Article

Unpaired Remote Sensing Image Dehazing Using Enhanced Skip Attention-Based Generative Adversarial Networks with Rotation Invariance

Yitong Zheng, Jia Su *, Shun Zhang , **Mingliang Tao** and **Ling Wang**

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China;
2016100530@mail.nwpu.edu.cn (Y.Z.); szhang@nwpu.edu.cn (S.Z.); mltao@nwpu.edu.cn (M.T.);
lingwang@nwpu.edu.cn (L.W.)
* Correspondence: jiasu1011@126.com

Abstract: Remote sensing image dehazing aims to enhance the visibility of hazy images and improve the quality of remote sensing imagery, which is essential for various applications such as object detection and classification. However, the lack of paired data in remote sensing image dehazing enhances the applications of unpaired image-to-image translation methods. Nonetheless, the considerable parameter size of such methods often leads to prolonged training times and substantial resource consumption. In this work, we propose SPRGAN, a novel approach leveraging Enhanced Perlin Noise-Based Generative Adversarial Networks (GANs) with Rotation Invariance to address these challenges. Firstly, we introduce a Spatial-Spectrum Attention (SSA) mechanism with Skip-Attention (SKIPAT) to enhance the model's ability to interpret and process spectral information in hazy images. Additionally, we have significantly reduced computational overhead to streamline processing. Secondly, our approach combines Perlin Noise Masks in pre-training to simulate real foggy conditions, thereby accelerating convergence and enhancing performance. Then, we introduce a Rotation Loss (RT Loss) to ensure the model's ability to dehaze images from different angles uniformly, thus enhancing its robustness and adaptability to diverse scenarios. At last, experimental results demonstrate the effectiveness of SPRGAN in remote sensing image dehazing, achieving better performance compared to state-of-the-art methods.

Keywords: dehaze; GANs; Perlin Noise; rotation loss; transformer-guided

1. Introduction

In optical remote sensing, the negative influence of fog often compromises the quality of the image content, limiting its applicability across diverse applications, especially object detection. In order to guarantee optimal performance and processing efficacy in these applications, it is of the utmost importance to develop algorithms that are capable of mitigating the effects of haze [1]. Therefore, remote sensing image haze removal is a crucial and indispensable pre-processing task. This prior process is commonly referred to as image dehazing, a technology that can provide clear, haze-free images, particularly in the context of remote sensing for ship detection. Specifically for ship detection and earth observation tasks, dehazing techniques help to improve the sharpness and contrast of images, thereby enhancing the performance of detection algorithms. Advances in image dehazing techniques greatly enhance the potential for extracting usable information from remote sensing data, particularly in challenging atmospheric conditions, as shown in Figure 1.
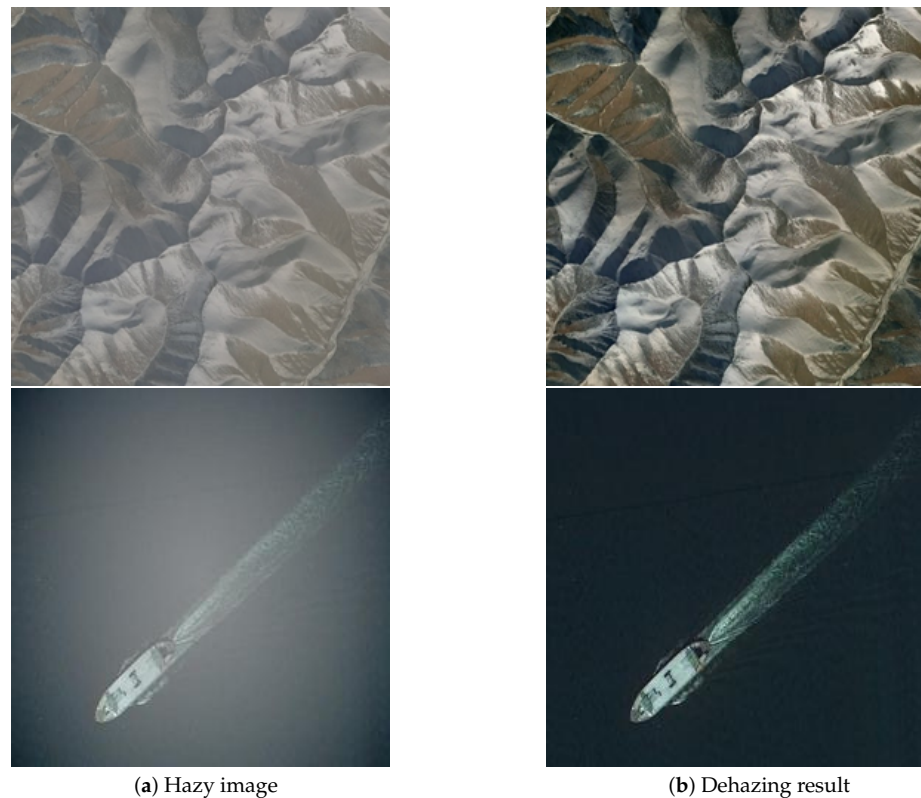
(**a**) Hazy image   (**b**) Dehazing result

Figure 1. Dehazing results on remote sensing hazy images. The hazy images are on the left and the dehazed images are on the right of the figure.

In recent years, significant progress has been made in the field of image dehazing, driven by the development of advanced machine learning and deep learning techniques. Prior information-based methods, such as histogram equalization [2] and contrast enhancement [3], have proven inadequate for handling complex haze conditions and diverse remote sensing environments. These conventional approaches often fail to preserve important image details and can introduce artifacts, leading to poor results in critical applications such as object detection and environmental monitoring. One of the most well-known prior-based techniques is the Dark Channel Prior (DCP) [4], which assumes that in most haze-free patches of an image, at least one color channel has some pixels with very low intensities. This assumption helps to estimate the thickness of the haze and subsequently remove it from the image. Another popular method is the Color Attenuation Prior (CAP) [5], which exploits the difference in attenuation among different color channels to estimate the depth information and remove haze effectively.

With the advent of deep learning, many neural network-based dehazing models have been proposed that offer significant improvements over traditional methods. Convolutional Neural Networks (CNNs) [6] have shown a remarkable ability to learn complex features from blurred images and restore clear images. However, many of these models are designed for paired image datasets, where hazy and clear image pairs are required for training. This requirement poses a significant challenge in remote sensing, where the acquisition of such paired datasets is often impractical.

To address the limitations of paired image dehazing models, unpaired image-to-image translation models such as CycleGAN [7] have emerged as a promising solution. Another notable model is Dehaze-AGGAN [8], which introduces an attention-guided generative adversarial network to better focus on relevant features in hazy images. UVCGAN [9] further enhances the dehazing process by incorporating a unified variational framework. These models use adversarial training to learn mappings between different image domains without the need for paired samples. This flexibility is particularly advantageous in remote sensing, where the acquisition of paired hazy and clear image datasets can be challenging.

However, despite their potential, CycleGAN and its variants have a significant computational cost. For better performance, these models often require large parameter sizes, leading to prolonged training times and increased resource consumption. The increase in computational requirements affects the efficiency of image dehazing algorithms, particularly in satellite image processing, where computational resources are limited. Moreover, while CycleGAN can produce visually convincing results, it often struggles with retaining fine details and handling varying haze densities effectively. The model may not consistently capture subtle features and textures, leading to the potential loss of important information in the dehazed images. This is particularly problematic in remote sensing applications, where the clarity and precision of image details are crucial for accurate analysis and decision-making.

Our proposed model, SPRGAN, builds on the strengths of unpaired image translation models while introducing novel enhancements to further improve performance. A key innovation in SPRGAN is the integration of the Spatial-Spectrum Attention (SSA) mechanism with Skip-Attention (SKIPAT). This combination allows the model to effectively capture both spatial and spectral features of hazy images. The SSA mechanism focuses on essential regions in the frequency domain, ensuring accurate haze removal and detail preservation. Meanwhile, the SKIPAT mechanism reduces the computational load while maintaining high model accuracy, improving efficiency without compromising performance. Another important aspect of our approach is the incorporation of Perlin Noise-Based Masks (PNM). These masks simulate realistic hazy conditions and improve the model's ability to generalize across different hazy scenarios. The introduction of PNM during the pre-training phase not only accelerates convergence speed but also improves overall performance. Additionally, SPRGAN features Rotation Loss (RT Loss) within its Transformer architecture. This innovative loss function ensures consistency between dehazed images and their rotated counterparts, further refining the model's capability to produce high-quality, haze-free images. RT Loss promotes the stability and reliability of the model, making it more adept at handling diverse and complex haze patterns.

The main contributions of this paper are summarized as follows:

1. **Proposing SPRGAN with Skip-Attention:** The research introduces an advanced model, SPRGAN, which incorporates a Spatial-Spectrum Attention (SSA) mechanism with Skip-Attention (SKIPAT). The skip-attention mechanism reduces the calculation complexity of the model, while the SSA mechanism enhances the model's ability to interpret and process spectral information in hazy images.

2. **Proposing of PNM for model pre-training:** this research introduces a novel approach by incorporating the Perlin Noise Mask (PNM) pre-training method during model pre-training, which effectively simulates hazy conditions, empowering the model to concurrently strengthen its super-resolution and dehazing capabilities.

3. **Integration of RT Loss within the Transformer Architecture:** the incorporation of RT Loss into the Transformer architecture, which enhances the core objectives of the SPRGAN model, is a pioneering aspect of this research, further justifying the selection of this enhanced framework for remote sensing image dehazing.

4. **Extensive experimental validation:** In order to validate the effectiveness of the proposed methods, extensive experiments were conducted. These experiments provide critical insights into the performance and robustness of the SPRGAN, PNM pre-training, and RT Loss integration. We also test the processing efficiency of our model on a widely used edge computing platform.

In Section 2, we provide a comprehensive overview of related works in the field of image dehazing. In Section 3, we comprehensively introduce our proposed methods, providing a clear explanation of the underlying principles. Section 4 presents a detailed analysis, including comparisons, object detection results, efficiency analysis, and an ablation study. Finally, Section 5 offers conclusions drawn from our research findings.

2. Related Works

Efforts to address the challenges of image dehazing have been ongoing for several years, with significant progress made in recent times. These methods can be broadly divided into two main categories: prior information-based methods and learning-based methods.

2.1. Prior Information-Based Methods

Prior information-based methods have been fundamental in the field of image dehazing. These methods rely on specific assumptions about the scene or haze characteristics to guide the dehazing process. One of the most notable approaches in this category is the Dark Channel Prior (DCP) [4], which is based on the observation that in most haze-free patches of outdoor images, at least one color channel has some pixels with very low-intensity values. This dark channel property helps estimate the thickness of the haze, allowing for effective removal. Another influential method is the Color Attenuation Prior (CAP) [5], which exploits the difference in attenuation among different color channels to estimate depth information. Similarly, the Haze Line Prior [10–12] uses the linear relationship of colors in the presence of haze to improve dehazing accuracy. These methods are primarily based on the following atmospheric scattering model:

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

where $I(x)$ corresponds to the hazy image, while $J(x)$ represents the image without haze. The variable A represents the global atmospheric light, and $t(x)$ denotes the transmission mapping:

$$t(x) = \exp^{-\beta d(x)} \tag{2}$$

with β representing the scattering coefficient and $d(x)$ the distance from the observer to the object.

While these methods have been effective in estimating the transmission map and atmospheric light, they often struggle to achieve accurate results due to the inherent challenges associated with prior assumptions. They can be particularly sensitive to variations in scene depth and the presence of bright objects.

2.2. Learning-Based Methods

2.2.1. Supervised Learning Methods

With the advent of deep learning, learning-based methods have emerged as powerful tools for image dehazing. These methods leverage large datasets and neural network architectures to learn the complex relationships between hazy and clear images.

Neural networks offer a promising solution to address these challenges by estimating mapping parameters or directly generating dehazed images [6,13–20]. Research in computer vision, image processing, computational photography, and graphics, has resulted in robust dehazing systems in supervised settings, as demonstrated in various works [21–28]. Zheng [21] utilized a single image dehazing model based on a deep Convolutional Neural Network (CNN) to enhance the quality of remote sensing images. Li [22] proposed an atmospheric optical depth (AOD) estimation model to estimate the transmission map and atmospheric light. Qu [24] introduced an enhanced dehazing model based on a generative adversarial network (GAN) to improve the quality of remote sensing images. Zhu [28] proposed an atmospheric light estimation model to estimate the atmospheric light in hazy images. Ren [25] proposed a gated fusion network to enhance the quality of remote sensing images. Tian [27] proposed a variational prior model to estimate the transmission map and atmospheric light in hazy images.

