Spatial–Frequency Fusion Network With Learnable Fractional Fourier Transform for Remote Sensing Imaging Enhancement

Wenyu Xu, Maohan Liang, Yuxu Lu, Ruobin Gao, *Member, IEEE*, and Dong Yang

***Abstract*—Atmospheric haze significantly degrades the quality of remote sensing images, reducing visibility, distorting spectral information, and impairing downstream tasks such as land cover classification and infrastructure layout analysis. To overcome these challenges, this article proposes a novel spatial–frequency fusion network (termed SFFNet) with a learnable fractional Fourier transform for efficient remote sensing imaging enhancement. In the spatial domain, the SFFNet uses a multiscale spatial pyramid pooling block to capture both fine-grained details and global contextual information, while residual connections ensure robust feature learning and spatial detail preservation. In the frequency domain, a self-learned fractional Fourier transform module adaptively extracts haze-relevant features, leveraging a learnable parameter to dynamically adjust the fractional order of the transform. Furthermore, an attentive frequency gate selectively emphasizes critical frequency features based on the local features of the input image. To effectively address the challenges of nonuniform haze distribution, a self-attention-guided fusion mechanism is introduced, synergistically integrating spatial and frequency information. In addition, a hierarchical feature fusion strategy progressively refines multiscale features throughout the dehazing process, ensuring comprehensive and accurate haze removal. Experimental results on both synthetic and real-world remote sensing datasets show that the SFFNet achieves significant improvements in quantitative metrics and visual quality. Moreover, the SFFNet demonstrates strong practical potential in remote sensing object detection by improving accuracy and robustness.**

***Index Terms*—Deep network, fractional Fourier transform, imaging enhancement, remote sensing, spatial–frequency fusion.**

I. Introduction

HAZE substantially degrades the quality of remote sensing images, leading to diminished visibility, distorted spectral information, and impaired scene interpretability [1], [2]. The degradation of image quality presents considerable challenges for applications such as land cover classification, remote sensing object detection [3], sea–land port segmentation [4], and environmental monitoring [5]. As shown in Fig. 1, remote sensing images differ from natural images by capturing large-scale landscapes that often feature complex atmospheric interactions, including nonuniform haze distribution and varying haze densities [6]. The distinctive features of remote sensing visual data demand specialized methods that can handle atmospheric interference at satellite imaging scales.

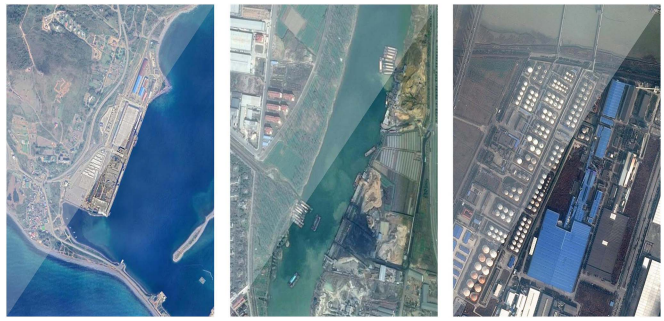

Fig. 1. Examples of scene restoration from three real-world degraded remote sensing images of maritime ports. The upper triangles in three images are degraded patterns, and the corresponding restored patterns by our SFFNet are shown in the lower triangles.

Traditional methods mainly rely on physical imaging models, such as the atmospheric scattering model (ASM) [7]. Specifically, building on the ASM, the dark channel prior (DCP) [8] has demonstrated promising performance by leveraging natural scene statistics, where small patches in haze-free images often have at least one color channel with very low intensity. However, DCP-based [9], [10] methods often suffer from issues such as artifacts, overdehazing in sky or sea regions, and limited effectiveness in handling complex lighting conditions, depth variations, and noise interference [11]. To address these limitations, alternative physical model-based methods [12], [13], [14] have integrated algorithms, such as edge-preserving filtering and low-rank decomposition, to improve image clarity and computational efficiency. Nevertheless, traditional methods remain constrained by their dependence on carefully designed priors, which may lack robustness and generalization capability in diverse remote sensing scenarios.

Received 10 April 2025; revised 8 June 2025; accepted 25 June 2025. Date of publication 3 July 2025; date of current version 25 July 2025. This work was supported by the Research Grants Council of Hong Kong under Grant PolyU 15201722. *(Corresponding author: Dong Yang.)*

Wenyu Xu, Yuxu Lu, and Dong Yang are with the Department of Logistics and Maritime Studies and the Maritime Data and Sustainable Development Centre, The Hong Kong Polytechnic University, Hong Kong (e-mail: wendy.xu@connect.polyu.hk; yuxulouis.lu@connect.polyu.hk; dong.yang@polyu.edu.hk).

Maohan Liang is with the Department of Civil and Environmental Engineering, National University of Singapore, Singapore 119077 (e-mail: mhliang@nus.edu.sg).

Ruobin Gao is with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: gaor0009@163.com).

Digital Object Identifier 10.1109/JSTARS.2025.3585939

Learning-driven methods excel in image restoration by modeling complex degradations and recovering fine details. Specifically, early methods [15], [16], [17], [18] integrated physical models with convolutional neural networks (CNNs) to estimate haze parameters or generate haze-free images directly. Building on these foundations, Transformer-based methods [6] have achieved better performance by extracting long-range dependencies and utilizing multiscale feature representations. However, they are often associated with high computational costs and depend on large-scale labeled datasets, which may not always be available for remote sensing imagery. In addition, while few-shot [19], semisupervised [20], and unsupervised learning [21] methods attempt to address the limitations of labeled data, they frequently introduce artifacts and struggle to match the performance of supervised methods, particularly in complex remote sensing scenes.

To address the challenges of atmospheric haze in remote sensing images, we propose a novel spatial–frequency fusion network (SFFNet) for adaptive and robust imaging enhancement. The SFFNet integrates spatial-domain enhancement, frequency-domain modeling, and spatial–frequency fusion into the framework, effectively addressing the limitations of existing methods. Specifically, the SFFNet captures multiscale spatial features using a spatial pyramid pooling block (SPPB) with residual connections to preserve spatial details, while a self-learned fractional Fourier transform module adaptively models haze-relevant frequency features. To combine spatial and frequency information effectively, a self-attention-guided fusion mechanism is proposed, enabling SFFNet to handle nonuniform haze distributions and achieve accurate haze removal. In addition, we propose a hybrid loss function to further improve the visual performance of the restored image. The contributions of our work are summarized as follows.

