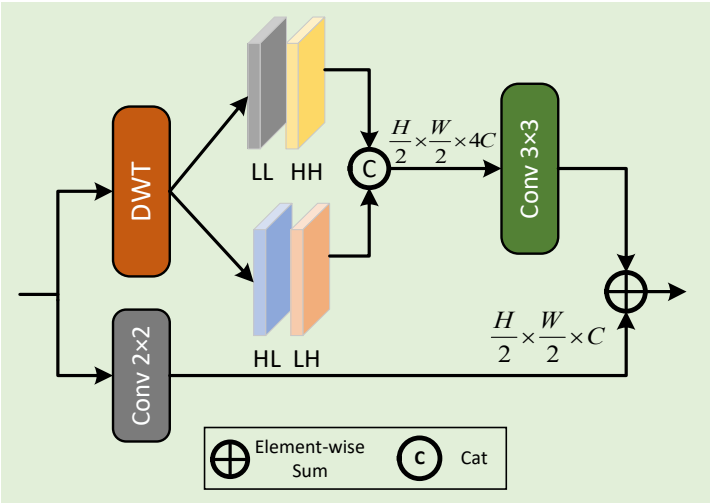


Figure 2. Structure of WDM.

3.2. Discrete Wavelet Block

Our FFEB first extracts features in the frequency domain from the wavelet-downsampled features, then performs spatial-domain feature extraction before producing the final output, as shown in Figure 1.

Wavelet transform can reduce the spatial dimensions by half with each transformation without sacrificing information, unlike other techniques, such as Fast Fourier Transform

(FFT) and Discrete Cosine Transform (DCT), which may result in information loss. Haar wavelet transform converts the input into four sub-bands, i.e.,

$$f_{LL},\, f_{HL},\, f_{LH},\, f_{HH} = \mathrm{DWT}(\widehat{x}). \tag{3}$$

In the discrete wavelet block (DWB), we first apply the Haar discrete wavelet transform to obtain four sub-band frequency features. The LL sub-band typically contains most of the signal energy, while the other three sub-bands capture edge and detail information. The LL sub-band is processed using a small AOD network for feature extraction, while the other three sub-bands undergo refinement using dilated residual blocks to enhance high-frequency features. Finally, the inverse wavelet transform is applied to produce the output feature map.

Drawing inspiration from AOD [39], we designed a lightweight Small AOD Dehazing Block (SAOD) grounded in the analytical formulation presented in Equation (4), with its structural details illustrated in Figure 3. First, according to the physical model, the clear image (J) is expressed as follows:

$$\begin{aligned} J(x) &= K(x)I(x) - K(x) + b \\ K(x) &= \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1} \end{aligned} \tag{4}$$

In the equation, $K(x)$ represents a parameter that fuses $t(x)$ and A from the atmospheric scattering model into a single term, and b denotes the bias term.

Considering b as a bias, we adopt a learning-based approach to estimate this bias. To begin with, we employ global average pooling to compress the feature dimensions and filter out repetitive or non-informative content from the representation space. GAP computes the average value of the feature map across its spatial dimensions, resulting in a one-dimensional feature vector that aligns with the characteristics of the bias value. This vector then undergoes a 1 × 1 convolution for feature transformation, followed by sigmoid activation to obtain the bias (b). The estimated b is represented as follows:

$$b = \sigma(\mathrm{Conv}(\mathrm{LeakyRelu}(\mathrm{Conv}(\mathrm{GAP}(f_{LL}))))). \tag{5}$$

However, since the transmission map (K) is non-homogeneous, applying GAP would result in information loss. Consequently, we employ stacked convolutional layers with a 3 × 3 kernel size to facilitate feature learning, as elaborated below:

$$K = \sigma(\mathrm{Conv}(\mathrm{LeakyRelu}(\mathrm{Conv}(\mathrm{Conv}(f_{LL}))))). \tag{6}$$

Therefore, through SAOD, we obtain the $f_{LL}^{'}$ feature as follows:

$$f_{LL}^{'} = K \otimes f_{LL} \ominus K \oplus b \tag{7}$$

For the remaining three sub-bands, we employ dilated residual blocks to refine the features. These blocks use dilated convolutions with dilation rates of 1, 2, and 1; a kernel size of 3; and a stride of 1, as illustrated in Figure 3. Subsequently, the refined features are restored to the spatial domain using the inverse wavelet transform, as described below:

$$\begin{aligned} f_{HL}^{'},\, f_{LH}^{'},\, f_{HH}^{'} &= \mathrm{DilateRes}(f_{HL},\, f_{LH},\, f_{HH}), \\ \widehat{x} &= \mathrm{IDWT}(f_{LL}^{'},\, f_{HL}^{'},\, f_{LH}^{'},\, f_{HH}^{'}). \end{aligned} \tag{8}$$

In this formulation, $\sigma(\cdot)$ is the sigmoid nonlinearity, and $\mathrm{Conv}(\cdot)$ indicates a standard convolutional layer. While 1×1 convolutions are designed for manipulation of channel-wise information and scaling dimensions, 3×3 kernels are favored for their efficiency in extracting spatial context and identifying localized features, including edge and texture patterns.

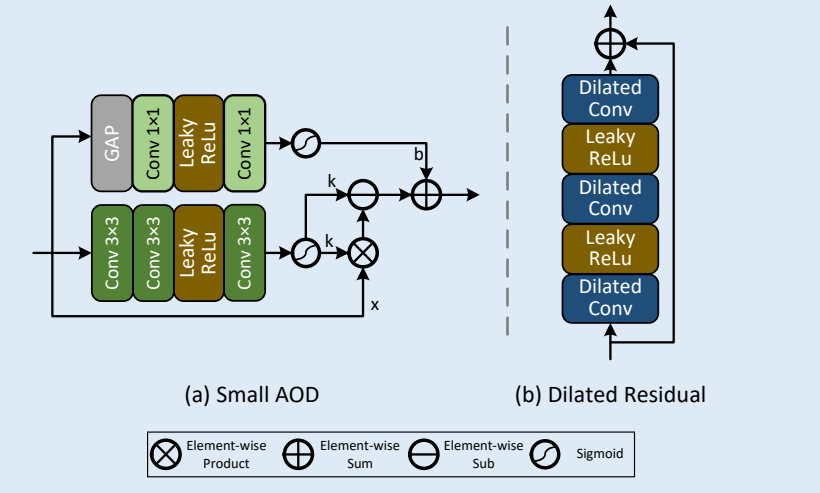


Figure 3. Structure of SAOD and dilated residual block.