However, these methods often require large amounts of paired data, which can be challenging and costly to obtain.

2.2.2. Semi-Supervised Learning Methods

To address the dependency on paired data, recent efforts have shifted towards unpaired image dehazing models. For instance, CycleGAN [7] has been utilized to learn mappings between hazy and dehazed images. DualGAN [29], DiscoGAN [30], and SMAP-GAN [31] adopt the CycleGAN architecture and fine-tune specific aspects to achieve improved results. However, in the context of remote sensing, CycleGAN often introduces noise into dehazed images and has minimal impact on structural similarity (SSIM) and peak signal-to-noise ratio (PSNR). Additionally, in the context of image dehazing, CycleGAN struggles to prioritize critical foggy areas within the input data, leading to noise pollution and feature loss.

To overcome these challenges, attention-guided image-to-image translation methods have gained popularity, employing attention modules to enhance image clarity. These modules utilize masks to highlight crucial features within input images, ensuring that the model concentrates its efforts effectively. For instance, Liang [32] utilized object mask annotations for model training, while [33] employed an additional network to acquire attention masks. Tang [34] utilized the CycleGAN structure and attention mechanism for image translation tasks. Similarly, Zheng [8] leveraged attention mechanisms to enhance the performance of dehazing models. However, these methods may struggle to capture the spectrum information necessary for accurate dehazing. Despite the effectiveness of attention-guided image-to-image translation methods in improving image clarity, there remains a critical need for further innovation in this domain. This motivation leads us to the adoption of the Enhanced Transformer-Guide Generative Adversarial Network (TGGAN) in this study.

To address the limitations of existing unpaired image dehazing models, we propose SPRGAN, a novel approach that leverages Enhanced Skip-Attention-Based Generative Adversarial Networks with Rotation Invariance. Our findings indicate that, in the spectrum domain, haze also leads to blurriness and loss of essential information. Through experimental analysis of image spectral characteristics under varying hazy conditions, we propose a Spatial-Spectrum Attention (SSA) mechanism. This mechanism enables the model to comprehend the relationships within the three-channel frequency domain information of the input. It allows the model to recover the spectral features of the hazy image through the spectrum encoder block. However, it is crucial to recognize that the training and convergence speed of the SSA can be noticeably slow. To address this issue, we have introduced a Skip-Attention Mechanism (SKIPAT) and Perlin Noise-Based Masks (PNM) to enhance the model's convergence speed and efficiency. Finally, the introduction of Rotation Loss (RT Loss) within the Transformer architecture represents a pioneering step in this research.

3. Proposed Method

In Section 3, we introduce our proposed methods in detail, explaining the underlying principles clearly. We discuss the key innovations of our research, which include the SSA-SKIPAT Generator, PNM pre-training, and RT Loss integration (as shown in Figure 2). These advancements improve spectral understanding, hasten convergence, and enhance image clarity, showing promise for superior remote sensing applications.
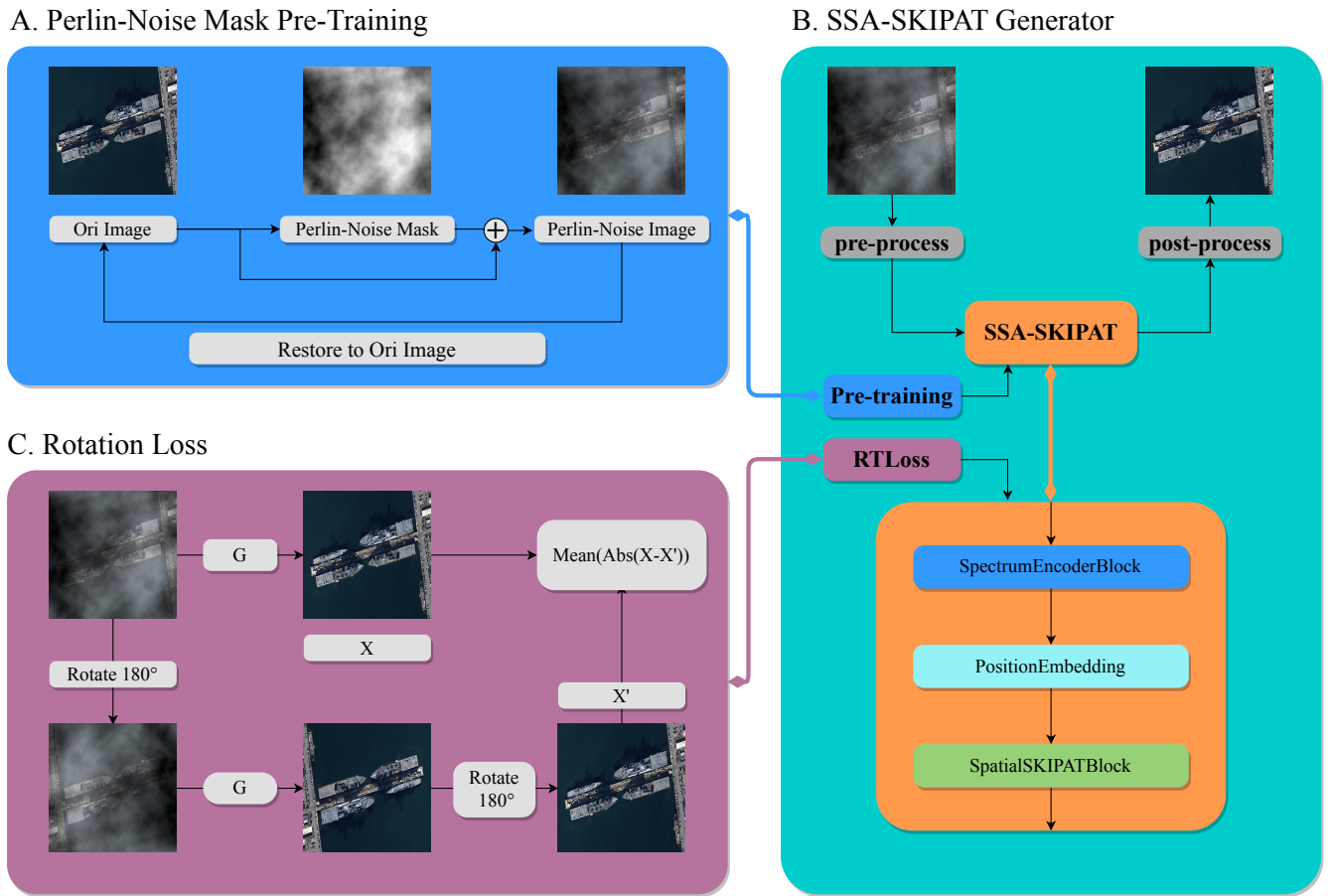
A. Perlin-Noise Mask Pre-Training

B. SSA-SKIPAT Generator

C. Rotation Loss

Figure 2. Framework of our innovation. (**A**) Perlin Noise Mask pre-training; (**B**) SSA-SKIPAT generator; (**C**) Rotation Loss.

3.1. SSA-Enhanced Generator with Skip-Attention

SPRGAN follows a structural framework similar to CycleGAN [7] for unpaired image dehazing, which uses two pairs of generators and discriminators to facilitate unpaired image-to-image translation. In the SSA-enhanced generator, we employ the UNet architecture [35], where the encoding path extracts features from the input via four convolutional layers with downsampling. The model incorporates a residual structure (skip connection [36]), facilitating the transfer of features between encoding and decoding layers. Additionally, features from the lowest encoding layer are fed into the Spatial-Spectrum Attention Vision Transformer with Skip-Attention (SSA-SKIPAT) (Figure 3A). The SSA-SKIPAT module adopts a fusion technique that combines information from both the spectrum and spatial domains to effectively learn feature information. Simultaneously, the SKIPAT module reduces the computational complexity of MSA while maintaining the accuracy of the original model.

In the encoding block of the UNet architecture, the pre-processing layer initially converts the input image into a tensor with dimensions (w_0, h_0, f_0). Subsequently, this tensor undergoes a series of transformations through four convolutional and pooling layers, resulting in an output feature with dimensions $(w_0/16, h_0/16, 8f_0)$. The output of the encoding block is then passed as input to the SSA-SKIPAT. The SSA-SKIPAT (Figure 3B) consists of three spectrum encoder blocks (Figure 3C), a position embedding block, and three spatial SKIPAT blocks (Figure 3D). Finally, the post-processing layer converts the output of the SSA-SKIPAT into the output image with dimensions $(w_0, h_0, 3)$.
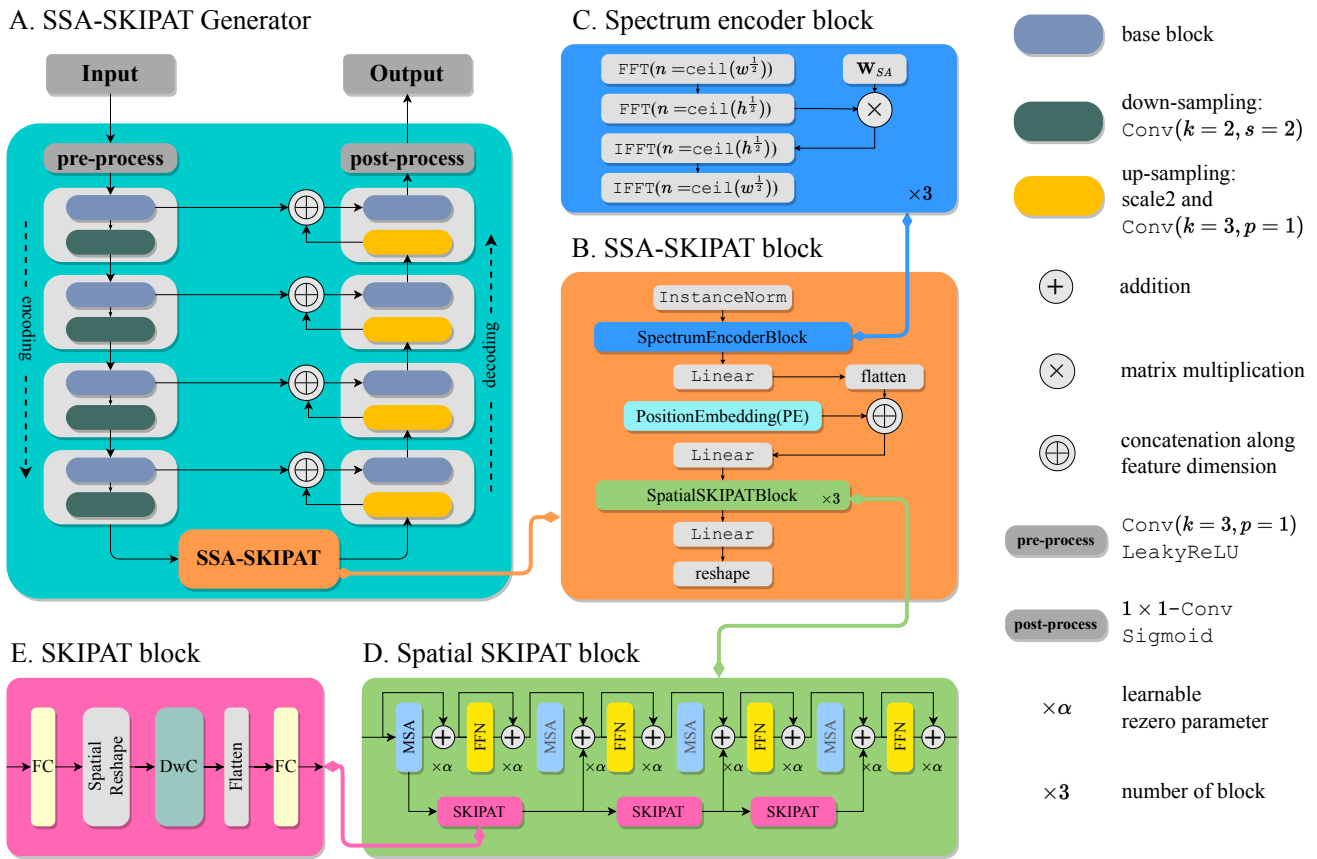
Figure 3. Framework of the proposed SPRGAN. (**A**) SSA-SKIPAT generator; (**B**) Spatial-Spectrum Attention SKIPAT block; (**C**) Spectrum encoder block; (**D**) Spatial SKIPAT block; (**E**) SKIPAT block.