1) We propose a framework integrating spatial-domain enhancement, frequency-domain modeling, and spatial–frequency fusion, providing an adaptive and robust solution for effective haze removal in remote sensing imagery.
2) We propose a self-learned fractional Fourier transform module for adaptive frequency-domain feature extraction (FdE), coupled with a self-attention-guided fusion mechanism to effectively address nonuniform haze distributions.
3) Experiments on synthetic and real-world remote sensing datasets demonstrate SFFNet's performance, achieving better quantitative metrics and generating visually improved results with enhanced scene interpretability.

The rest of this article is organized as follows. In Section II, we systematically review related dehazing work. The imaging degradation model is given in Section III. Section IV is the principle of our SFFNet. Experimental details and results are provided in Section V. Finally, Section VI concludes this article and outlines future work.

II. Related Work

A. Traditional Methods

Traditional image dehazing methods mainly depend on physical models and handcrafted priors, leveraging assumptions about scene features and atmospheric conditions to estimate and remove haze. The DCP [8] is used to estimate transmittance by analyzing the statistical properties of dark pixels in natural scenes. While effective in many cases, the DCP is susceptible to introducing artifacts or overdehazing, especially in sky regions. To address these limitations, Yu and Liao [22] enhanced DCP using edge-preserving filters, which improves image clarity while minimizing artifacts. Other statistical model-based methods have also been proposed. For example, the color line prior method [23] estimates transmittance by analyzing the linear distribution of colors in degraded images. In contrast, Tan [24] focused on local contrast maximization to enhance visual clarity, though it often results in overenhancement. In addition, optimization-based strategies have been explored to improve dehazing performance. For example, Meng et al. [13] proposed edge-aware optimization of transmittance to preserve fine details, and Berman et al. [14] proposed the nonlocal dehazing method, which leverages the similarity of color distributions for more refined results. Liang et al. [25] proposed a remote sensing dehazing method using heterogeneous priors for robust atmospheric light and transmission estimation. For real-time applications, methods such as rank-one prior (ROP)+ [10] simplify dehazing by employing low-rank matrix decompositions, enabling faster processing. Meanwhile, Ancuti and Ancuti [26] integrated multiexposure imaging techniques to enhance image quality. However, traditional methods struggle to effectively manage complex scenes with nonuniform haze, varying lighting conditions, and significant depth variations in remote sensing imagery.

B. Learning-Driven Methods

Physical model-based learning methods combine an ASM with learnable parameters, using parameters such as transmittance and atmospheric light. For instance, DehazeNet [15] estimated the transmittance map via CNNs and integrated it with a physical model to generate haze-free images. Li et al. [16] proposed the all-in-one dehazing network (AOD-Net) to directly output haze-free images through a parameterized physical model, skipping separate parameter estimation. GridDehazeNet [27] enhanced adaptability to complex scenes by combining a physical model with a grid structure. Lihe et al. [28] combined residual learning and the ASM for efficient physics-aware haze removal. End-to-end learning methods bypass physical models, directly mapping hazy to haze-free images using deep learning frameworks like CNNs [29] or Transformers. Transformer-based methods excel in capturing long-range dependencies and improving dehazing via global modeling. For instance, PCSformer [6] integrated physical priors with the Transformer framework, balancing detail preservation and haze removal. Unsupervised and semisupervised methods are particularly effective when labeled data are scarce. Cycle-Dehaze [21] achieved unsupervised dehazing through cycle consistency loss but may introduce artifacts. The domain adaptation method [30] addressed data distribution gaps by transferring synthetic training results to real data. Liang et al. [31] proposed a self-supervised method to reduce reliance on labeled data using internal tasks like image reconstruction but lag behind supervised methods in image quality. Wang et al. [20]

proposed an unsupervised contrastive learning-based dehazing framework, which uses unpaired data and self-contrastive loss to solve domain shift and achieve efficient dehazing. However, existing learning-based dehazing methods often rely on extensive labeled datasets, struggle with artifact generation in unsupervised settings, and are challenging to generalize effectively to real-world scenarios.

C. Deep Learning With Fourier Transform

Frequency-domain features extracted via Fourier transform, combined with the nonlinear modeling of deep learning, improve image restoration by effectively preserving global structures and recovering fine details [32], [33]. Low-frequency components typically capture the global structural information of an image, while high-frequency components encode finer details and texture information. For example, Jiang et al. [34] proposed a frequency-domain generative adversarial network (GAN), which significantly outperformed traditional methods in super-resolution tasks by effectively enhancing high-frequency components. Several studies have proposed spatial–frequency joint modeling frameworks that integrate information from both spatial and frequency domains to improve restoration performance. For example, Liu et al. [35] combined density-guided transformers and frequency dual-path enhancement to effectively remove nonhomogeneous haze in remote sensing images. Zhou et al. [36] proposed an efficient image restoration method that leveraged the Fourier transform for global modeling, reducing computational complexity while enhancing performance. Zheng et al. [37] proposed the Transformer-guided cycle-consistent generative adversarial network (CycleGAN) framework, incorporating frequency-domain attention, semitransparent mask pretraining, and total variation loss to enhance remote sensing image dehazing. Sun et al. [38] integrated Fourier and wavelet-based heterogeneous enhancement to effectively fuse CNN and Transformer features, significantly improving imaging quality. Similarly, Wu et al. [39] proposed a frequency self-prompting method that dynamically utilizes frequency properties to guide a universal restoration network. However, efficiently extracting and seamlessly integrating spatial- and frequency-domain features from images using neural networks remains a significant challenge.