3.3. Multi-Dimensional Attention Block

Within the MAB framework, a depthwise convolution (DWConv) is first applied, ensuring the preservation of feature details while improving computational efficiency. We then apply normalization to ensure consistent feature scales and facilitate more rapid network training. To enhance feature diversity, we use convolutions with a kernel size of 1 and 3 to increase the number of channels. The convolution with a kernel size of 1 mainly handles channel transformation and dimensional adjustment, while the convolution with a kernel size of 3 captures local patterns. After extracting and normalizing the features, they are fused to provide a more comprehensive input representation. Finally, another convolution with a kernel sizeof 3 is applied to further refine the features.

$$\begin{aligned} x_0 &= \mathrm{BN}(\mathrm{DWConv}(\widehat{x})), \\ x_1 &= \mathrm{Conv3}(\mathrm{Relu}(\mathrm{Conv}(x_0) \oplus \mathrm{Conv}(x_0))). \end{aligned} \tag{9}$$

According to FFA-Net [22], pixel attention is essential for isolating scale-relevant structures through the enhancement of significant pixel information. Pixel-level attention is designed to pinpoint and enhance critical spatial areas within an image—an ability especially valuable in dehazing tasks, where preserving local visibility is vital. To complement this, we incorporate a Channel Attention (CA) module operating in parallel. While the spatial branch emphasizes localized clarity, the channel branch selectively boosts haze-relevant responses based on global contextual cues. This combined attention strategy enables the model to effectively integrate detailed textures with high-level semantic features.

$$\begin{aligned} x_a &= \mathrm{PA}(x_1) \oplus \mathrm{CA}(x_1), \\ x_b &= \sigma(\mathrm{FAM}(x_1)). \end{aligned} \tag{10}$$

The outputs from the dual attention modules are fused through element-wise summation, yielding a richer and more informative feature representation. Inspired by the strategy proposed by Ma et al. [43], we apply a frequency-aware modulation using a Frequency Attention Module (FAM) scaled via a sigmoid activation to adaptively adjust the feature

intensity and improve representational expressiveness. To further refine the features, a 1 × 1 convolution is employed to reduce dimensionality, condense critical information, and mitigate overfitting risks. This is followed by a 3 × 3 convolution to expand contextual understanding, alongside a residual connection that preserves the original signal. The final output effectively combines both refined enhancements and retained inputs, ensuring robustness for downstream tasks.

$$y_{out} = \mathrm{DWConv}(\mathrm{Conv}(x_a \otimes x_b)) \oplus \widehat{x}. \tag{11}$$

In this expression, $\sigma(\cdot)$ denotes the sigmoid activation, while $\otimes$ indicates an element-wise multiplication. The sigmoid operation is mathematically defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{12}$$

3.4. Loss Function

While L_2 loss is widely adopted in dehazing tasks, empirical results from recent image restoration studies reveal that L_1 loss can achieve more favorable performance, particularly in terms of PSNR and SSIM. Therefore, we adopt the simpler L_1 loss as our primary objective. To further enhance the restoration of frequency details, we introduce a frequency loss by applying Fourier Transform (FT) to both the output and ground-truth images and computing the L_1 loss between their real and imaginary components.

$$Loss = \|J - GT\|_1 + \lambda \cdot \left[\|J_{real} - GT_{real}\|_1 + \left\|J_{imag} - GT_{imag}\right\|_1 \right] \tag{13}$$

where J denotes the output remote sensing image after dehazing by DWTMA-Net; GT refers to its corresponding ground-truth counterpart; J_{real} represents the real part of the generated image; GT_{real} represents the real part of the ground-truth image; J_{imag} represents the imaginary part of the generated image; GT_{imag} represents the imaginary part of the ground-truth image; and λ represents the weight, which is 0.1.

4. Results

4.1. Datasets

We evaluate the performance of the proposed DWTMA-Net on two synthetic remote sensing (RS) haze datasets—SateHaze1k [30] and HRSD [31]—as well as one real-world UAV-based hazy dataset, HazyDet [32]. SateHaze1k is divided into three subsets corresponding to different haze densities: thin, moderate, and thick. Each subset includes 320 training samples and 45 testing images. Thin haze scenes are generated using haze masks extracted from real cloud formations, while moderate haze images blend characteristics of mist and medium haze. Thick haze is simulated using transmittance maps to represent dense atmospheric conditions.

The HRSD dataset consists of two subsets: LHID and DHID. LHID contains 30,517 training images and 500 test images, generated through the atmospheric scattering model to simulate varying levels of haze, thereby improving the model's robustness across different haze intensities. In comparison, DHID includes 14,990 images that are synthesized using real haze maps, offering a more authentic representation of haze characteristics. Among these, 14,490 images are designated for training, while 500 are reserved for testing. The inclusion of both synthetic and real haze features in these subsets provides a comprehensive platform for evaluating the dehazing capabilities of DWTMA-Net.

In order to showcase the generalization ability of our model in practical settings, we assess its performance using the newly introduced HazyDet dataset, which consists of

drone-captured images affected by haze. The dataset consists of a training set with 8000 images, a validation set with 1000 images, and a test set with 2000 images. It contains a mix of authentic hazy images captured under natural fog conditions, as well as artificially created hazy images generated through the Atmospheric Scattering Model (ASM). Additionally, HazyDet features a dedicated real hazy drone detection test set (RDDTS) designed to assess model robustness in practical scenarios. Figure 4 provides examples of training samples from the dataset.



Figure 4. Example of training samples from the HazyDet dataset.

4.2. Implementation Details

Our DWTMA-Net was trained and tested using the PyTorch framework (version 1.13.1) on a system equipped with four NVIDIA GeForce GTX 1080 Ti GPUs. To enhance the training data, we applied random rotations of 90°, 180°, and 270°, as well as horizontal flipping. The input images were RGB remote sensing data resized to 240 × 240 pixels. For the FFEB, we set the configuration as [N1, N2, N3, N4, N5] = [2, 2, 4, 2, 2], with respective embedding channels of [24, 48, 96, 48, 24]. The batch size was set to 16 for each sub-dataset. The initial learning rate was initialized at 2.0×10^{-4} and reduced progressively to zero using a cosine annealing scheduler.