Spectrum encoder block. The spectrum encoder block (Figure 3C) is composed of a 2D Fast Fourier Transform (FFT), a spectrum attention weight matrix (three dimensions) and a 2D Inverse Fast Fourier Transform (IFFT). The 2D FFT of a spatial domain signal (image) is calculated using the following formula:

$$F(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x,y) \cdot e^{-j2\pi\left(\frac{ux}{N} + \frac{vy}{M}\right)} \tag{3}$$

where $F(u,v)$ represents the frequency domain representation. $f(x,y)$ is the input spatial domain signal/image. (u,v) are the frequency domain coordinates. N and M are the dimensions of the input signal/image.

Then, we use a weight matrix W that modulates the spectral components of the image.

$$F_{\text{att}}(u,v) = F(u,v) \cdot W(u,v) \tag{4}$$

Here, $F(u,v)$ denotes the Fourier coefficients at frequency (u,v). $W(u,v)$ represents the attention weight matrix applied to the Fourier coefficients. By multiplying $F(u,v)$ with $W(u,v)$, we modify the spectral representation of the image. $F_{att}(u,v)$ represents the frequency domain representation. This allows us to selectively emphasize or de-emphasize certain frequencies, effectively enhancing desired features or suppressing noise and unwanted artifacts.

At last, we use the 2D IFFT to reconstruct the spatial domain signal from its frequency domain representation, which is calculated as follows:

$$f(x,y) = \frac{1}{NM} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} F_{att}(u,v) \cdot e^{j2\pi(\frac{ux}{N} + \frac{vy}{M})} \tag{5}$$

where $f(x,y)$ is the reconstructed spatial domain signal/image.

We employ 2D FFT to convert the input image features from the spatial domain to the frequency domain and utilize the spectrum attention weight matrix to learn the relationship between the three-channel frequency domain information of the image. By restoring the amplitude of features at different positions in the 2D FFT spectrum of the hazy image, we can effectively reconstruct and restore the image. To recover the original image's feature information, we apply different weights to various positions in the 2D FFT spectrum and perform a 2D inverse Fourier transform (IFFT) to return to the image domain. Similarly to spatial domain attention mechanisms, through network training, these weights can adaptively restore the critical regions in the spectrum image. The Fourier Transform is preferred for dehazing tasks due to its ability to separate and manipulate the image's low-frequency (general structures) and high-frequency (details and noise) components in the frequency domain. This method facilitates noise reduction by isolating and suppressing irrelevant high-frequency elements, enhances detail preservation by targeting specific frequency bands, and allows for efficient filtering, such as using high-pass filters to enhance edges and low-pass filters to smooth hazy regions. This approach, combined with trainable parameter matrices, offers more effective and efficient image processing than direct spatial domain methods.

At the same time, the SSA mechanism we use faces the problem of excessive computational complexity. To address this issue, we introduce the SKIPAT mechanism, which skips the MSA block and directly passes the input features to the FFN block, reducing the computational complexity and improving the model's convergence speed.

Skip-Attention block. The MSA block in ViT encodes the similarity of each patch to every other patch as an $n \times n$ attention matrix. This operator is computationally expensive, with a complexity of $O(n^2)$. As ViT scales, i.e., as n increases, the complexity grows quadratically and this operation becomes a bottleneck. As per the analysis in the paper [37], with the increasing number of MSA layers, the ability to extract target features does not significantly improve, but the computational load increases. Therefore, we are committed to finding the most cost-effective method that maximizes the effectiveness of MSA without substantially increasing computational complexity. To address these issues, we introduce the SKIPAT mechanism as shown in Figure 3D,E, which skips the MSA block and directly passes the input features to the FFN block. The SKIPAT parametric function consists of two linear layers and an interposed depthwise convolution (DwC) [38], as follows:

$$\hat{S}_i^{MSA} = FC_2(DwC(FC_1(S_{i-1}^{MSA}))) \tag{6}$$

where S_{i-1}^{MSA} is the input feature of the $i-1$th SKIPAT block. FC_1 and FC_2 are linear layers. DwC is the depthwise convolution operation, which is used to reduce the number of parameters and computational complexity. $\hat{S}_i^{MSA}$ is the output feature of the ith SKIPAT block.

The patch embeddings are then input to the first linear layer FC1: $FC_1 : \mathbb{R}^{n \times d} \to \mathbb{R}^{n \times 2d}$, where n is the number of patches and d is the dimension of the patch embeddings. Subsequently, the depthwise convolution DwC: $\mathbb{R}^{\sqrt{n} \times \sqrt{n} \times 2d} \to \mathbb{R}^{\sqrt{n} \times \sqrt{n} \times 2d}$ which is applied to the output of the first linear layer, followed by the second linear layer FC2: $FC_2 : \mathbb{R}^{n \times 2d} \to \mathbb{R}^{n \times d}$. The output of the second linear layer is the output of the SKIPAT block, which is then passed to the FFN block. We use three SKIPAT blocks in each Spatial encoder block, and the SPRGAN model consists of three Spatial encoder blocks. Experimental results show that the SKIPAT mechanism can effectively improve the model's convergence speed while maintaining performance.

3.2. Self-Supervised Pre-Training with Perlin Noise-Based Masks (PNM)

In traditional self-supervised pre-training, models are typically trained using raw image data alongside opaque masks for image inpainting tasks. However, for specific tasks like remote sensing image dehazing, this approach may be limiting. To address this limitation and better capture the complexity of remote sensing images, we introduce an innovative technique involving the use of Perlin Noise-Based Masks (PNM) Figrue 4.
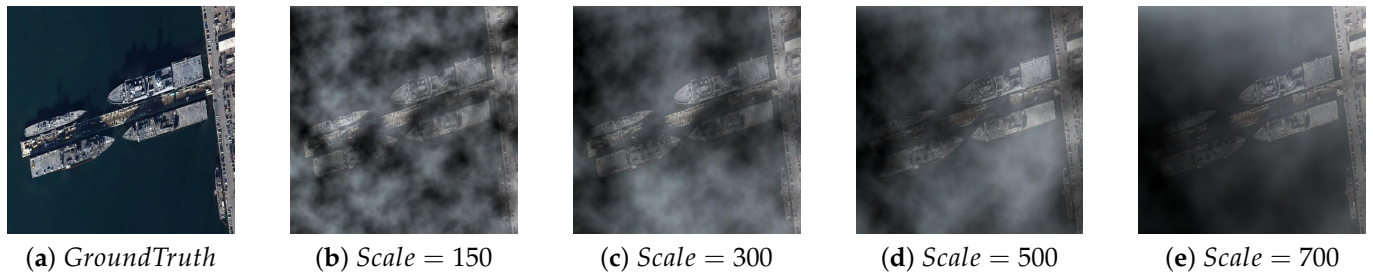

| (a) *GroundTruth* | (b) *Scale* = 150 | (c) *Scale* = 300 | (d) *Scale* = 500 | (e) *Scale* = 700 |

Figure 4. The Impact of Perlin Noise Masks on images. We generate Perlin Noise Masks with varying *Scale* values (150, 300, 500, 700).

PNM represents a unique type of mask utilized during pre-training in conjunction with the original images. Rather than employing traditional opaque masks, PNM introduces variability and complexity by incorporating Perlin Noise patterns. Perlin Noise is a type of gradient noise used in computer graphics to create natural-looking textures and smooth transitions. It is generated by combining multiple layers of noise at different frequencies and amplitudes.

The Perlin Noise formula for generating two-dimensional noise is given by:

$$N(x,y) = \sum_{i=0}^{n} \sum_{j=0}^{n} A_{i,j} \cdot \text{fade}(x-i) \cdot \text{fade}(y-j) \cdot G_{i,j} \tag{7}$$

where $N(x,y)$ represents the value of two-dimensional Perlin Noise at point (x,y). $A_{i,j}$ is a two-dimensional array representing the amplitude of the Perlin Noise, $G_{i,j}$ is a two-dimensional array representing the gradient vector of the Perlin Noise, n is the order of the Perlin Noise, and fade is a fade function used for smooth transitions. Specifically, the fade function $\text{fade}(t)$ can be expressed as:

$$\text{fade}(t) = 6t^5 - 15t^4 + 10t^3 \tag{8}$$

In practice, Perlin Noise is often computed using interpolation functions such as linear interpolation, cubic interpolation, etc. These interpolation functions help create smooth transitions between discrete noise values, resulting in continuous Perlin Noise patterns.

The formula for adding Perlin Noise to an image is:

$$\text{noisy\_image}(x,y) = \text{image}(x,y) + \text{scale} \cdot N(x,y) \tag{9}$$

Here, $\text{noisy\_image}(x,y)$ represents the value of the noisy image at point (x,y), $\text{image}(x,y)$ represents the value of the original image at point (x,y), and scale is the scaling factor of the Perlin Noise. By changing the scale value, we can adjust the intensity of the Perlin Noise added to the image. Specifically, we generate Perlin Noise patterns with dimensions matching those of the input images. Each pixel in the noise pattern corresponds to a transmission coefficient, determining the opacity of the corresponding pixel in the original image. By applying Perlin Noise patterns as masks, we introduce semi-random variations in the opacity of different image regions, effectively simulating the diverse atmospheric conditions encountered in remote sensing imagery.

In the Perlin Noise Mask (PNM) PNM pre-training step, we pre-train the SPRGAN generators on an image inpainting task. We use the PNM to generator masked image

from the original unmasked image. Then, the generator is trained to predict the original unmasked image using pixel-wise L_1 loss. After initial pretraining, our method employs an unpaired training strategy using CycleGAN's cycle consistency and adversarial losses. This allows the model to generate and reconstruct clear images from hazy inputs without needing direct pairs. This approach is highly suitable for real-world applications like urban surveillance, agricultural monitoring, coastal surveillance, and traffic monitoring, where paired hazy and clear images are challenging to obtain. The model's robustness and adaptability ensure reliable dehazing performance across diverse scenarios.

During the Perlin Noise Mask (PNM) pre-training step with SPRGAN, the generators are initially trained on an image inpainting task to learn to reconstruct images effectively. This pre-training leverages techniques in the frequency domain, such as Fourier Transform, to process hazy images. Fourier Transform decomposes the image into frequency components, where higher frequencies correspond to fine details and noise, and lower frequencies represent general image structures. By manipulating these components, filters can selectively enhance details and suppress noise. After applying these transformations, inverse Fourier Transform reconstructs the processed image into the spatial domain, yielding clearer images with improved detail visibility and reduced haze effects.

Throughout the pre-training process, the model learns to handle these Perlin Noise-Based Masks, thereby enhancing its ability to address the challenges posed by haze and improve both super-resolution and dehazing capabilities. The advantage of using PNM lies in its ability to introduce realistic variability and complexity into the pre-training process, better simulating real-world atmospheric conditions encountered in remote sensing images. By incorporating Perlin Noise patterns, the model gains the capacity to adapt to diverse hazy environments, leading to improved performance and accelerated convergence speed in subsequent tasks.

3.3. Enhanced Objective with Rotation Loss

In remote sensing image dehazing, we often encounter scenarios where the orientation of the input image may vary. Rotation Loss calculates the mean absolute difference between corresponding pixels in the original haze-free image and the haze-free image obtained after rotating the original image by 180 degrees and passing it through the generator. It quantifies the pixel-level difference between the two images, reflecting variations in color, texture, and other pixel attributes.

This Rotation Loss helps evaluate the model's ability to maintain consistency in dehazing performance across different orientations of the input images, thus enhancing the model's robustness and generalization capability. Additionally, it provides insights into how well the model preserves structural information during the dehazing process.

$$\mathcal{L}_{RT}(x,y) = \text{mean}(|G(x,y) - R(G(R(x,y)))|) \tag{10}$$

where $G(x,y)$ represents the intensity value of the pixel at position (x,y) in the original haze-free image generated by the generator. R represents the operation of rotating the image by 180 degrees. $|\cdot|$ denotes the absolute value. $\text{mean}(\cdot)$ computes the mean value over all pixel coordinates. With the inclusion of the RT Loss, we can formulate our objective as follows:

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_{cycle} * \mathcal{L}_{cycle} + \lambda_{idt} * \mathcal{L}_{idt} + \lambda_{RT} * \mathcal{L}_{RT} \tag{11}$$

where $\mathcal{L}_{GAN}$, $\mathcal{L}_{cycle}$, $\mathcal{L}_{idt}$ and $\mathcal{L}_{RT}$ are adversarial loss, cycle-consistency loss, identity-consistency and total variation loss, respectively. The parameters λ_{cycle}, λ_{idt} and λ_{RT} are to control the relative importance of the three objectives.