III. Physical Imaging Model

In remote sensing image processing, atmospheric scattering effects, such as haze and aerosol interference, significantly impact image quality. These effects cause signal attenuation and spectral distortion, obscuring surface details and reducing visibility [6]. The imaging model of remote sensing images can be described through the ASM [7], which considers the propagation features of light in the atmosphere. Based on the ASM, the hazy imaging model can be expressed as

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

where $I(x)$ represents the observed remote sensing image, $J(x)$ is the potentially clear image, A is the atmospheric light value, and $t(x)$ is the transmission rate. The first term $J(x) \cdot t(x)$ on the right-hand side of the equation represents the direct transmission component, describing the portion of scene-reflected light reaching the sensor after atmospheric attenuation. The second term $A \cdot (1 - t(x))$ represents the atmospheric scattered light component, describing the light intensity gain caused by atmospheric scattering. The transmission rate $t(x)$ is a key parameter describing the degree of light attenuation during atmospheric propagation and can be given as

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

where β is the atmospheric scattering coefficient, and $d(x)$ represents the distance from the scene point to the camera. The transmission rate has an exponential relationship with the scattering coefficient and propagation distance, reflecting the attenuation features of the atmosphere on light.

Haze variation in remote sensing is caused by spatially nonuniform atmospheric conditions, with the scattering coefficient β differing across regions [40]. In addition, terrain-induced distance variations $d(x)$ cause heterogeneous transmission rates $t(x) = e^{-\beta d(x)}$, unlike the uniform haze typically found in terrestrial settings. It leads to uneven image degradation, with longer atmospheric paths and multispectral distortions further complicating dehazing compared to land-based scenes.

IV. Spatial–Frequency Fusion Network

A. Overview

As shown in Fig. 2, our SFFNet is specifically proposed to address the challenges associated with integrating spatial and frequency information for efficient remote sensing image restoration. The architecture consists of three primary components, i.e., spatial-domain feature extraction (SdE), FdE, and spatial–frequency feature fusion (SFF). In addition, a hybrid loss function, which incorporates perceptual, frequency-aware, and reconstruction constraints, is used to further enhance dehazing performance while effectively preserving structural details.

B. Spatial-Domain Feature Extraction

Remote sensing images affected by haze are characterized by spatial degradation, including blurred edges, loss of fine-grained details, and a decline in structural consistency. These challenges arise from the inherent complexity of haze, which exhibits spatial nonuniformity and multiscale degradation patterns. To address these issues, we propose an SdE module that captures both local details and global contextual information. It is suggested through a similar SPPB [41], which processes the input feature map x using multiple convolutional branches with different receptive fields. The SPPB consists of multiple branches, which can extract features at different scales and are fused to form the spatial feature representation, i.e.,

$$F_{\text{SPPB}}(x) = F_1(x) + F_3(x) + F_3^{(3)}(x) + F_3^{(9)}(x) \tag{3}$$

where $F_1(x)$, $F_3(x)$, $F_3^{(3)}(x)$, and $F_3^{(9)}(x)$ represent the outputs of the 1×1, 3×3, 3×3 dilated (i.e., rate $= 3$), and 3×3 dilated (i.e., rate $= 9$) convolutions, respectively. The dilation rates of 3 and 9 are specifically chosen to progressively expand

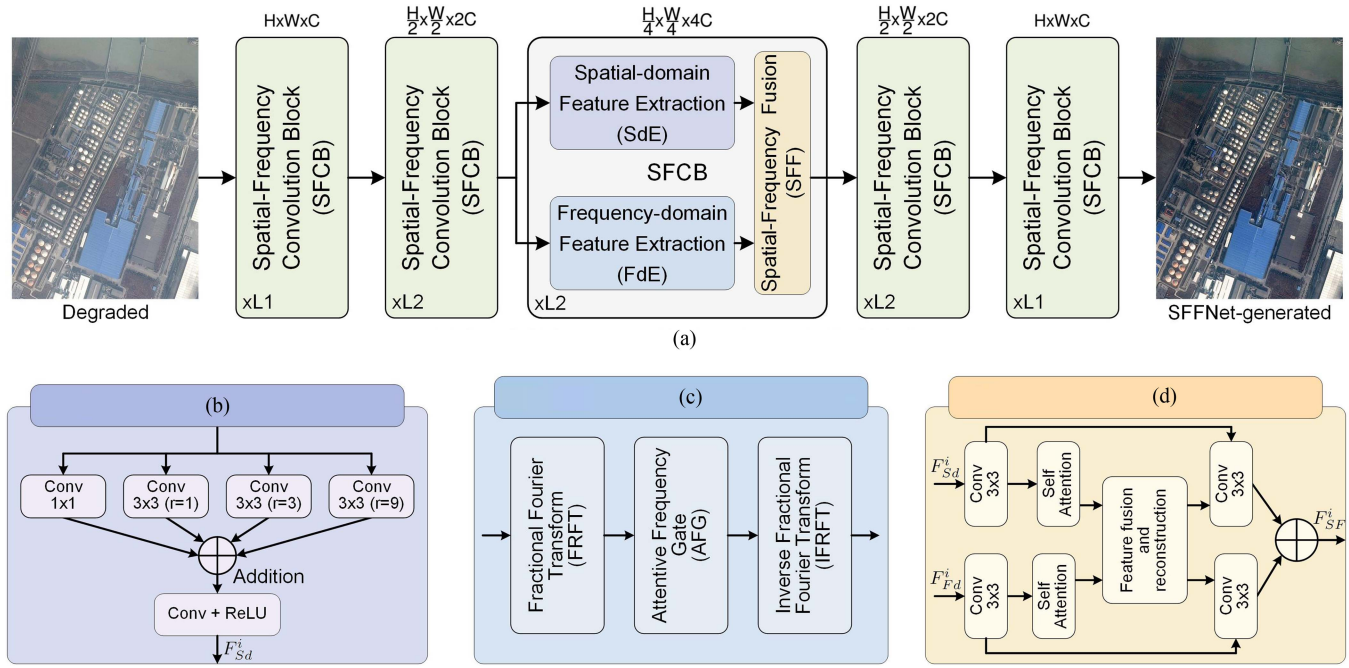

Fig. 2. (a) Overall architecture of the SFFNet. Spatial–frequency convolutional block (SFCB) is the main component of SFFNet and mainly consists of three parts: (b) spatial-domain feature extraction module, (c) frequency-domain feature extraction module, and (d) spatial–frequency feature fusion module.