We assessed the generalization capacity and performance of DWTMA-Net through comprehensive comparisons across various tasks. To maintain consistency, we utilized the official codebases released by the respective authors during the training process.

4.3. Quantitative Evaluations

The quantitative evaluation results on the HRSD and SateHaze1k datasets are shown in Tables 1 and 2, with performance assessed using PSNR and SSIM metrics, as well as the average results across various haze density levels. Additionally, Table 3 presents the results on the HazyDet dataset, further highlighting the model's ability to generalize for UAV-based image dehazing tasks.

Table 1. Comparative analysis on the **SateHaze1k** dataset, where **bold** indicates the optimal method and underline signifies the second best.

Method	Thin Haze			Moderate Haze			Thick Haze			Average		
	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE
DCP [34]	20.15	0.8645	17.98	20.51	0.8932	17.09	15.77	0.7117	17.73	18.81	0.8241	17.60
AOD-Net [39]	15.97	0.8169	18.66	15.39	0.7442	17.28	14.44	0.7013	17.91	15.27	0.7541	17.95
FCTF-Net [44]	19.13	0.8532	18.77	22.32	0.9107	17.75	17.78	0.7617	18.14	19.74	0.8419	18.22
GridDehaze-Net [20]	19.81	0.8556	18.77	22.75	0.9085	16.35	17.94	0.7551	18.69	20.17	0.8397	17.94
FFA-Net [22]	24.04	0.9130	17.09	25.62	0.9336	16.80	21.70	0.8422	17.35	23.79	0.8963	17.08
MixDehaze-Net [40]	22.12	0.8822	18.04	23.92	0.9040	**16.08**	19.96	0.7950	17.94	22.00	0.8604	17.35
OK-Net [41]	20.68	0.8860	17.78	25.39	0.9406	17.47	20.21	0.8186	18.57	22.09	0.8817	17.94
Dehazeformer [24]	24.90	0.9104	16.88	27.13	0.9431	16.70	22.68	0.8497	17.64	24.90	0.9011	17.07
VmambaIR [26]	20.81	0.8753	18.28	24.34	0.9132	16.61	20.04	0.8045	17.96	21.73	0.8643	17.62
FCDM [27]	18.94	0.8486	18.08	17.36	0.8753	16.81	16.97	0.7530	18.09	17.76	0.8256	17.66
MMPD-Net [45]	25.16	0.9227	16.76	27.30	0.9454	16.76	22.85	0.8571	17.96	25.10	0.9084	17.16
DWTMA-Net	**25.59**	**0.9229**	**16.71**	**27.53**	**0.9459**	16.80	**22.88**	**0.8576**	**17.33**	**25.33**	**0.9088**	**16.95**

Table 2. Comparative analysis on the **HRSD** dataset, where **bold** indicates the optimal method and underline signifies the second best.

Method	LHID			DHID			Average		
	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE
DCP [34]	21.34	0.7976	19.84	19.15	0.8195	18.99	20.25	0.8086	19.42
AOD-Net [39]	21.91	0.8144	19.53	16.03	0.7291	18.91	18.97	0.7718	19.22
FCTF-Net [44]	28.55	0.8727	19.27	22.43	0.8482	18.61	25.49	0.8605	18.94
GridDehaze-Net [20]	25.80	0.8584	19.54	26.77	0.8851	18.93	26.29	0.8718	19.24
FFA-Net [22]	29.33	0.8755	19.02	24.62	0.8657	18.50	26.98	0.8706	18.76
MixDehaze-Net [40]	29.47	0.8631	19.26	27.36	0.8864	18.89	28.42	0.8748	19.08
OK-Net [41]	29.03	0.8766	18.73	27.80	0.8973	18.17	28.42	0.8870	18.45
MMPD-Net [45]	29.76	0.8771	18.98	28.23	0.8977	18.29	29.00	0.8874	18.64
FCDM [27]	15.16	0.6459	18.79	17.13	0.6978	20.43	16.15	0.6719	19.61
DWTMA-Net	**29.86**	**0.8828**	**18.70**	**28.34**	**0.8981**	**18.15**	**29.10**	**0.8905**	**18.43**

Table 3. Comparative analysis on the **HazyDet** dataset, where **bold** indicates the optimal method and underline signifies the second best.

Method	HazyDet		
	PSNR	SSIM	NIQE
DCP [34]	17.03	0.8024	12.30
AOD-Net [39]	18.99	0.7808	12.27
FCTF-Net [44]	24.89	0.8552	12.23
GridDehaze-Net [20]	26.66	0.8801	11.33
FFA-Net [22]	27.12	0.8782	11.31
MixDehaze-Net [40]	28.75	0.9068	11.30
OK-Net [41]	27.76	0.8875	11.36
DWTMA-Net	**29.02**	**0.9108**	**11.23**

As shown in Table 1, our proposed DWTMA-Net achieves significant PSNR and SSIM improvements across different haze levels, including light, moderate, and dense conditions. Since dense haze severely degrades image quality, presenting particularly difficult restoration challenges for remote sensing data, our method exhibits marginally lower performance in such conditions. Traditional methods such as DCP and AOD-Net perform poorly overall, while FFA-Net, MixDehaze-Net, OK-Net, and Dehazeformer demonstrate moderate performance. MMPD-Net achieves relatively strong results. Nevertheless, DWTMA-Net

still outperforms other models on several key metrics. Notably, our method shows excellent performance, with average PSNR and average SSIM gains of 0.23 and 0.0004 higher than the second-ranked MMPD-Net, respectively. These findings confirm DWTMA-Net's effectiveness and robustness in addressing remote sensing image dehazing challenges.

Table 2 demonstrates DWTMA-Net's superior performance on two benchmark synthetic datasets (LHID and DHID). On LHID, it outperforms MMPD-Net (previous state of the art), with gains of 0.1 dB in PSNR and 0.0057 in SSIM. On the DHID dataset, PSNR is increased by 0.11 dB and SSIM by 0.0004. It is worth noting that MMPD-Net ranks second across all metrics, with results close to ours, which may be attributed to its use of multi-scale convolutions and feature dimensionality expansion. Overall, these results demonstrate the strong dehazing capability and robust performance of DWTMA-Net in remote sensing image dehazing tasks.