The core idea behind Rotation Loss (RT Loss) is to enforce a constraint ensuring that the dehazing results remain consistent regardless of geometric transformations applied to the input images. When an image is geometrically transformed (e.g., rotated) and then input into the model, the resulting dehazed image should be nearly identical to the

image obtained by first inputting the original image into the model and then applying the same transformation to the output. This constraint ensures consistent performance across different orientations. Additionally, simply augmenting the dataset with rotated images only ensures the model is trained on various angles, but does not guarantee uniform dehazing quality. Using RT Loss, the training process becomes more efficient, as it achieves better outcomes with fewer batches per epoch, enhancing the model's robustness and adaptability to diverse scenarios.

4. Experiments and Results

In this section, we present an evaluation of the single image dehazing capabilities of our model through a series of experiments. First, we compare our results with state-of-the-art single image dehazing approaches. Secondly, we evaluate the object detection performance of our model on the RSD dataset, RICE dataset and RESIDE dataset [39]. Then, we analyze the convergence behavior of the proposed algorithm. Finally, we perform an ablation study to further elucidate the effectiveness and contributions of different components within our approach.

4.1. Datasets

We use the RSD dataset [8] for our experiments. This dataset is notable for its comprehensive collection of extensively annotated remotely sensed ship images paired with their corresponding dehazed versions. It serves a dual purpose, allowing evaluation of both single-image dehazing and military/civilian ship detection. The dataset consists of a total of 7000 simulated and hazy ship images, containing over 10,000 ship targets. In particular, 3500 paired images, including both hazed and haze-free versions, are synthetically generated using *atmospheric scattering* model. In addition, the dataset's test set includes over 300 real-life hazy weather images, allowing for a robust evaluation of our model's generalization capabilities.

To further confirm the effectiveness and versatility of our proposed method, we performed experiments on the RICE dataset [40], which is tailored for cloud removal in remote sensing images.

RESIDE [39] is an indoor dataset, which contains synthetic and real-world hazy images, called Realistic Single Image Dehazing. This dataset highlights various data sources and image content and is divided into five subsets, each for different training or evaluation purposes. Various dehazing algorithm evaluation criteria are provided, including full reference metrics, no-reference metrics, subjective evaluation, and task-driven evaluation. By evaluating our results alongside the latest state-of-the-art approaches on these datasets, we ensure a thorough assessment of our method's performance.

4.2. Experiment Details
4.2.1. Training Strategy

In our approach, the generator undergoes pre-training via a self-supervised image inpainting process. Following this, we introduce Perlin Noise Mask to dim these patches, better simulating hazy weather conditions. We utilize the Adam optimizer and a cosine annealing learning-rate scheduler, along with standard data augmentation techniques during training. Specifically, we pre-train one generator using the RSD dataset, enhancing its ability to restore ship targets effectively.

Following pre-training, we proceed with adversarial training on the entire model using the RSD dataset. We implement our model using the PyTorch framework and conduct approximately 500 to 1000 epochs for each dataset. Our training strategy involves utilizing the Adam optimizer with a learning rate initially set at 0.0001 for the first half of the training duration. Subsequently, we linearly anneal it to zero during the second half of the training process.

4.2.2. Parameter Setting

Furthermore, we incorporate three data augmentations: resizing, random cropping to dimensions of 256×256, and random horizontal flipping. We set the batch size to 6 and utilize a single NVIDIA 4090 GPU with 24 GB of VRAM. Our optimization process employs the Adam optimizer with momentum terms set to $\beta_1 = 0.5$ and $\beta_2 = 0.999$.

We determine the values of two critical hyperparameters: $Scale = 500$ and $\lambda_{RT} = 5e - 2$. Figure 5 demonstrates that when $Scale = 500$, we achieve the fastest model convergence speed and the best PSNR and SSIM results (PSNR = 28.10, SSIM = 0.8674). This evidence strongly supports the notion that employing Perlin Noise Masks during pre-training offers significant performance advantages. Similarly, Figure 6 illustrates that when $\lambda_{RT} = 5e - 2$, we attain the swiftest model convergence and the highest PSNR and SSIM results (PSNR = 28.14, SSIM = 0.8768). This provides clear evidence that incorporating the RT Loss during image restoration training, in contrast to the original objective ($\lambda_{RT} = 0$), leads to substantial improvements in image quality. It is worth noting that as PSNR increases, the quality of the dehazed images improves, leading to an increase in SSIM. In conclusion, we set $Scale = 500$ and $\lambda_{RT} = 5e - 2$.
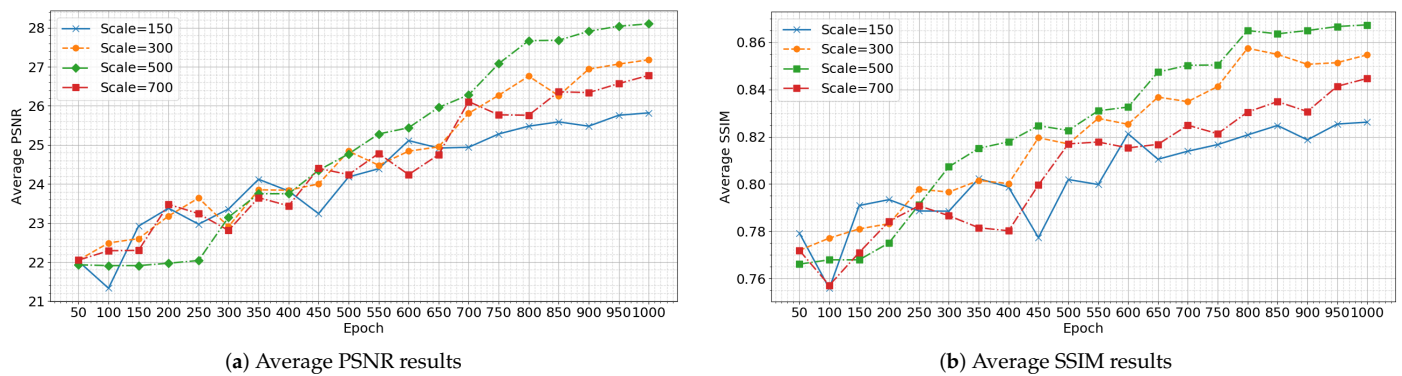

Figure 5. Average PSNR (**a**) and SSIM (**b**) results of dehazed images which are generated by models with different *Scale*, which denote the scale of cloud generated by Perlin Noise Masks.
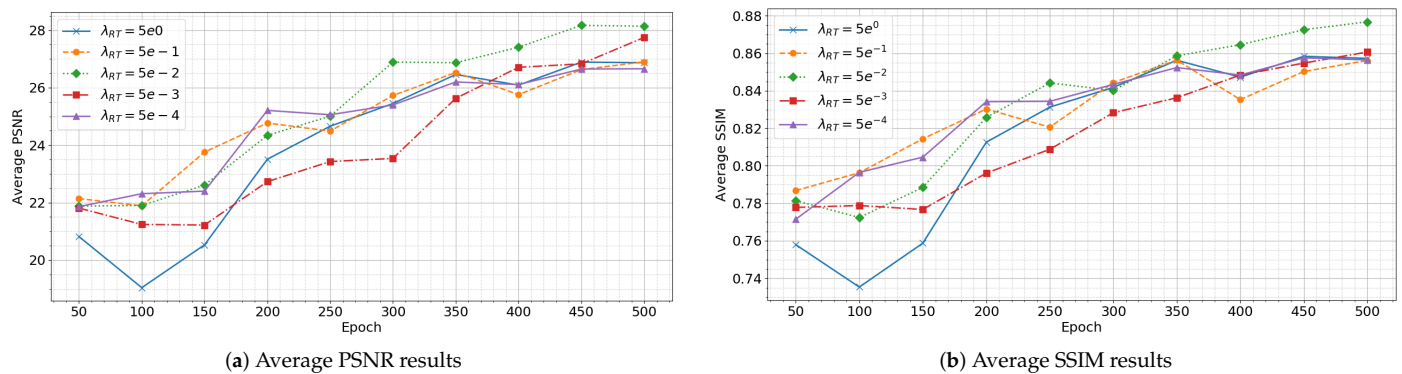

Figure 6. Average PSNR (**a**) and SSIM (**b**) results of dehazed images which are generated by models with different λ_{RT}.

In our dehazing task, we emphasize cyclic consistency, thus opting for a configuration where λ_{cycle} carries greater weight, specifically using the 10:0.5 ratio. We set λ_{cycle} to 10 and λ_{idt} to 0.5, which were found to yield the best results. λ_{RT} is set to 0.05 to balance the rotational consistency constraint. These values ensure the best balance between the different loss components and the overall model performance.

4.2.3. Competing Models

We have conducted experiments on the RSD and RICE datasets to evaluate the performance of our method in comparison to other state-of-the-art techniques in the field of

unpaired single image dehazing. The selected methods for comparison include Attention-GAN [34], CycleGAN [41], and UVCGAN [9].

4.2.4. Evaluation Metrics

When performing experiments or meeting other requirements, image processing procedures may affect image quality or change its content. These changes can be measured using image quality metrics. PSNR/SSIM, as discussed by A. Hore [42], is suitable for this purpose. In this study, we use PSNR/SSIM to assess image quality.

PSNR is a widely used metric to evaluate the quality of reconstructed images. It is defined as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$$

where MAX is the maximum possible pixel value of the image. MSE is the Mean Squared Error between the original and reconstructed images. The MSE is calculated as:

$$\text{MSE} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i,j) - K(i,j)]^2$$

where $I(i,j)$ is the pixel value of the original image. $K(i,j)$ is the pixel value of the reconstructed image. M and N are the dimensions of the images. PSNR is a widely used metric for evaluating image quality, measuring the ratio between the maximum possible power of a signal and the power of corrupting noise. Higher PSNR values indicate better image quality, with a maximum value of infinity for identical images. The PSNR is calculated using the mean squared error (MSE) between the original and dehazed images. A higher PSNR value indicates that the dehazed image is closer to the original clear image, reflecting improved image quality.

SSIM is a metric that evaluates the perceived quality of digital images. It considers changes in structural information, luminance, and contrast. SSIM is defined as follows:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where μ_x and μ_y are the average values of images x and y. σ_x^2 and σ_y^2 are the variances of images x and y. σ_{xy} is the covariance of images x and y. C_1 and C_2 are constants to stabilize the division when the denominators are close to zero. SSIM is another popular metric for evaluating image quality, measuring the similarity between two images based on luminance, contrast, and structure. SSIM values range from -1 to 1, with 1 indicating identical images. SSIM is calculated using luminance, contrast, and structure comparison between the original and dehazed images. A higher SSIM value indicates that the dehazed image is more similar to the original clear image in terms of structure and content. In addition, we evaluate mean average precision (mAPs) values using YOLOv8l [43].