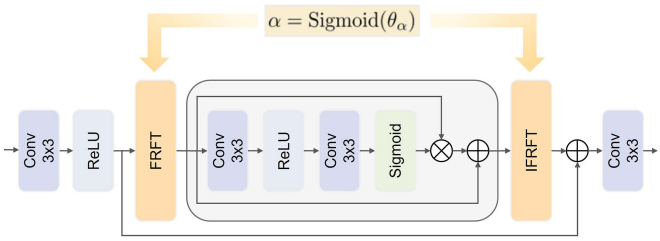

Fig. 3. Pipeline of self-learned FdE module. It mainly consists of three parts: FrFT for extracting frequency-domain information, AFG for enhancing important frequency features, and IFrFT for mapping frequency-domain features back to the spatial domain.

the receptive field, enabling the capture of medium-range and long-range spatial dependencies that are crucial for modeling the nonuniform haze distribution patterns commonly observed in large-scale remote sensing images.

To preserve the original spatial details, a residual connection is added, which can be given as

$$F_{\text{sd}} = \mathcal{R}(\mathcal{C}(F_{\text{SPPB}}(x)) + x) \tag{4}$$

where F_{sd} represents the final spatial-domain features, $\mathcal{R}$ is rectified linear unit, and $\mathcal{C}$ is a convolutional operation.

C. Frequency-Domain Feature Extraction

Remote sensing images contaminated by haze exhibit complex nonstationary frequency features across scales, which traditional Fourier analysis struggles to capture due to its focus on global frequencies. To address this, as shown in Fig. 3, we propose a self-learned FdE module that adaptively learns the optimal fractional order α for each input, leveraging fractional calculus to flexibly model multiscale localized haze features.

1) Self-Learned FrFT: Given the input feature map x, the α-order fractional Fourier transform is defined as

$$F_\alpha(x) = \int K_\alpha(t,u)x(t)dt \tag{5}$$

where the kernel function $K_\alpha(t,u)$ with scaling factor C_α, for $\alpha \neq 0, 1$, is formulated as

$$K_\alpha(t,u) = C_\alpha \exp[j\pi(t^2\cot(\alpha\pi/2) - 2tu\csc(\alpha\pi/2) + u^2\cot(\alpha\pi/2))] \tag{6}$$

where t represents the variable in the input domain (e.g., spatial coordinates), and u represents the variable in the output domain (i.e., fractional frequency domain). For $\alpha \in (0,1)$, this kernel enables a continuous interpolation between the time and frequency domains. When $\alpha = 0$, FrFT degenerates to the original signal, and when $\alpha = 1$, it becomes the conventional Fourier transform. The scaling factor C_α ensures energy preservation across transformations. The scaling factor C_α ensures energy preservation across transformations

$$C_\alpha = \sqrt{1 - j\cot(\alpha\pi/2)}. \tag{7}$$

Unlike prior methods that rely on fixed transform orders, our module incorporates a learnable fractional order α, which is dynamically optimized during training. Instead of preselecting fixed values for α, the network learns to determine the optimal fractional order for each input image. To constrain α to the range $[0,1]$, it is parameterized using a sigmoid function

$$\alpha = \text{Sigmoid}(\theta_\alpha) \tag{8}$$

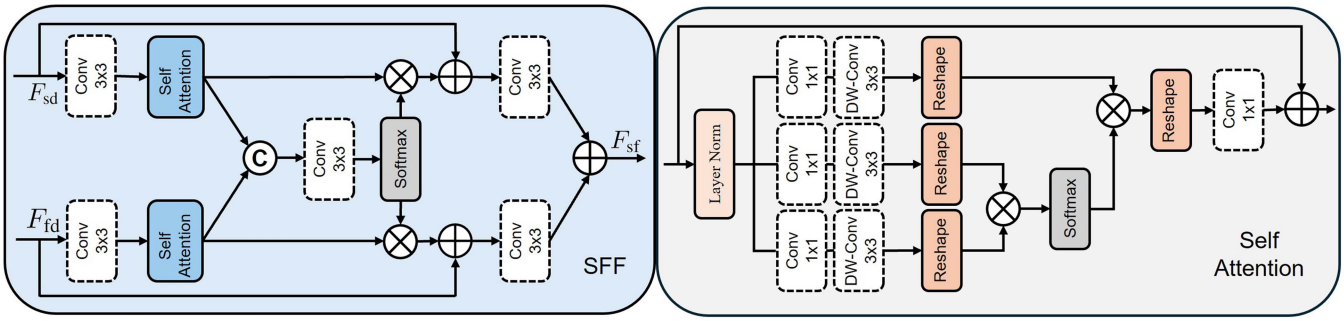

Fig. 4. Pipeline of SFF module. It leverages self-attention mechanisms in both spatial and frequency domains, followed by an adaptive fusion strategy.

where θ_α is a trainable parameter initialized to a reasonable value. Compared to the traditional Fourier transform, which maps signals strictly between the time domain ($\alpha = 0$) and the frequency domain ($\alpha = 1$), the fractional fourier transform (FrFT) offers a continuous interpolation between these domains through the fractional order $\alpha \in (0, 1)$. It allows FrFT to capture intermediate representations that combine both time- and frequency-domain features, providing greater flexibility in modeling haze patterns with spatially varying and scale-dependent features.

To separately analyze haze density and distribution features, we decompose the transformed feature into magnitude and phase components, i.e.,

$$F_\alpha(I) = |F_\alpha(I)| \odot \exp(j\phi_\alpha(I)) \tag{9}$$

where $|F_\alpha(I)|$ represents the magnitude spectrum reflecting haze density variations, and $\phi_\alpha(I)$ encodes spatial distribution information.

2) Attentive Frequency Gate: To further enhance relevant frequency components while preserving key image details via identity mapping, we propose an attentive frequency gate (AFG) module that dynamically modulates frequency components based on local image properties. The AFG module first transforms the input features X into the frequency domain using an FrFT with a learnable order $\alpha \in [0, 1]$. The real and imaginary parts are concatenated to form X_{frft}. To model complex inter-channel dependencies, we use a lightweight two-layer neural network with dimensionality reduction, i.e.,

$$z = W_2(\mathcal{R}(W_1(X_{\text{frft}}))) \tag{10}$$

where W_1 reduces channel dimensionality for computational efficiency, and W_2 restores channel dimensionality for attention weighting. The channel reduction ratio of 4 provides an optimal balance between model capacity and computational overhead. The channelwise attention weights are computed through sigmoid normalization. The final frequency-modulated features are obtained through

$$Y = X \odot \sigma(z) + X \tag{11}$$

where $\odot$ denotes elementwise multiplication. The residual connection ensures stable gradient flow during training while preserving original feature information. Through the adaptive modulation mechanism, our AFG module effectively enhances relevant frequency components while maintaining crucial image details through identity mapping.