To demonstrate that the model not only performs well on specific datasets but also adapts to more complex and diverse data from real-world applications, showcasing its strong adaptability, robustness, and generalization ability, we added a UAV-based image dataset for comparison. As shown in Table 3, our model achieves significantly superior performance across all metrics. Significant quality improvements are observed, with our model exceeding the second-best method's performance by 0.27 dB PSNR and 0.0233 SSIM.

We use two key metrics to evaluate the computational efficiency and memory requirements of the proposed model: FLOPs and the number of parameters. FLOPs represent the number of floating-point operations required for a single forward pass, reflecting the computational cost. The number of parameters indicates the total trainable weights, reflecting memory usage. Fewer FLOPs and fewer parameters make the model more suitable for deployment in resource-constrained or real-time scenarios, as shown in Table 4. Our SAOD is a simplified version of AOD, with 170.66M FLOPs and 684B parameters.

Table 4. Comparison of **FLOPs** and **parameters** across models.

Method	FLOPs	Parameters
DCP	-	-
AOD-Net	457.70 (M)	1.76 (K)
FCTF-Net	40.19 (G)	163.48 (K)
GridDehaze-Net	85.72 (G)	955.75 (K)
FFA-Net	624.20 (G)	4.68 (M)
MixDehaze-Net	114.30 (G)	3.17 (M)
OK-Net	158.20 (G)	4.43 (M)
MMPD-Net	298.19 (G)	8.66 (M)
DWTMA-Net	188.72 (G)	8.34 (M)

4.4. Qualitative Evaluations

The experiments utilize three benchmark datasets covering satellite (SateHaze1k), surface (HRSD), and aerial (HazyDet) hazy scenarios.

A performance comparison of various dehazing approaches on the light-haze test set is presented in Figure 5. The figure reveals that DCP and AOD-Net demonstrate constrained dehazing capability, leaving substantial haze remnants and apparent chromatic aberrations in output images. GridDehaze-Net and OK-Net are able to remove most of the haze, but a small amount still lingers. Although FFA-Net and MixDehaze-Net achieve better dehazing results, they fall short in color restoration. In particular, the areas highlighted by red boxes fail to accurately reproduce the colors of the reference images. In contrast, both MMPD-Net and DWTMA-Net demonstrate strong dehazing capabilities and produce results that closely resemble the ground truth. Specifically, in the red-box regions, DWTMA-Net achieves more accurate color recovery, whereas MMPD-Net tends to render the grass and

trees with slightly lighter tones compared to the reference image. The quantitative metrics displayed below the images further highlight the differences in performance among the methods. Overall, our method outperforms the others in terms of dehazing effectiveness, color fidelity, and frequency information restoration, resulting in images that are more visually aligned with real-world scenes.

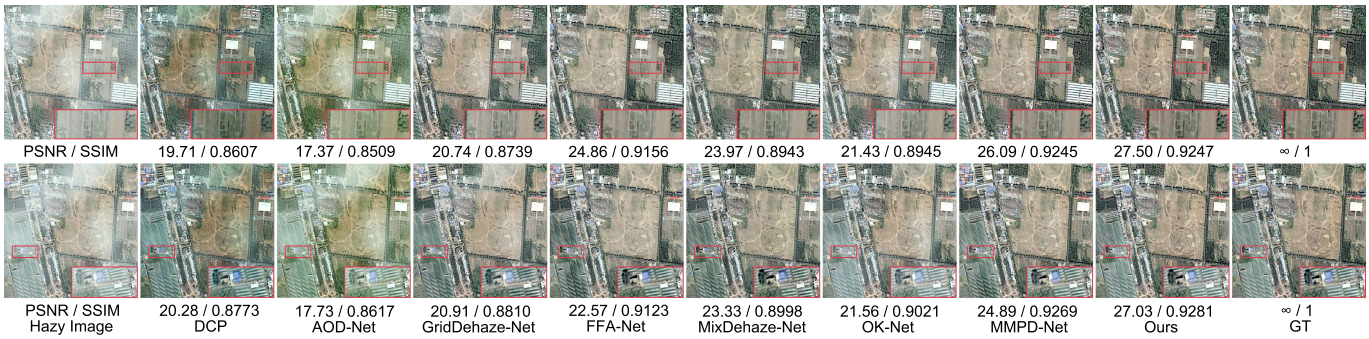


Figure 5. Qualitative analysis of two lightly hazy samples from the Haze1k-thin collection.

Figure 6 evaluates multiple dehazing approaches using moderately hazy remote sensing imagery, where atmospheric interference significantly degrades image features. While DCP fails to adequately remove haze residues, AOD-Net partially restores visibility but introduces undesirable darkening effects. In contrast, GridDehaze-Net, FFA-Net, MixDehaze-Net, OK-Net, and MMPD-Net show substantially improved restoration quality. However, noticeable differences from the ground truth still exist. For instance, in the second image, the region highlighted by the red box in the result from GridDehaze-Net appears blurred and faded, and the high-frequency details in the first image are not well restored. In contrast, FFA-Net, MixDehaze-Net, OK-Net, and MMPD-Net deepen the high-frequency features in the red-box region of the first image but show varying degrees of color distortion in the red-box region of the second image. According to the performance metrics displayed below each image, our method demonstrates the best dehazing performance. Overall, our approach exhibits superior dehazing capability and more accurate color restoration in high-frequency regions, which is largely attributed to the incorporation of our frequency information enhancement module.

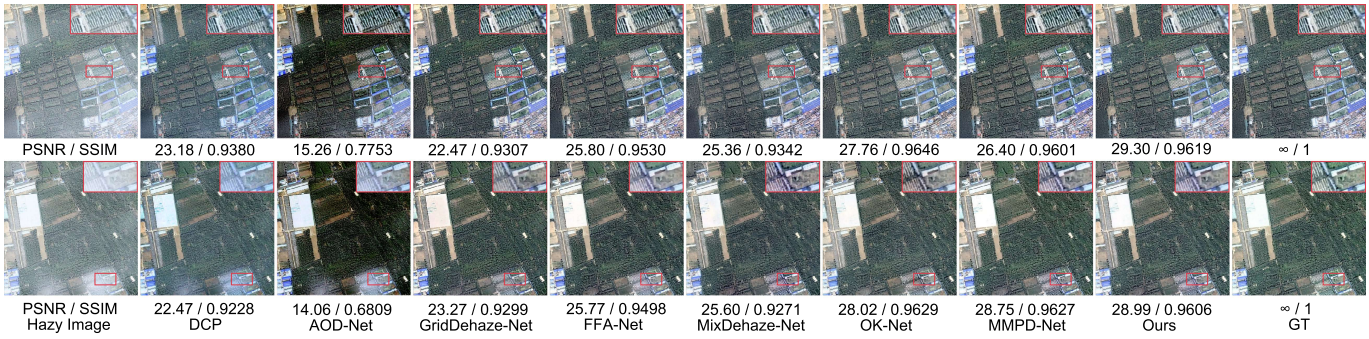


Figure 6. Qualitative analysis of two moderately hazy samples from the Haze1k-moderate collection.