4.3. Comparison with State-of-the-Art Approach

We conducted a comparative analysis of our proposed model against other state-of-the-art methods in the single image dehazing domain, utilizing the RSD dataset. The selected methods for experimentation included CycleGAN [7], UVCGAN [9], and Dehaze-AGGAN [8]. The results of our comparisons are visually presented in Figures 7–9, while quantitative metrics in terms of average PSNR/SSIM values can be found in Table 1. Furthermore, we performed object detection using the YOLOv8l [43] detector after the dehazing process, and the corresponding results are shown in Figure 10 and quantified in Table 2.
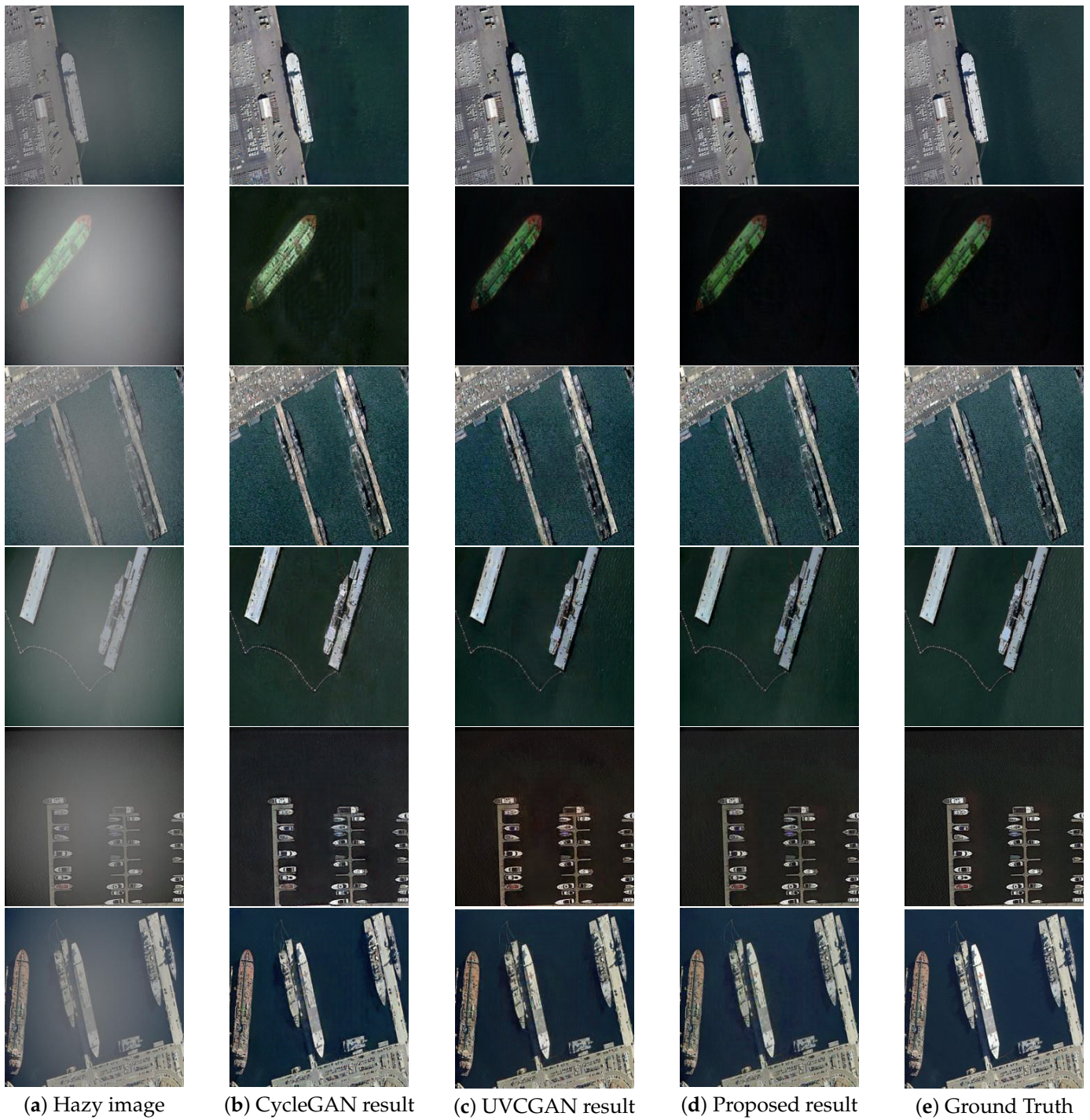
(**a**) Hazy image   (**b**) CycleGAN result   (**c**) UVCGAN result   (**d**) Proposed result   (**e**) Ground Truth

Figure 7. Qualitative results on RSD datasets. We test our dataset on three baseline models including ours. The resulting order in each row is hazy image, CycleGAN [7] result, UVCGAN [9] result, our result and the Ground Truth.
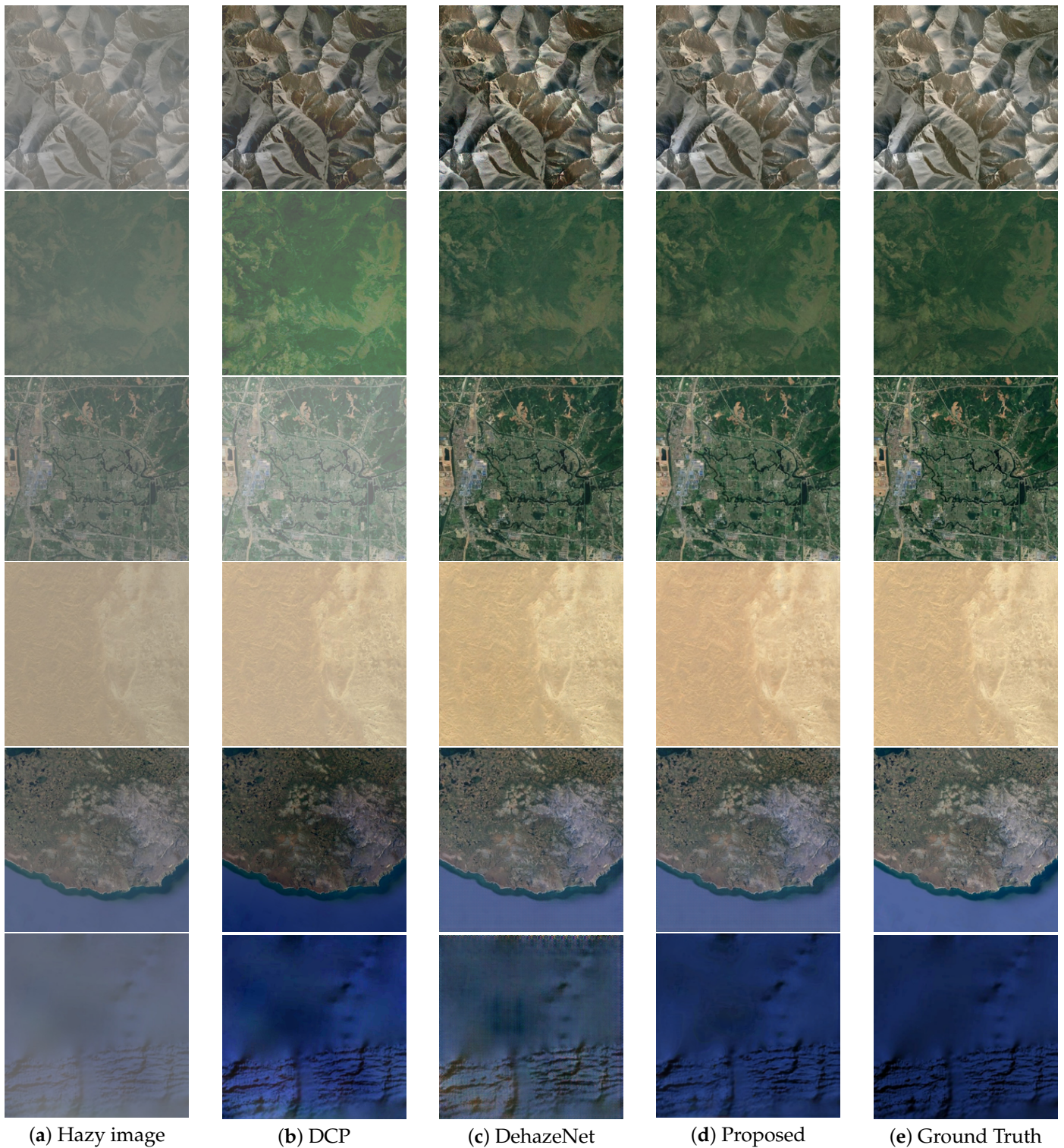
(**a**) Hazy image   (**b**) DCP   (**c**) DehazeNet   (**d**) Proposed   (**e**) Ground Truth

Figure 8. Qualitative results on RICE datasets. We selected 6 images of different scenes (mountains, forests, urban areas, deserts, coasts, deep sea (from top to bottom)) from the RICE dataset to show the performance of the three models. The result order in each row is hazy image, DCP [4] result, DehazeNet [19] result, our result and Ground Truth.
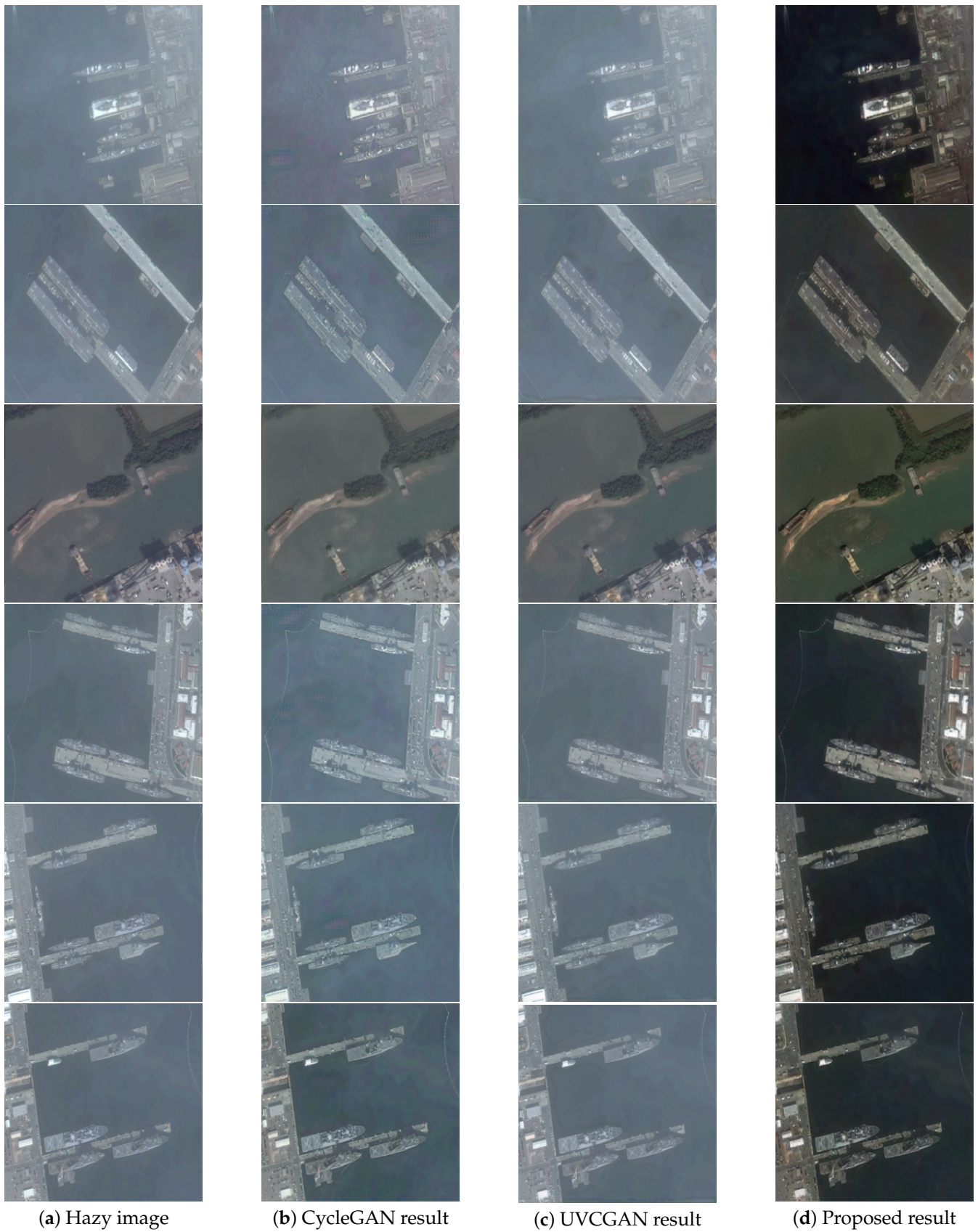
| (**a**) Hazy image | (**b**) CycleGAN result | (**c**) UVCGAN result | (**d**) Proposed result |