D. Spatial–Frequency Feature Fusion

Hazy remote sensing images exhibit both spatial-domain degradation (e.g., blurred details and edge loss) and frequency-domain distortions (e.g., reduced contrast and uneven energy distribution). To address these challenges, we propose an SFF module, shown in Fig. 4. This module leverages self-attention-guided fusion mechanisms (SAMs) in both spatial and frequency domains. Given spatial-domain features F_{sd} and frequency-domain features F_{fd}, the SFF module first processes the inputs through initial convolution transformations. Then, the SAM is used independently to the spatial and frequency features and can extract refined features by capturing long-range dependencies within each domain, i.e.,

$$F_{\text{s}}^{\text{att}} = \text{SAM}_s(\mathcal{C}_s(F_{\text{sd}})) \tag{12}$$

$$F_{\text{f}}^{\text{att}} = \text{SAM}_f(\mathcal{C}_f(F_{\text{fd}})) \tag{13}$$

where F_s^{att} and F_f^{att} are the output attention feature maps. $\text{SAM}(\cdot)$ computes attention weights and refines the input features, which can be given as

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V \tag{14}$$

where Q, K, and V are the query, key, and value matrices, projected from F using learnable convolutional layers, respectively. The term $\sqrt{d}$ is a scaling factor, with d being the dimensionality of the query and key vectors. The attention output refines the input feature map by weighting the value representations based on the similarity between queries and keys.

After obtaining the self-attending features, the spatial and frequency features are concatenated and passed through a fusion layer, i.e.,

$$F_{\text{fusion}} = \text{Softmax}(\mathcal{C}_{sf}([F_{\text{s}}^{\text{att}}, F_{\text{f}}^{\text{att}}])) \tag{15}$$

where $[\cdot, \cdot]$ denotes channelwise concatenation. The fused features are then used to enhance the original spatial and frequency features via elementwise multiplication and residual connections, which can be given as

$$F_{\text{s}}' = \mathcal{C}_{s'}(F_{\text{s}}^{\text{att}} \odot F_{\text{fusion}} + F_{\text{sd}}) \tag{16}$$

$$F'_{\mathrm{f}} = \mathcal{C}_{f'}(F_{\mathrm{f}}^{\mathrm{att}} \odot F_{\mathrm{fusion}} + F_{\mathrm{fd}}). \tag{17}$$

Finally, the enhanced spatial and frequency features are combined to generate the output F_{sf} of the SFF module, i.e.,

$$F_{\mathrm{sf}} = F'_{\mathrm{s}} + F'_{\mathrm{f}}. \tag{18}$$

The SFF module is essential as spatial features capture fine-grained local details, while frequency features model global haze patterns, providing complementary information necessary for haze removal in complex remote sensing scenes.

E. Encoder–Decoder Architecture

Our network uses an encoder–decoder framework with skip connections [42], where the proposed SdE, FdE, and SFF are strategically integrated. The encoder E progressively reduces spatial dimensions while increasing feature channels, enabling the network to extract both local and global haze patterns. The decoder D gradually recovers spatial details through transposed convolutions. The overall architecture is given as

$$F_e^l = E_l(F_e^{l-1}, \mathrm{SFFNet}_l(F_e^{l-1})), \quad l \in [1, L] \tag{19}$$

where F_e^l represents the encoded features at level l, and L denotes the total number of encoding levels. The decoder operates symmetrically to the encoder through skip connections

$$F_d^l = D_l([F_d^{l+1}, F_e^l]), \quad l \in [L-1, 1] \tag{20}$$

where $[\cdot,\cdot]$ denotes channelwise concatenation, and F_d^l represents the decoded features at level l. This multiscale architecture enables our network to process haze features at different spatial scales through the encoding–decoding process, maintain fine details via skip connections, and progressively fuse and refine features during reconstruction.

F. Hybrid Loss Function

We propose a multiscale hybrid loss function for image dehazing that combines supervision from pixel $\mathcal{L}_{L1}$, perceptual $\mathcal{L}_{\mathrm{per}}$, and frequency domains $\mathcal{L}_{\mathrm{fre}}$. Our loss function is motivated by the observation that effective dehazing requires both local detail preservation and global contrast enhancement. The total loss $\mathcal{L}_{\mathrm{total}}$ is given as

$$\mathcal{L}_{\mathrm{total}} = \lambda_1 \mathcal{L}_{L1} + \lambda_p \mathcal{L}_{\mathrm{per}} + \lambda_f \mathcal{L}_{\mathrm{fre}} \tag{21}$$

where λ_1, λ_p, and λ_f are weighting coefficients set to 1.0, 0.1, and 0.1, respectively, through empirical validation.

1) Pixelwise $L1$ Loss: The $L1$ loss provides fundamental supervision in pixel space; mathematically, we have

$$\mathcal{L}_{L1} = \frac{1}{N}\sum_{i=1}^{N} |I_{\mathrm{re}}^i - I_{\mathrm{gt}}^i| \tag{22}$$

where I_{re} and I_{gt} denote the predicted dehazed image and ground truth, respectively. Unlike $L2$ loss, $L1$ loss exhibits better convergence properties and is more robust to outliers, making it particularly suitable for dehazing where local intensity variations can be significant. The linear penalization of $L1$ loss also helps preserve sharp edges and prevent oversmoothing in the dehazed results.

2) Perceptual Loss: To capture high-level semantic information and structural features, we use a perceptual loss based on VGG16 network [43] features, i.e.,

$$\mathcal{L}_{\mathrm{per}} = \sum_l \frac{\alpha_l}{C_l H_l W_l} \|\phi_l(I_{\mathrm{re}}) - \phi_l(I_{\mathrm{gt}})\|_1 \tag{23}$$

where ϕ_l represents features from layer l of the VGG16 network, and α_l are layer-specific weights. The multilayer feature extraction provides hierarchical supervision: lower layers capture local textures and patterns, while higher layers encode semantic information. This hierarchical supervision is crucial for maintaining perceptual quality in regions with varying haze densities.