Figure 7 compares restoration results across different approaches using severely degraded images from the dense-haze dataset. Due to the severe impact of dense haze, a significant amount of frequency information and fine details is lost in this dataset. DCP and AOD-Net fail to completely eliminate atmospheric interference, leaving substantial haze contamination and introducing significant chromatic aberrations in the processed images. GridDehaze-Net and OK-Net introduce noticeable color shifts in their outputs; for example, compared to the reference image, the green lawns in the two restored images

appear either overly darkened or overly lightened. Color fidelity analysis reveals that FFA-Net, MMPD-Net, and MixDehaze-Net all introduce chromatic aberrations, rendering lawns in un-naturally pale or oversaturated green tones. In contrast, our method demonstrates consistently robust performance across all dehazing indicators in both images, producing results that closely resemble the ground truth in terms of overall structure, frequency details, and spatial information.

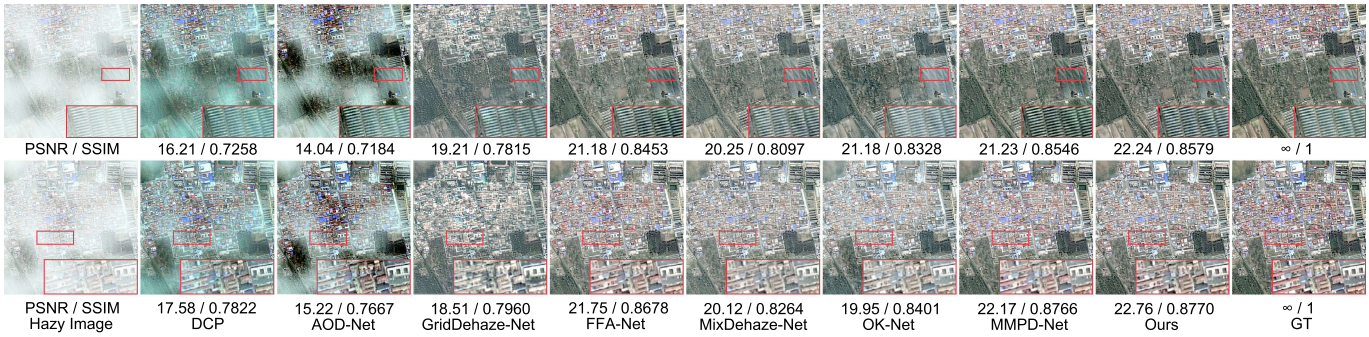


Figure 7. Qualitative analysis of two densely hazy samples from the Haze1k-thick collection.

Figure 8 compares the performance of multiple dehazing algorithms on the LHID dataset. DCP's output displays excessive color saturation, resulting in critical detail loss, whereas AOD-Net generates underexposed reconstructions that degrade visual clarity. Although GridDehaze-Net, FFA-Net, MixDehaze-Net, OK-Net, and MMPD-Net achieve perceptually reasonable results, their PSNR/SSIM scores indicate substantial deviations from reference data. The proposed DWTMA-Net outperforms these approaches by effectively restoring haze-obscured high-frequency components, delivering superior sharpness and enhanced image quality.

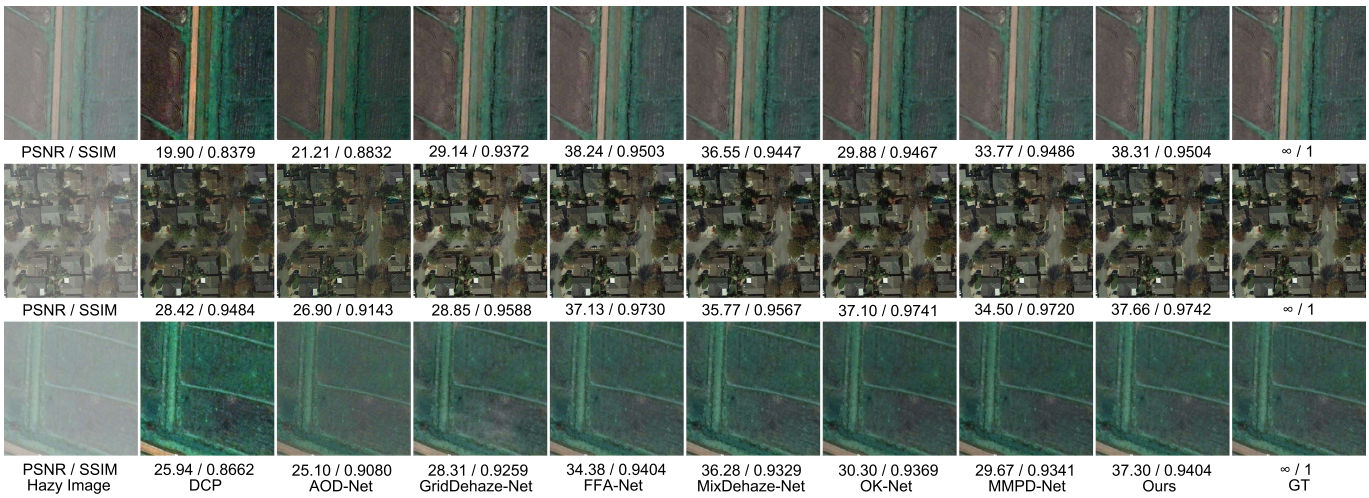


Figure 8. Qualitative analysis of three samples from the LHID collection.