Figure 9. Qualitative results on RSD datasets. We test real hazy images on three baseline models, including ours. The resulting order in each row is hazy image, CycleGAN [7] result, UVCGAN [9] result and our result.
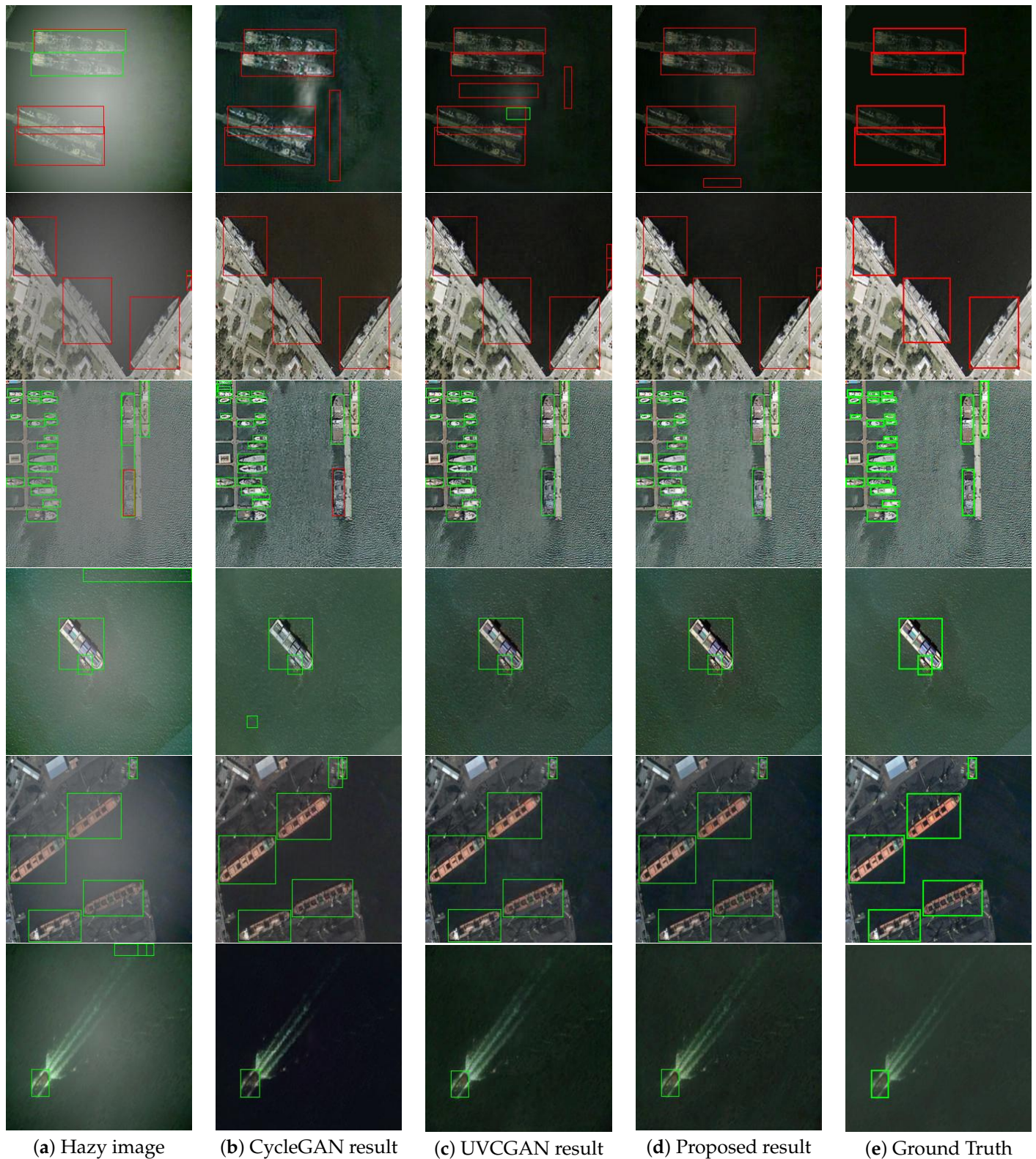
(**a**) Hazy image (**b**) CycleGAN result (**c**) UVCGAN result (**d**) Proposed result (**e**) Ground Truth

Figure 10. Object detection results on dehazing results. We tested the dehazing results with the baseline detector YOLOv8l [43]. The result order in each row is hazy image, CycleGAN [7] result, UVCGAN [9] result, our result and the Ground Truth.

Table 1. Average PSNR and SSIM results on RSD datasets (unpaired single image dehazing methods).

Methods	Metrics	Test
None	PSNR | SSIM	13.22 | 0.6523
CycleGAN [7]	PSNR | SSIM	23.67 | 0.8211
Dehaze-AGGAN [8]	PSNR | SSIM	24.11 | 0.8356
UVCGAN [9]	PSNR | SSIM	26.31 | 0.8641
Proposed	PSNR | SSIM	**28.31 | 0.8806**

Table 2. Object detection performance (mAP and the improvement of hazy images) of ship on RSD dataset.

Methods	mAP (%)	Gain	mAP 50–95 (%)	Gain
None	98.49	-	93.41	-
CycleGAN [7]	98.60	0.11	94.36	0.95
Dehaze-AGGAN [8]	98.72	0.23	94.71	1.30
UVCGAN [9]	99.28	0.79	95.81	2.40
Proposed	**99.43**	0.94	**96.12**	2.71
Ground Truth	99.89	1.40	97.01	3.60

4.3.1. Qualitative Results

Figure 7 provides a visual comparison between CycleGAN [7], UVCGAN [9], and SPRGAN. CycleGAN [7], a widely recognized model for unpaired image translation, exhibits certain limitations in effectively removing haze from images. While it succeeds in generating dehazed images, the extent of haze removal and the preservation of image details may not be optimal. UVCGAN [9], on the other hand, demonstrates improved performance compared to CycleGAN. It effectively reduces the haze in images, resulting in clearer and more visually appealing outputs. However, it may still encounter challenges in completely eliminating haze artifacts and preserving fine details in the images. In contrast, SPRGAN, our proposed model, surpasses both CycleGAN and UVCGAN in terms of haze removal and image clarity. Leveraging the Spatial-Spectrum Attention mechanism and Perlin Noise-Based Masks, SPRGAN excels in effectively restoring hazy images with minimal artifacts and superior detail preservation. Additionally, the integration of RT Loss further enhances its performance, ensuring accurate haze removal and enhanced visual quality. Overall, SPRGAN demonstrates superior performance compared to existing models, making it a promising solution for unpaired image dehazing tasks.

In Figure 8, we compare the results of our model with DCP [4] and DehazeNet [19] on the RICE dataset. We select six images of different scenes (mountains, forests, urban areas, deserts, coasts, deep sea) from the RICE dataset to test the performance of the three models. Especially in the coast scene, the DCP [4] has poor performance in color restoration and detail recovery, while DehazeNet [19] still retains some hazy and noisy areas in the image. In all six of the scenes, our model has the best performance in terms of color restoration, detail recovery, and noise reduction due to the use of Spectrum Attention and TV Loss.

Figure 9 exhibits qualitative outcomes derived from the evaluation of real hazy images. The performance of UVCGAN [9] and CycleGAN [7] in eliminating fog is insufficient, as evidenced by the residual haziness in the results. Conversely, SPRGAN demonstrates remarkable efficacy in haze removal, showcasing significant clarity improvement, even when initially trained on simulated images. This observation underscores the versatility and robustness of our model, indicating its suitability for diverse real-time remote sensing applications.

4.3.2. Quantitative Results

Tables 1 and 3 display the average PSNR and SSIM results for the RSD and RICE dataset. In the first row of the results, you can find the average PSNR and SSIM values

calculated directly between each hazy image and its corresponding ground truth image. It is evident that our proposed method gets the best results in terms of PSNR and SSIM values (28.31 | 0.8806 and 33.42 | 0.9629). Table 1 shows the average PSNR and SSIM results on the RSD dataset for different unpaired single image dehazing methods. The data show that the images without any dehazing treatment have the lowest PSNR and SSIM values of 13.22 and 0.6523, respectively. After processing by CycleGAN, Dehaze-AGGAN, and UVCGAN, there are improvements in both PSNR and SSIM, reaching 23.67 | 0.8211, 24.11 | 0.8356, and 26.31 | 0.8641, respectively. However, the proposed method surpasses these results with higher PSNR and SSIM values, reaching 28.31 and 0.8806, respectively, demonstrating its superiority in single image dehazing tasks.

Table 3. Average PSNR and SSIM results on RICE datasets.

Methods	Metrics	Test
None	PSNR	SSIM	16.63	0.7391
DCP [4]	PSNR	SSIM	17.96	0.8427
CycleGAN [7]	PSNR	SSIM	28.12	0.9189
DehazeNet [19]	PSNR	SSIM	29.48	0.9210
Dehaze-AGGAN [8]	PSNR	SSIM	30.19	0.9356
SPAGAN [44]	PSNR	SSIM	30.23	0.9572
pix2pix [45]	PSNR	SSIM	31.03	0.9124
UVCGAN [9]	PSNR	SSIM	32.09	0.9491
Proposed	PSNR	SSIM	**33.42**	**0.9629**

Quantitative analysis of the data in Table 3 provides several insights into the performance of different dehazing methods on the RICE dataset. UVCGAN [9] achieves a respectable PSNR of 32.09 dB and SSIM of 0.9491, indicating a good performance in terms of image quality improvement. However, our proposed method outperforms UVCGAN and all other methods compared, achieving a remarkable PSNR of 33.42 dB and SSIM of 0.9629. This significant improvement demonstrates the effectiveness of our approach in improving image clarity and fidelity. In comparison, CycleGAN [7] achieves a lower PSNR of 28.12 dB and SSIM of 0.9189, suggesting inferior performance compared to UVC-GAN and our proposed method. Notably, our method outperforms even state-of-the-art approaches such as SPAGAN [44] and pix2pix [45], indicating its superiority in dehazing remote sensing images on the RICE dataset.

4.4. Object Detection Results

We conduct object detection experiments on the RSD dataset to further evaluate the effectiveness of our proposed method. RSD is a fine-grained categorization dataset that distinguishes between military and civilian ship targets. We train the YOLOv8l [43] detector on the RSD dataset using the pre-trained parameters and standard optimization method outlined in the official YOLOv8l [43] model. After the dehazing process, we test the detection results of the hazy images, CycleGAN [7], UVCGAN [9], our proposed method and the ground truth. We use the mean Average Precision (mAP) as the evaluation metric for object detection results, which is calculated based on the intersection over union (IoU) threshold of 0.5 (mAP) and 0.5–0.95 (mAP 50–95(%)).

4.4.1. Qualitative Results

Figure 10 shows a visual comparison of object detection results on the RSD dataset, including Hazy images, CycleGAN [7] results, UVCGAN [9] results, our results, and the Ground Truth. It is important to note that the YOLOv8 model excels at detecting small objects, so it is crucial to avoid false negatives and false positives in performance comparisons. The detection results of the hazy images, CycleGAN [7], and UVCGAN [9] show limited false and missing detections. In testing the detection of small objects, our

proposed method exhibits no missing detections, with only a few false detections. The proposed method demonstrates remarkable performance in reducing false and missing detection errors, closely approaching ground truth results. This outcome highlights the efficacy of the approach in enhancing object detection accuracy, even under hazy conditions, and achieving results comparable to those in clear conditions.

4.4.2. Quantitative Results

Table 2 displays the outcomes of our experiments, focusing on mean Average Precision (mAP), and reveals the substantial improvements achieved when compared to hazy images. In our approach, we surpass all four comparison methods, achieving a 2.71 increase in mAP (50–90) when contrasted with results from hazy images. These results serve as compelling evidence, demonstrating the efficacy of our method in effectively restoring critical image features. Furthermore, our approach has a notably positive influence on the recognition of remote sensing images.

4.5. Cross-Dataset Experiments

We have added the cross-dataset experiments with SPAGAN in the revised manuscript.