3) Frequency-Domain Loss: To ensure proper recovery of different frequency components, we suggest a frequency-domain loss, which can be given as

$$\mathcal{L}_{\mathrm{fre}} = \mathcal{L}_{\mathrm{mag}} + 0.5\mathcal{L}_{\mathrm{phase}} \tag{24}$$

where the magnitude spectrum loss is defined as

$$\mathcal{L}_{\mathrm{mag}} = \|\mathbf{W} \odot \log(1 + |\mathcal{F}(I_{\mathrm{re}})|) - \log(1 + |\mathcal{F}(I_{\mathrm{gt}})|)\|_1 \tag{25}$$

where $\mathcal{F}$ denotes the 2-D Fourier transform and $\mathbf{W}$ is a frequency weighting matrix suggested to emphasize different frequency components

$$\mathbf{W} = 0.3W_{\mathrm{low}} + 0.5W_{\mathrm{mid}} + 0.2W_{\mathrm{high}} \tag{26}$$

where W_{low}, W_{mid}, and W_{high} are binary masks in the frequency domain, defined using radial frequency $f = \sqrt{f_x^2 + f_y^2}$, where $(f_x, f_y) \in [-1, 1]^2$. Specifically, $W_{\mathrm{low}} = 1$ for $f \le 0.2$, $W_{\mathrm{mid}} = 1$ for $0.2 < f \le 0.6$, and $W_{\mathrm{high}} = 1$ for $f > 0.6$, otherwise 0, emphasizing low-frequency global contrast, mid-frequency structures, and high-frequency details, respectively.

The phase spectrum loss complement magnitude supervision is given as

$$\mathcal{L}_{\mathrm{phase}} = \|\angle\mathcal{F}(I_{\mathrm{re}}) - \angle\mathcal{F}(I_{\mathrm{gt}})\|_1 \tag{27}$$

where $\angle\mathcal{F}$ represents the phase spectrum operation. Phase information is crucial for preserving structural coherence and edge alignment in the restored image.

V. Experiments and Discussions

In this section, we provide an overview of the experimental setup, including training and testing datasets, the experimental platform, the evaluation metrics, and the competitive methods. We assess the performance of SFFNet against competitive methods on both remote scene-related datasets and standard benchmarks. Furthermore, we conduct an extensive series of ablation studies to validate the contributions of individual network modules. Finally, we explore the application of SFFNet to advanced vision tasks, accompanied by an analysis of its runtime efficiency and computational complexity.

A. Implementation Details

1) Datasets: The limited availability of real-world paired data (i.e., clear and low-visibility images) poses a significant challenge to the training of learning-based image restoration

TABLE I
DETAILS OF TRAINING AND TESTING DATASETS USED IN OUR EXPERIMENTS

Datasets	Train	Test	RSS	Paired	Real
CDD-11 [44]	1183	200		✓	
RESIDE [45]	0	200		✓	
DOTA [46]	1140	200	✓	✓	✓

RSS: remote sensing scene.

networks. To overcome this limitation, we utilize the dataset for object detection in aerial images (DOTA) [46] and the composite degradation dataset (CDD-11) [44] to synthesize low-visibility images. To evaluate the robustness and generalization performance of our proposed method, we incorporate the classic image dehazing benchmark dataset, i.e., realistic single image dehazing (RESIDE) [45]. We also select real remote sensing hazy images from the DOTA dataset to verify the performance of the proposed method in practical applications. More detailed information can be found in Table I.

2) Experimental Platform: The network is trained for 50 epochs using the Adam optimizer with an initial learning rate of 0.001. The learning rate is reduced by a factor of 0.1 every 15 epochs to ensure effective convergence. All experiments are conducted in a Python 3.9 environment utilizing the PyTorch framework. The training process is performed on a high-performance PC equipped with an Intel(R) Core(TM) i9-12900K CPU @ 2.30 GHz and an Nvidia GeForce RTX 4090 GPU, enabling accelerated computations.

3) Evaluation Metrics: To quantitatively assess the effectiveness of visibility enhancement, we utilize both referenced and no-referenced evaluation metrics. Reference-based metrics, which require a ground truth image for comparison, include peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), feature similarity index (FSIM), visual saliency-induced index (VSI), and lightness order error (LOE). These metrics measure the fidelity of the enhanced image in terms of signal clarity and structural similarity. For most of these metrics (e.g., PSNR, SSIM, FSIM, and VSI), higher values indicate superior quality, while for LOE, lower values denote better performance as it measures the consistency of lightness order. No-reference metrics, which do not require a reference image, include the natural image quality evaluator (NIQE) and the perceptual image quality evaluator (PIQE). These metrics assess the perceptual quality based on intrinsic image features, with lower values indicating better quality.

4) Competitive Methods: To assess the restoration performance on remote sensing images, we will compare SFFNet against several state-of-the-art methods, which include traditional methods such as DCP [8], ROP [10], luminance and dark channel prior (LDCP) [47], and contrast enhancement and exposure fusion (CEEF) [48], as well as learning-based methods like multi-scale convolutional neural network (MSCNN) [49], DehazeNet [15], AODNet [16], all-in-one image restoration network (AirNet) [50], DeFormer [51], all-in-one scene recovery network (AoSRNet) [52], and compensation atmospheric scattering model (CASM) [53]. To ensure fairness and impartiality, all implementations are sourced directly from the original code provided by the respective authors.

B. Synthetic DOTA Degradation

1) Quantitative Analysis: To evaluate the performance of our SFFNet, we select 200 synthesized degraded images from the DOTA dataset. According to Table II, the SFFNet ranks first in all indicators compared with other competing methods. In particular, the two reference evaluation indicators (i.e., SSIM and PIQE) have achieved obvious performance improvements. Some methods often result in overenhancement or poor lighting consistency. Learning-based methods such as MSCNN, AirNet, and CASM may perform well on specific metrics but lack overall robustness compared to traditional methods such as DCP and ROP. In contrast, our SFFNet demonstrates balanced and comprehensive performance across all metrics, effectively restoring image details, enhancing structural fidelity, and improving perceptual quality.