Figure 9 benchmarks dehazing performance on the DHID dataset, featuring uniformly dense haze in remote sensing imagery. DCP's reconstructions exhibit severe luminance suppression and detail loss, failing to recover critical high-frequency components. AOD-Net produces even darker outputs, with a noticeable black mask overlaying the images. Although GridDehaze-Net, MixDehaze-Net, OK-Net, FFA-Net, and MMPD-Net are capable of effectively removing haze and largely restoring the overall scene, their outputs still exhibit subtle black artifacts in fine-detail regions when compared to the reference images. In contrast, our proposed DWTMA-Net demonstrates superior performance in both color

accuracy and high-frequency detail restoration, particularly in areas such as roads and rooftops, as further confirmed by the quantitative metrics shown below the images.

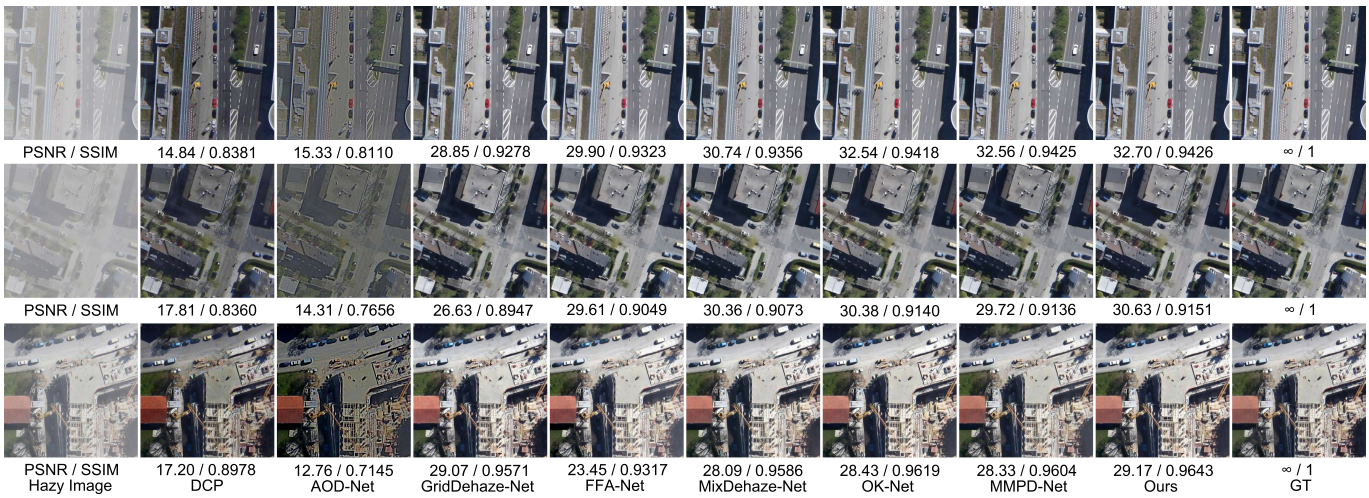


Figure 9. Qualitative analysis of three samples from the DHID collection.

Figure 10 benchmarks dehazing performance across UAV imagery under varying atmospheric conditions, from light to dense haze. In these examples, light haze slightly obscures critical information in UAV images, affecting recognition and tracking accuracy. Moderate haze interferes with the identification of certain regions of the UAV, while heavy haze severely hampers information extraction, significantly impacting the UAV's operational capabilities. Therefore, effective haze removal from UAV images is of great importance.

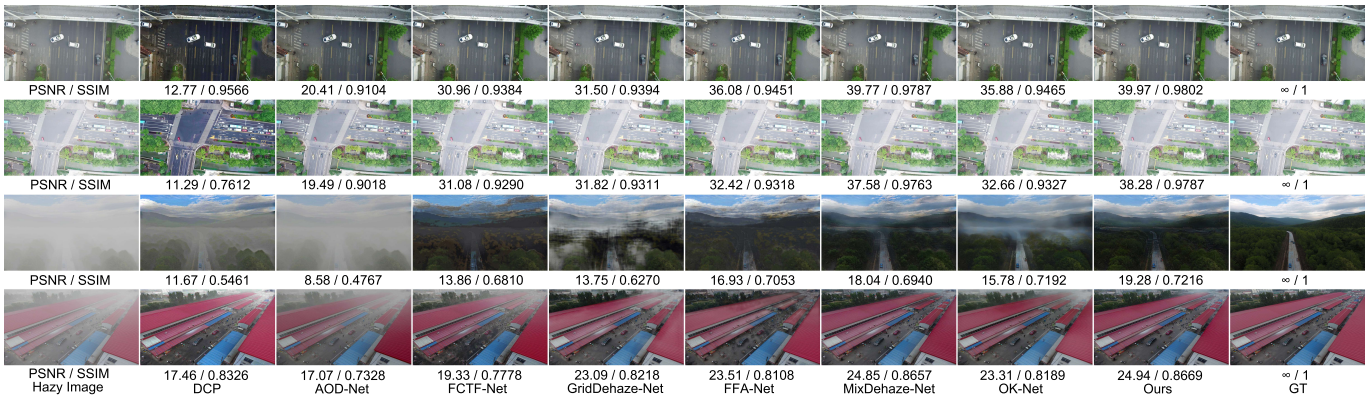


Figure 10. Qualitative analysis of four samples from the HazyDet collection.

As for the results, DCP performs poorly, introducing noticeable color distortion and leaving substantial residual haze in the third and fourth images. AOD-Net shows limited performance under light haze and fails to remove haze effectively in the third and fourth images. GridDehaze-Net and OK-Net perform relatively well in the first and second images but leave obvious haze residues in the third and fourth images, indicating incomplete dehazing. Although FCTF-Net and FFA-Net manage to remove haze in the third image, the resulting images become heavily blurred, making it difficult to recognize objects. MixDehaze-Net successfully removes a large portion of the haze and preserves the original features, yet a small amount of haze remains in the third image.

In contrast, our proposed method achieves the best dehazing performance across all haze levels. It nearly restores all image details under light haze, effectively recovers obscured information under moderate haze, and successfully reconstructs images under

heavy haze conditions, producing visually impressive results. These outcomes highlight the strong generalization ability and robustness of our method across various haze intensities.

To demonstrate the dehazing performance of our network in real-world scenarios, we conducted tests on a real hazy remote sensing dataset provided in [46]. As shown in Figure 11, our method effectively removes haze while preserving edge and texture details. This test effectively evaluates the robustness and dehazing capability of our method in practical environments.