4.5.1. Qualitative Results

Figure 11 shows the cross-dataset experiments with SPAGAN: Figure 11a shows dehazing results through the model which is trained in the RSD dataset and tested in the RSD dataset. In Figure 11b, the training dataset is RESIDE and the testing dataset is RSD. Figure 11c's training dataset is RESIDE and its testing dataset is RESIDE. At last, in Figure 11d, the training dataset is RSD and the testing dataset is RESIDE. According to the qualitative results, the models trained by indoor and remote datasets can clear the haze from both datasets. It is obvious that the model trained by RESIDE generates brighter images than the model trained by RSD. On the other hand, the model trained by RSD can restore more details, which proves that our proposed method has strong robustness and adaptability in different environments.
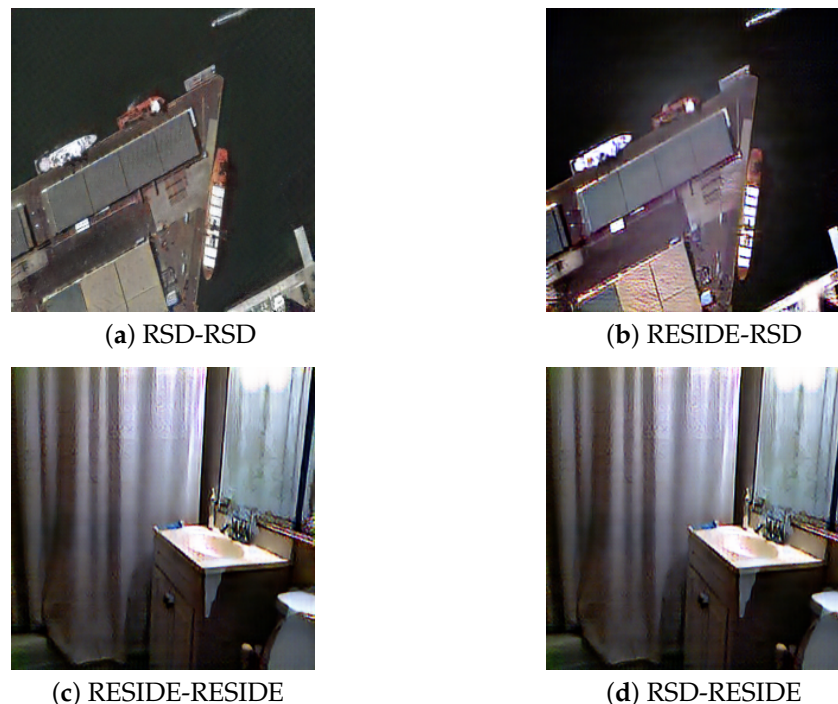

(**a**) RSD-RSD

(**b**) RESIDE-RSD

(**c**) RESIDE-RESIDE

(**d**) RSD-RESIDE

Figure 11. Comparative qualitative results between single and cross dataset experiments via SPAGAN.

4.5.2. Quantitative Results

Table 4 shows the average PSNR and SSIM results for the cross-dataset experiments. According to the results, the model trained by RSD dataset and tested by RSD dataset has the best performance in terms of PSNR and SSIM values (28.31 | 0.8806). Moreover, the model trained by RESIDE dataset and tested by RESIDE dataset has the second-best performance in terms of PSNR and SSIM values (27.19 | 0.8776). These results demonstrate the robustness and adaptability of our proposed method in different environments.

Table 4. Average PSNR and SSIM results for cross-dataset experiments.

Methods	Metrics	Test
RSD-RSD	PSNR	SSIM	28.31	0.8806
RESIDE-RSD	PSNR	SSIM	27.48	0.8654
RESIDE-RESIDE	PSNR	SSIM	27.19	0.8776
RSD-RESIDE	PSNR	SSIM	26.90	0.8612

4.6. Results on Ablation Study

To assess the impact of our model improvements on overall performance, we conducted an ablation study with five parts: (1) Model A: models without SKIPAT, PNM pre-training and RT Loss; (2) Model B: models with SKIPAT; (3) Model C: models with SKIPAT and PNM; (4) Model D: models with SKIPAT and RT Loss; (5) Model proposed: models with SKIPAT, RT Loss and PNM.

4.6.1. Qualitative Results

We present the qualitative results in Figure 12. Figure 12a,g show the hazy image and ground truth, respectively. Figure 12b–e display the dehazing results of Model A, B, C, and D, while Figure 12f displays the dehazing results of our proposed model.

In the first image, our proposed method outperforms the other models in faithfully restoring the orange color on the top of the cargo ship. In addition, we observe that models B, C, and D (with SKIPAT) have more vivid colors compared to model A (without SKIPAT), thanks to the frequency domain information extracted by SSA. Furthermore, in the second and third images, our proposed model comes close to the ground truth in terms of recovering details of the ships and the coastline. These results confirm the effectiveness and adaptability of our proposed method.
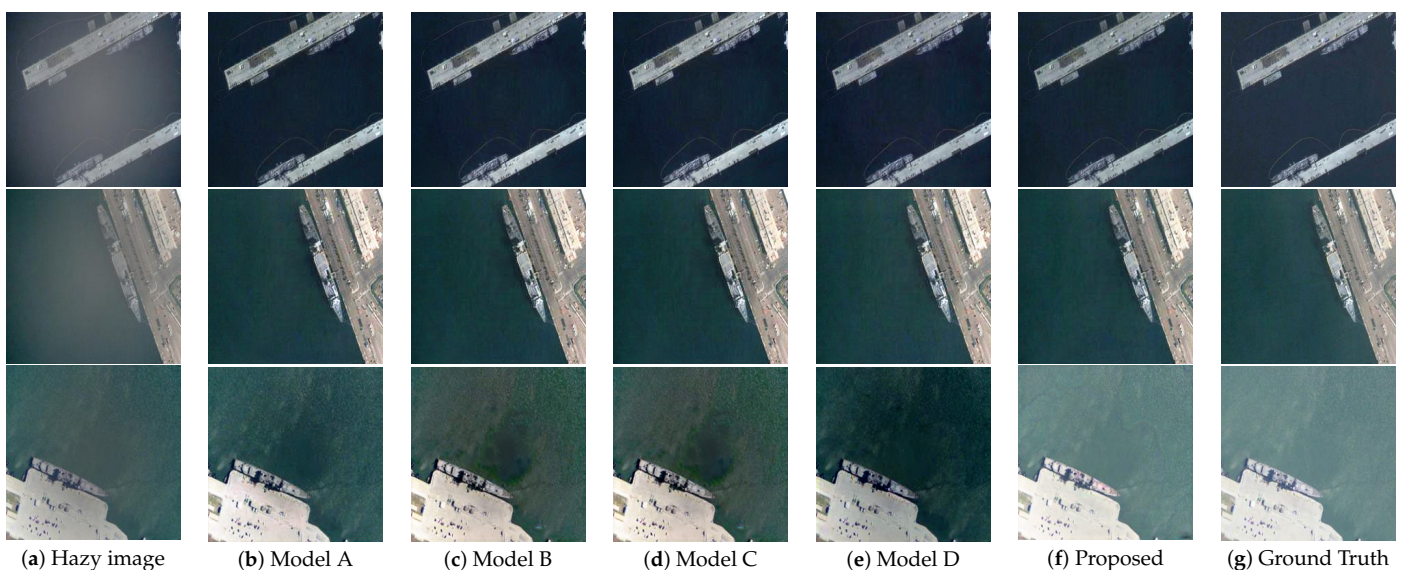

(**a**) Hazy image (**b**) Model A (**c**) Model B (**d**) Model C (**e**) Model D (**f**) Proposed (**g**) Ground Truth

Figure 12. Qualitative results of the ablation study.

4.6.2. Quantitative Results

Table 5 shows the PSNR/SSIM values for models A, B, C, D and the proposed method on the RSD test dataset. By analyzing the data, we can see that the SSA-SKIPAT module significantly improves the PSNR performance of the model, while moderately improving the SSIM performance. Models with SKIPAT show an increase in PSNR from 26.41 to 27.43 and an increase in SSIM from 0.8641 to 0.8658. In addition, the inclusion of RT Loss leads to a further increase in PSNR from 27.43 to 28.14 and SSIM from 0.8658 to 0.8768. On the other hand, the inclusion of PNM pre-training leads to a slight increase in PSNR from 27.43 to 28.10 and SSIM from 0.8658 to 0.8674. Finally, the proposed model, which includes all three components, achieves the highest PSNR and SSIM values of 28.31 and 0.8806, respectively. In summary, the incremental addition of SKIPAT, RT Loss and PNM has significantly improved the performance of the model. These improvements are in line with our expectations and provide strong evidence of the effectiveness of these methods.

Table 5. Average PSNR and SSIM results on RSD datasets in ablation study.

Methods	SKIPAT	PNM	RT Loss	Metrics	Test
Model A				PSNR | SSIM	26.41 | 0.8641
Model B	✓			PSNR | SSIM	27.43 | 0.8658
Model C	✓	✓		PSNR | SSIM	28.10 | 0.8674
Model D	✓		✓	PSNR | SSIM	28.14 | 0.8768
Proposed	✓	✓	✓	PSNR | SSIM	**28.31 | 0.8806**

4.7. Convergence and Efficiency of Proposed Algorithm

To assess the convergence of our method, we track the PSNR and SSIM performance over different training epochs. Figure 13 shows the evolution of the average PSNR and SSIM values for the dehazed images throughout the training process. Initially, the PSNR increases from 20.17 to its peak of 28.31, along with an increase in the SSIM from 0.7371 to 0.8796 as the number of epochs increases. However, beyond approximately 800 epochs, we observe fluctuations in the model's performance, with minimal changes in both PSNR and SSIM values. Therefore, we conclude that the model achieves convergence after approximately 800 epochs.
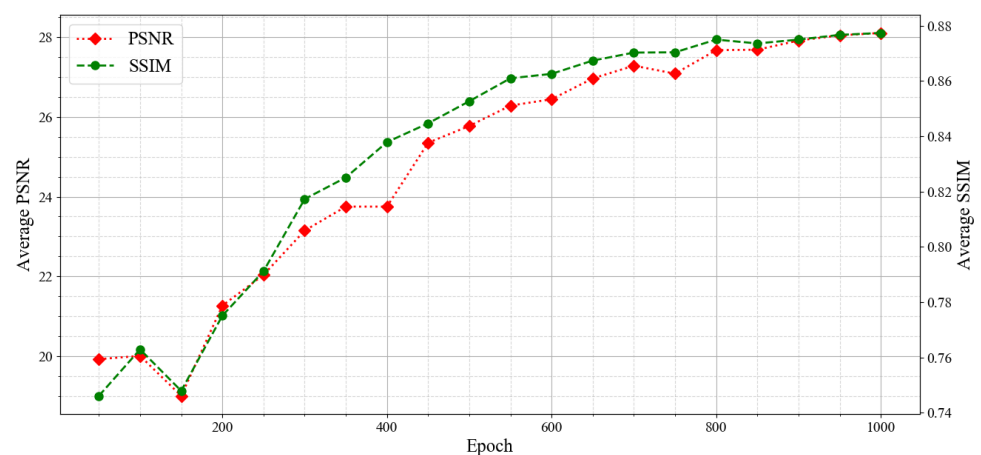

Figure 13. Evolution of average PSNR and SSIM with epochs.

We assess the parameters and efficiency of our proposed method and other baseline models in Table 6. By analyzing the data provided in Table 6, we observe that Cycle-GAN [7] and Dehaze-AGGAN [8] have similar parameter counts, around 55M and 60M, respectively, with PSNR values of 23.67 and 24.11, and SSIM values of 0.8211 and 0.8356,

respectively. UVCGAN [9], on the other hand, has a higher parameter count of 68M, with better performance, achieving a PSNR of 26.31 and an SSIM of 0.8641.

Table 6. Training time and parameters of CycleGAN [7], Dehaze-AGGAN [8], UVCGAN [9] and our methods.

Model Name	Training Time (h)	Para.	PSNR ǀ SSIM
CycleGAN [7]	40	55M	23.67 ǀ 0.8211
Dehaze-AGGAN [8]	45	60M	24.11 ǀ 0.8356
UVCGAN [9]	60	68M	26.31 ǀ 0.8641
Proposed	48	56M	**28.31** ǀ **0.8806**

In comparison, our proposed method has a parameter count of 56M, slightly higher than CycleGAN, but lower than UVCGAN. However, it achieves the highest PSNR value of 28.31 and SSIM value of 0.8806 among all the compared methods. This indicates that our method achieves superior performance in terms of image quality while maintaining a reasonable parameter count, showcasing its efficiency and effectiveness in single image dehazing tasks.

We test the processing efficiency of our model on the NVIDIA Jetson AGX Xavier L4T, a widely used edge computing platform with the following specifications: an 8-core ARM v8.2 64-bit CPU running at up to 2.26 GHz, a Volta GPU with 512 CUDA cores and 64 Tensor Cores running at 1.37 GHz, and 32 GB of LPDDR4x memory. This setup allows us to evaluate our model's performance and scalability in real-world edge computing scenarios, ensuring it can handle the demanding processing requirements typical of satellite operations. Table 7 shows that FPS increases from 3.43 (without SKIPAT) to 14.35 (with SKIPAT) and further to 18.18 (with SKIPAT and TensorRT). These results highlight the significant performance improvements achieved through combined optimizations, demonstrating the model's enhanced efficiency for real-time satellite data processing on the NVIDIA Jetson AGX Xavier L4T. The FPS of 18.18 achieved by the fully optimized model ensures efficient and timely data handling, crucial for real-time satellite operations. Thus, the optimized model is well-suited for applications demanding high processing speeds.