2) Qualitative Analysis: To compare the visual performance of our SFFNet, six synthetic hazy images are selected from the DOTA dataset. As shown in Fig. 5, the hazy images capture in the real world are generally low in brightness and contrast, which destroys the texture details of the image. DCP, LDCP, AirNet, and CEEF can enhance the contrast and clarity of the image, but there is color distortion and overexposure. Although the color distribution of ROP, MSCNN, and DehazeNet is close to the actual value, the image is locally blurred and the dehazing effect is not obvious enough. The colors after learning-based methods are somewhat color-degraded, the image is oversmoothed, and the edge texture information is locally lost. Our SFFNet effectively achieves dehazing without compromising the underlying texture or color features, delivering visual results that closely resemble a real clear image.

C. Standard Dataset Degradation

1) Quantitative Analysis: We provide the average results of each method on the CDD-11 and RESIDE datasets in Table III. DCP-generated results have relatively low PSNR and SSIM, indicating low similarity with the real image. The images processed by DeFormer and CEEF have the lowest PSNR and SSIM, indicating poor restored image quality. Combining the two datasets, LDCP and AoSRNet have relatively excellent PSNR and SSIM results. The quantitative evaluation results of CASM on the two datasets are insufficient, which may be mainly due to its insufficient generalization ability for such scenarios. Our method achieves the best quantitative results on both datasets, fully demonstrating the superior robustness and effectiveness of our SFFNet in remote sensing image dehazing compared to other methods.

2) Qualitative Analysis: We show the visual results of all methods in Fig 6. The DeFormer-generated results show obvious black shadows and significant color distortion, resulting in relatively poor quality. While the image processed by AirNet achieves better dehazing performance, there are still slight haze residues in some complex scenes. The AoSRNet-generated results are whiter overall and lose a lot of texture details.

TABLE II
COMPARISON OF DEHAZING RESULTS (MEAN ± STD) WITH PSNR, SSIM, FSIM, VSI, LOE, NIQE, AND PIQE ON DOTA-BASED SYNTHETIC DATASETS

	PSNR ↑	SSIM ↑	FSIM ↑	VSI ↑	LOE ↓	NIQE ↓	PIQE ↓
DCP [8]	20.39±3.02	0.851±0.066	0.934±0.031	0.975±0.012	414.81±171.03	3.25±0.79	7.81±2.73
ROP [10]	16.51±3.77	0.811±0.111	0.912±0.056	0.968±0.021	413.57±158.77	3.24±0.78	8.19±2.61
LDCP [47]	17.39±4.92	0.740±0.219	0.896±0.066	0.962±0.024	570.14±331.70	3.16±0.79	7.92±2.61
CEEF [48]	20.70±2.30	0.812±0.054	0.906±0.028	0.965±0.011	404.05±129.51	3.49±0.90	9.05±2.89
MSCNN [49]	16.75±3.46	0.876±0.070	0.968±0.016	0.988±0.006	369.21±170.13	3.28±0.81	7.68±3.22
DehazeNet [15]	13.99±2.49	0.806±0.082	0.936±0.028	0.979±0.012	379.32±185.14	3.32±0.86	7.07±3.74
AODNet [16]	21.30±2.40	0.908±0.042	0.925±0.025	0.982±0.011	432.27±180.26	3.31±0.83	9.35±3.09
AirNet [50]	21.61±3.21	0.922±0.065	0.977±0.014	0.992±0.005	371.86±181.61	3.60±0.80	6.92±2.82
DeFormer [51]	17.80±2.75	0.776±0.112	0.921±0.035	0.970±0.015	451.25±193.81	3.24±0.77	6.91±2.93
AoSRNet [52]	20.89±3.48	0.902±0.061	0.966±0.023	0.989±0.009	504.69±250.11	3.26±0.79	7.19±2.75
CASM [53]	18.57±3.43	0.884±0.064	0.955±0.019	0.983±0.006	393.12±183.10	3.55±0.84	7.64±3.41
SFFNet	**22.62±3.30**	**0.934±0.038**	**0.980±0.015**	**0.994±0.005**	**325.47±174.30**	**3.15±0.72**	**5.98±2.40**

The best results are in **bold**, and the second best are with underline.