Figure 11. Visual comparison of four images within the real dataset.

4.5. Ablation Study

To validate the contributions of our proposed components, we performed a comprehensive ablation analysis using the light-haze dataset, evaluating the individual and combined effects of three key modules: the wavelet downsampling module (WDM), Discrete Wavelet Block (DWB), and multi-scale attention block (MAB). For computational efficiency during training, we processed 80 × 80 pixel image patches sampled from the original RGB inputs while maintaining hyperparameters and training protocols identical to those of our complete model implementation.

Our baseline architecture builds upon the fundamental structure of Star [43], incorporating its core modules and basic attention mechanisms as the foundational learning blocks. In our modified design, the original DM module is replaced with the proposed WDM, and the DWB module is removed. Table 5 quantitatively evaluates the individual contributions of all proposed components (WDM, DWB, and MAB), with each module showing statistically significant performance improvements that confirm their design efficacy. Figure 12 visually compares the ablation results of different modules, providing a clearer and more intuitive illustration of each module's role and contribution in enhancing the overall network performance.

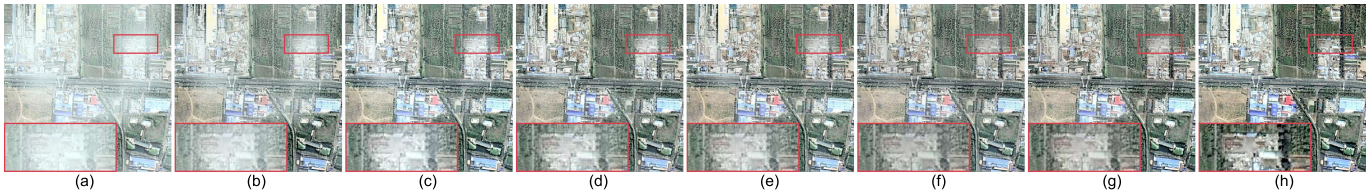


Figure 12. Visual comparison of ablation experiments on the thin dataset. (**a**) Hazy image; (**b**) Base(DM); (**c**) Base(DM) + DWB; (**d**) Base(DM) + MAB; (**e**) Base(DM) + DWB + MAB; (**f**) Base + MAB + WDM; (**g**) Base + DWB + MAB + WDM; (**h**) ground truth.

When the model is in the base state with standard downsampling, the image quality is the poorest, with the lowest PSNR and SSIM values. The red-box regions are heavily obscured by haze, and the overall image appears noticeably pale. Introducing the DWB module to the base network significantly improves image quality, increasing the PSNR by 2.06 and SSIM by 0.0349. Similarly, integrating the MAB module brings substantial enhancement, with a PSNR increase of 3.37 and an SSIM improvement of 0.0291, resulting in clearer details in the red-box areas. Further combining DWB and MAB in the FFEB module leads to additional gains in performance, with improved color restoration and a notable enhancement in overall visual quality. Although minor blurring and slight whitening still

remain, FFEB effectively recovers both frequency and spatial information. The contribution of DWB is clearly demonstrated in quantitative form, as the PSNR and SSIM values drop significantly when this module is removed. To objectively assess the WDM's superiority, we replaced standard downsampling layers with our wavelet-based module in identical network architectures. The results show a further PSNR gain of 0.82 and an SSIM increase of 0.0103, with significant visual enhancement in the red-box areas. The resulting images are noticeably sharper, demonstrating the effectiveness of the WDM in extracting frequency and spatial features and improving overall image quality.

Table 5. The ablation experiments performed on the thin-haze subset reveal comparative method effectiveness, with optimal results highlighted in boldface.

Method	Thin Haze	
	PSNR	SSIM
Base(DM)	18.47	0.8547
Base(DM) + DWB	20.53	0.8896
Base(DM) + MAB	21.84	0.8838
Base(DM) + DWB + MAB	22.87	0.8940
Base + MAB + WDM	21.64	0.8860
Base + DWB + MAB + WDM	**23.69**	**0.9043**

The MAB module contains different attention mechanisms. To analyze their impact, we conducted ablation experiments, as shown in Table 6.

Table 6. The ablation experiments conducted on the light-haze subset reveal the impacts of different attention mechanisms in the MAB.

Method	Thin Haze	
	PSNR	SSIM
DWTMA-Net - CA	23.16	0.8966
DWTMA-Net - PA	20.04	0.8905
DWTMA-Net - FA	22.66	0.8910

5. Discussion

This paper proposes a network that combines frequency-domain and spatial-domain processing to address the issues of blurring and information loss in remote sensing images. Experimental validation on aerial imagery (UAV dataset) confirms the model's strong transferability, with quantitatively significant haze removal results.

Although the proposed model achieves significantly better dehazing performance compared to existing lightweight methods, it also incurs a relatively higher computational cost. It is particularly well-suited for applications that demand high image clarity and detail preservation, such as remote sensing under adverse weather conditions or UAV-based surveillance. However, to enable efficient deployment on resource-constrained edge devices, further model optimization is necessary. Future work will focus on techniques such as network pruning, quantization, and knowledge distillation to develop a more lightweight and efficient variant. By seeking a balanced trade-off between performance and complexity, this study lays a solid foundation for both practical deployment and future scalability of the model.

6. Conclusions

DWTMA-Net consists of a series of frequency feature extraction modules designed to simultaneously capture both frequency and spatial information. We designed an innovative

downsampling method that combines Haar discrete wavelet transform to extract frequency-domain features and convolution operations to capture spatial-domain features. The extracted features are then processed separately in the frequency and spatial domains. The discrete wavelet block handles the frequency-domain information, decomposing the features into four sub-band frequency characteristics using wavelet transform. Different recovery and refinement strategies are applied to each sub-band. Next, we apply a multi-dimensional attention mechanism to enhance the spatial-domain features. This mechanism extracts fine details through deep convolution layers and captures diverse features by expanding the number of channels. These diverse features are further optimized using channel attention, pixel attention, and Fourier frequency-domain attention, enhancing both local and global information, which improves image quality and strengthens the network's robustness. Experimental results show that our method achieves excellent performance on the SateHaze-1K, HRSD, and HazyDet datasets, effectively recovering image details in complex environments. However, under heavy haze conditions, the network still faces challenges in information recovery. To advance this research direction, two key objectives will be pursued: (1) the development of an enhanced lightweight architecture specifically optimized for remote sensing image dehazing under challenging conditions and (2) the construction of a comprehensive benchmark dataset addressing top-down imaging artifacts to facilitate community-wide progress.