Table 7. Comparison of FPS for proposed method without SKIPAT and proposed method with TensorRT accelerate, on NVIDIA Jetson AGX Xavier L4T.

Model Name	FPS
Proposed (without SKIPAT)	3.43
Proposed (with SKIPAT)	14.35
Proposed (with SKIPAT and TensorRT)	18.18

4.8. Testing the Dehazing Effect at Different Angles

To further validate the effectiveness of RT Loss, we have conducted additional experiments to test the dehazing performance at different angles. We have rotated the images in the dataset at various angles (e.g., 0, 90, 180, 270 degrees) and evaluated the dehazing results using the models trained with and without RT Loss.

Table 8 shows the average PSNR and SSIM results for different angles with and without RT Loss. The data show that the model trained with RT Loss consistently achieves higher PSNR and SSIM values across all angles tested, including 0, 90, 180, and 270 degrees. Specifically, without RT Loss, the model's performance degrades significantly at rotated angles, with the PSNR dropping from 28.10 at 0 degrees to 26.90 at 270 degrees. Similarly, SSIM decreases from 0.8674 to 0.8612. In contrast, the model with RT Loss maintains more stable and higher performance, with the PSNR only slightly fluctuating around 28.31 and the SSIM consistently above 0.8765 for all angles. This demonstrates that RT Loss not only improves the overall dehazing quality, but also enhances the model's robustness

to rotational variations, ensuring consistent dehazing performance regardless of image orientation.

Table 8. Average PSNR and SSIM results for different angles.

Methods	Metrics	PSNR	SSIM
Without RT Loss	0 degree	28.10	0.8674
	90 degree	27.48	0.8654
	180 degree	27.19	0.8676
	270 degree	26.90	0.8612
With RT Loss	0 degree	28.31	0.8806
	90 degree	27.99	0.8768
	180 degree	28.21	0.8801
	270 degree	27.83	0.8765

4.9. Visual Comparison of Dehazing Results

To further illustrate the effectiveness of our proposed method, we provide a visual comparison of the dehazing results.

Figure 14 shows the visual comparison of the dehazing results. We can observe that our proposed method effectively removes the haze from the images, revealing clear and detailed features that were previously obscured by haze. These results demonstrate the superior dehazing performance of our method, highlighting its potential for enhancing the quality and applicability of remote sensing imagery across critical domains such as environmental monitoring, land management, and target detection.
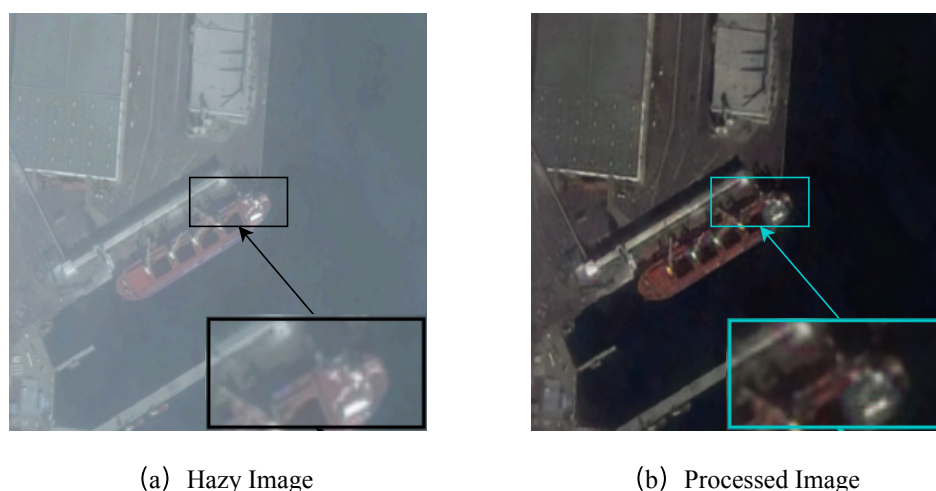

(a) Hazy Image   (b) Processed Image

Figure 14. Visual comparison of dehazing results. (a) Original hazy images. **(b)** Dehazed images by our method.

5. Conclusions

This paper introduces SPRGAN, a network that was specifically created for unpaired dehazing tasks on remote sensing images. It employs the Spatial-Spectrum Attention (SSA) mechanism with Skip-Attention (SKIPAT) to restore hazy images across a variety of domains. Furthermore, Perlin Noise-Based Masks (PNM) are introduced to effectively simulate foggy conditions, enhancing the model's super-resolution and dehazing capabilities during pre-training. Additionally, the integration of Rotational Loss (RT Loss) into the architecture further enhances the model's ability to handle haze-related challenges. This study showcases the efficacy of SPRGAN in improving image clarity and applicability in remote sensing applications, highlighting its potential for advancing the field of unpaired image dehazing.

The Spatial-Spectrum Attention with Skip-Attention (SSA-SKIPAT) mechanism, coupled with Perlin Noise Masks (PNM), has shown robust performance in enhancing image dehazing for remote sensing, particularly under haze and cloud occlusion scenarios. SSA-SKIPAT's effectiveness in capturing both spectral and spatial features highlights its potential for broader applications in image processing tasks requiring detailed feature extraction and context understanding. Future research directions include exploring multimodal fusion applications, adaptive SSA-SKIPAT variants for dynamic environmental conditions, transfer learning across datasets, and further efficiency optimizations of PNM and SSA-SKIPAT integration. This discussion not only reinforces the method's current strengths, but also charts a path for future innovations in remote sensing and related image processing domains.

There are several avenues for future research from our work. Firstly, while SPRGAN shows impressive performance in unpaired single image dehazing, further improvements are possible. Exploring new architectures and training strategies could increase the efficiency and effectiveness of the model in dealing with complex haze conditions. In addition, applying our approach to other domains, such as medical imaging or underwater photography, could provide new insights and broaden its utility. Furthermore, addressing computational constraints by optimizing model parameters and exploring parallel computing techniques will be crucial for deploying SPRGAN in real-world scenarios where computational resources may be limited. Finally, extending our analysis to evaluate the robustness and generalization capabilities of SPRGAN across different environmental conditions and sensor modalities will be essential for its adoption in practical applications. In conclusion, the future offers promising opportunities to further develop unpaired image dehazing and to exploit its potential in various domains.

Data Availability Statement: Data are contained within the article.

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

References

1. Zhao, T.; Wang, Y.; Li, Z.; Gao, Y.; Chen, C.; Feng, H.; Zhao, Z. Ship Detection with Deep Learning in Optical Remote-Sensing Images: A Survey of Challenges and Advances. *Remote Sens.* **2024**, *16*, 1145. [CrossRef]
2. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vision Graph. Image Process.* **1987**, *39*, 355–368. [CrossRef]
3. Arici, T.; Dikbas, S.; Altunbasak, Y. A histogram modification framework and its application for image contrast enhancement. *IEEE Trans. Image Process.* **2009**, *18*, 1921–1935. [CrossRef] [PubMed]
4. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [CrossRef] [PubMed]
5. Zhu, Q.; Mai, J.; Shao, L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **2015**, *24*, 3522–3533. [PubMed]
6. Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; Yang, M.H. Single image dehazing via multi-scale convolutional neural networks. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part II 14; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 154–169.
7. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
8. Zheng, Y.; Su, J.; Zhang, S.; Tao, M.; Wang, L. Dehaze-AGGAN: Unpaired remote sensing image dehazing using enhanced attention-guide generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]
9. Torbunov, D.; Huang, Y.; Yu, H.; Huang, J.; Yoo, S.; Lin, M.; Viren, B.; Ren, Y. Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 702–712.

10. Berman, D.; Treibitz, T.; Avidan, S. Non-local image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1674–1682.

11. Berman, D.; Treibitz, T.; Avidan, S. Air-light estimation using haze-lines. In Proceedings of the 2017 IEEE International Conference on Computational Photography (ICCP), Stanford, CA, USA, 12–14 May 2017; pp. 1–9.

12. Makarau, A.; Richter, R.; Müller, R.; Reinartz, P. Haze detection and removal in remotely sensed multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5895–5905. [CrossRef]

13. Wei, J.; Wu, Y.; Chen, L.; Yang, K.; Lian, R. Zero-shot remote sensing image dehazing based on a re-degradation haze imaging model. *Remote Sens.* **2022**, *14*, 5737. [CrossRef]

14. Xu, G.; Zhang, B.; Chen, J.; Wu, F.; Sheng, J.; Hong, W. Sparse Inverse Synthetic Aperture Radar Imaging Using Structured Low-Rank Method. *IEEE Trans. Geosci. Remote Sens.* **2021**. [CrossRef]

15. Guo, J.; Yang, J.; Yue, H.; Tan, H.; Hou, C.; Li, K. RSDehazeNet: Dehazing network with channel refinement for multispectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 2535–2549. [CrossRef]

16. Guo, J.; Yang, J.; Yue, H.; Hou, C.; Li, K. Landsat-8 OLI Multispectral Image Dehazing Based on Optimized Atmospheric Scattering Model. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 10255–10265. [CrossRef]

17. Shen, D.; Liu, J.; Wu, Z.; Yang, J.; Xiao, L. ADMM-HFNet: A Matrix Decomposition-Based Deep Approach for Hyperspectral Image Fusion. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–17. [CrossRef]

18. Yuan, J.; Cai, Z.; Cao, W. TEBCF: Real-World Underwater Image Texture Enhancement Model Based on Blurriness and Color Fusion. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]

19. 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]

20. 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.

21. Zheng, J.; Liu, X.Y.; Wang, X. Single image cloud removal using U-Net and generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **2020**. [CrossRef]

22. 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.

23. Li, R.; Pan, J.; Li, Z.; Tang, J. Single image dehazing via conditional generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8202–8211.

24. Qu, Y.; Chen, Y.; Huang, J.; Xie, Y. Enhanced pix2pix dehazing network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8160–8168.

25. Ren, W.; Ma, L.; Zhang, J.; Pan, J.; Cao, X.; Liu, W.; Yang, M.H. Gated fusion network for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3253–3261.

26. Zhang, X. Research on Remote Sensing Image De-haze Based on GAN. *J. Signal Process. Syst.* **2021**, *94*, 305–313. [CrossRef]

27. Tian, X.; Li, K.; Wang, Z.; Ma, J. VP-Net: An Interpretable Deep Network for Variational Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]

28. Zhu, Z.; Luo, Y.; Wei, H.; Li, Y.; Qi, G.; Mazur, N.; Li, Y.; Li, P. Atmospheric light estimation based remote sensing image dehazing. *Remote Sens.* **2021**, *13*, 2432. [CrossRef]

29. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.

30. Kim, T.; Cha, M.; Kim, H.; Lee, J.K.; Kim, J. Learning to discover cross-domain relations with generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1857–1865.

31. Chen, X.; Chen, S.; Xu, T.; Yin, B.; Peng, J.; Mei, X.; Li, H. SMAPGAN: Generative Adversarial Network-Based Semisupervised Styled Map Tile Generation Method. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4388–4406. [CrossRef]

32. Liang, X.; Zhang, H.; Xing, E.P. Generative semantic manipulation with contrasting gan. *arXiv* **2017**, arXiv:1708.00315.

33. Chen, X.; Xu, C.; Yang, X.; Tao, D. Attention-gan for object transfiguration in wild images. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 164–180.

34. Tang, H.; Liu, H.; Xu, D.; Torr, P.H.; Sebe, N. Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks. *arXiv* **2019**, arXiv:1911.11897.

35. 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; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

37. Venkataramanan, S.; Ghodrati, A.; Asano, Y.M.; Porikli, F.; Habibian, A. Skip-attention: Improving vision transformers by paying less attention. *arXiv* **2023**, arXiv:2301.02240.

38. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.

39. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [CrossRef] [PubMed]

40. Lin, D.; Xu, G.; Wang, X.; Wang, Y.; Sun, X.; Fu, K. A remote sensing image dataset for cloud removal. *arXiv* **2019**, arXiv:1901.00600.
41. Engin, D.; Genç, A.; Kemal Ekenel, H. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 825–833.
42. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
43. Ultralytics. Ultralytics YOLOv8. Available online: https://docs.ultralytics.com/ (accessed on 21 September 2023).
44. Yang, Y.; Wang, X.; Song, M.; Yuan, J.; Tao, D. Spagan: Shortest path graph attention network. *arXiv* **2021**, arXiv:2101.03464.
45. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.