Author Contributions: Methodology, L.W.; Validation, Y.L.; Resources, R.H.; Writing—original draft, X.G.; Writing—review & editing, H.Z.; Supervision, H.X. 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) and the Fundamental Research Funds for the Central Universities of China (SWU2009107).

Data Availability Statement: The StateHaze1k dataset, HRSD dataset, and HazyDet dataset are publicly available for research use only. For more information, please refer to the following links: StateHaze1k https://www.dropbox.com/s/k2i3p7puuwl2g59/Haze1k.zip?dl=0 (accessed on 7 June 2025), HRSD https://github.com/Shan-rs/DCI-Net (accessed on 7 June 2025), and HazyDet https://github.com/GrokCV/HazyDet (accessed on 7 June 2025).

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

References

1. 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]
2. 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]
3. 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]
4. Yan, Q.; Yang, K.; Hu, T.; Chen, G.; Dai, K.; Wu, P.; Ren, W.; Zhang, Y. From dynamic to static: Stepwisely generate HDR image for ghost removal. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 1409–1421. [CrossRef]
5. McCartney, E. *Optics of the Atmosphere: Scattering by Molecules and Particles*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1976.
6. Yan, Q.; Zhang, L.; Liu, Y.; Zhu, Y.; Sun, J.; Shi, Q.; Zhang, Y. Deep HDR imaging via a non-local network. *IEEE Trans. Image Process.* **2020**, *29*, 4308–4322. [CrossRef]
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, 3–8 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. Yan, Q.; Wang, H.; Ma, Y.; Liu, Y.; Dong, W.; Woźniak, M.; Zhang, Y. Uncertainty estimation in HDR imaging with Bayesian neural networks. *Pattern Recognit.* **2024**, *156*, 110802. [CrossRef]
11. Zhou, H.; Chen, Z.; Liu, Y.; Sheng, Y.; Ren, W.; Xiong, H. Physical-priors-guided DehazeFormer. *Knowl.-Based Syst.* **2023**, *266*, 110410. [CrossRef]
12. Liu, Y.; Wang, X.; Hu, E.; Wang, A.; Shiri, B.; Lin, W. VNDHR: Variational single nighttime image dehazing for enhancing visibility in intelligent transportation systems via hybrid regularization. *IEEE Trans. Intell. Transp. Syst.* 2025, *early access*.
13. Liu, Y.; Yan, Z.; Tan, J.; Li, Y. Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *33*, 1643–1657. [CrossRef]
14. 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 31st ACM International Conference on Multimedia, Ottawa, ON, Canada, 29 October–3 November 2023; pp. 4119–4128.
15. Li, C.; Hu, E.; Zhang, X.; Zhou, H.; Xiong, H.; Liu, Y. Visibility restoration for real-world hazy images via improved physical model and Gaussian total variation. *Front. Comput. Sci.* **2024**, *18*, 181708. [CrossRef]
16. Li, T.; Liu, Y.; Ren, W.; Shiri, B.; Lin, W. Single Image Dehazing Using Fuzzy Region Segmentation and Haze Density Decomposition. *IEEE Trans. Circuits Syst. Video Technol.* 2025, *early access*.
17. Chen, G.; Jia, Y.; Yin, Y.; Fu, S.; Liu, D.; Wang, T. Remote sensing image dehazing using a wavelet-based generative adversarial networks. *Sci. Rep.* **2025**, *15*, 3634. [CrossRef] [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. 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]
20. 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.
21. 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, Seattle, WA, USA, 13–19 June 2020; pp. 2157–2167.
22. 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.
23. 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.
24. Song, Y.; He, Z.; Qian, H.; Du, X. Vision transformers for single image dehazing. *IEEE Trans. Image Process.* **2023**, *32*, 1927–1941. [CrossRef]
25. Nie, J.; Xie, J.; Sun, H. Remote Sensing Image Dehazing via a Local Context-Enriched Transformer. *Remote Sens.* **2024**, *16*, 1422. [CrossRef]
26. 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. [CrossRef]
27. Wang, J.; Wu, S.; Yuan, Z.; Tong, Q.; Xu, K. Frequency compensated diffusion model for real-scene dehazing. *Neural Netw.* **2024**, *175*, 106281. [CrossRef]
28. Huang, Y.; Xiong, S. Remote sensing image dehazing using adaptive region-based diffusion models. *IEEE Geosci. Remote. Sens. Lett.* **2023**, *20*, 8001805. [CrossRef]
29. Yan, Q.; Hu, T.; Wu, P.; Dai, D.; Gu, S.; Dong, W.; Zhang, Y. Efficient Image Enhancement with A Diffusion-Based Frequency Prior. *IEEE Trans. Circuits Syst. Video Technol.* 2025, *early access*.
30. 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.
31. Zhang, L.; Wang, S. Dense haze removal based on dynamic collaborative inference learning for remote sensing images. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 5631016. [CrossRef]
32. Feng, C.; Chen, Z.; Kou, R.; Gao, G.; Wang, C.; Li, X.; Shu, X.; Dai, Y.; Fu, Q.; Yang, J. HazyDet: Open-source Benchmark for Drone-view Object Detection with Depth-cues in Hazy Scenes. *arXiv* **2024**, arXiv:2409.19833.
33. 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; pp. 1–8.
34. 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]
35. Fattal, R. Dehazing using color-lines. *Acm Trans. Graph. (TOG)* **2014**, *34*, 1–14. [CrossRef]

36. 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.
37. 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.
38. 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.
39. 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.
40. 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; pp. 1–10.
41. Cui, Y.; Ren, W.; Knoll, A. Omni-Kernel Network for Image Restoration. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, 20–27 February 2024; Volume 38, pp. 1426–1434.
42. Sui, T.; Xiang, G.; Chen, F.; Li, Y.; Tao, X.; Zhou, J.; Hong, J.; Qiu, Z. U-Shaped Dual Attention Vision Mamba Network for Satellite Remote Sensing Single-Image Dehazing. *Remote Sens.* **2025**, *17*, 1055. [CrossRef]
43. 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, 16–22 June 2024; pp. 5694–5703.
44. 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]
45. 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. [CrossRef]
46. 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]

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.