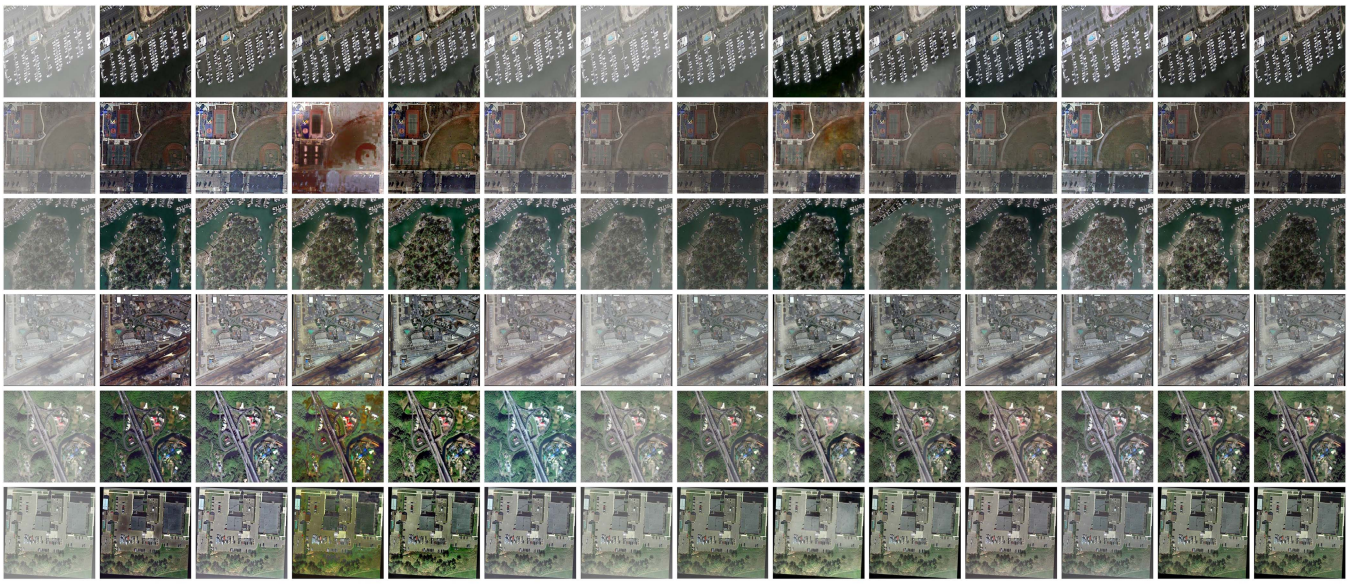

Fig. 5. Visual comparisons of dehazing results are presented for the DOTA dataset [46]. From left to right, the images include: (a) hazy input images, followed by restored images produced by (b) DCP [8], (c) ROP [10], (d) LDCP [47], (e) CEEF [48], (f) MSCNN [49], (g) DehazeNet [15], (h) AODNet [16], (i) AirNet [50], (j) DeFormer [51], (k) AoSRNet [52], (l) CASM [53], (m) our SFFNet, and (n) the corresponding ground truth.

TABLE III
COMPARISON OF (MEAN ± STD) VALUES OF PSNR AND SSIM ON STANDARD TEST DATASETS RESIDE [45] AND CDD-11 [44]

	Reside		CDD-11	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
DCP [8]	14.41±3.35	0.776±0.073	14.42±3.56	0.850±0.069
ROP [10]	19.08±3.60	0.865±0.059	17.83±3.11	0.899±0.056
LDCP [47]	19.64±3.50	0.869±0.060	18.91±3.74	0.892±0.066
CEEF [48]	13.80±2.59	0.789±0.067	13.32±2.48	0.798±0.066
MSCNN [49]	16.93±3.45	0.827±0.100	18.73±3.03	0.907±0.059
DehazeNet [15]	15.28±4.31	0.759±0.138	17.52±2.90	0.855±0.072
AODNet [16]	14.74±3.13	0.784±0.117	15.27±2.88	0.858±0.070
AirNet [50]	18.25±4.23	0.821±0.110	24.21±3.01	0.955±0.026
DeFormer [51]	17.60±4.32	0.850±0.095	12.82±2.56	0.766±0.096
AoSRNet [52]	20.56±4.57	0.893±0.067	19.05±4.22	0.896±0.086
CASM [53]	16.39±2.92	0.842±0.078	14.78±2.73	0.866±0.052
SFFNet	**21.33±4.21**	**0.905±0.056**	**25.32±2.87**	**0.961±0.027**

The best results are in **bold**, and the second best are with underline.

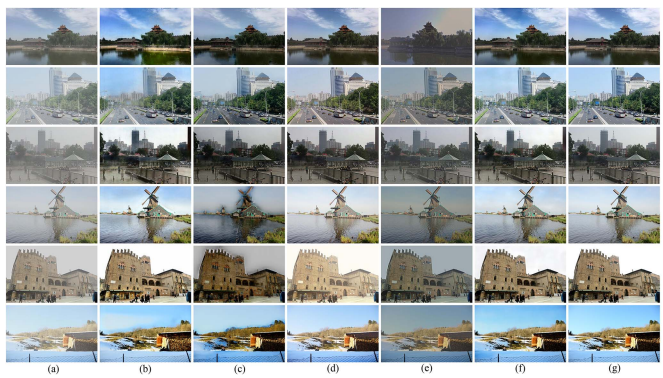

Fig. 6. Visual comparisons of dehazing results are presented for RESIDE [45] and CDD datasets [44]. From left to right, the images include: (a) hazy input images, followed by restored images produced by (b) AirNet [50], (c) DeFormer [51], (d) AoSRNet [52], (e) CASM [53], (f) our SFFNet, and (g) the corresponding ground truth.

Fig. 7. Visual results of our method for sandstorm image enhancement and underwater image enhancement.

Although CASM removes the interference of haze to a certain extent, the generated image has low contrast and looks unnatural. Compared with the original clear image, our method can handle challenging thick haze and nonuniform haze, produces fewer artifacts, and retains the texture details and local color information of the image. It further demonstrates the better dehazing generalization ability of our SFFNet under different scenarios.

D. Scenario Generalization Analysis

The proposed SFFNet's ability to generalize across tasks like sandstorm and underwater image enhancement, as shown in Fig. 7, is rooted in its spatial–frequency fusion framework, which handles diverse degradation patterns without fine-tuning. The SPPB captures multiscale spatial features, from local textures to global context, ensuring resilience to varying degradations like scattering in sandstorms or turbidity underwater. The self-learned FrFT module adaptively extracts frequency-domain features, addressing common degradation patterns (e.g., low-frequency contrast and high-frequency details). In addition, the self-attention-guided fusion mechanism dynamically integrates spatial and frequency features, prioritizing relevant information for different degradations. It allows the SFFNet to learn generalized feature representations, delivering enhanced contrast and fine details across diverse restoration tasks.

E. Real-World Degradation Analysis

Fig. 8 presents the qualitative comparison results of the real remote sensing hazy image. To highlight the restored details, a local region of the image has been enlarged. It is evident that the dehazing performance of DehazeNet and AirNet is insufficient in some cases, leaving noticeable haze residue in the images. While MSCNN demonstrates a slightly better dehazing effect than DehazeNet and AirNet, it often introduces significant color distortion. AODNet and DeFormer effectively remove haze but tend to darken the images, whereas AoSRNet causes a noticeable yellowish tint. The contrast of the image enhanced by convolutional block attention module (CBAM) is insufficient making the restored image darker. In contrast, our proposed SFFNet delivers superior dehazing performance by not only removing haze and restoring fine texture details but also maintaining excellent color fidelity. To further demonstrate the robustness of our method, as shown in Fig. 9, we also select real-world hazy remote sensing images from the DOTA dataset to show the restored visual effects.

TABLE IV
ABLATION STUDY ON SDE, FDE, AND SFF MODULES

BM	SdE	FdE	SFF	PSNR ↑	SSIM ↑
✓				20.22±3.59	0.898±0.047
✓	✓			20.57±3.44	0.918±0.043
✓		✓		20.16±3.37	0.911±0.051
✓	✓	✓		21.80±3.39	0.927±0.041
✓	✓	✓	✓	22.62±3.30	0.934±0.038

"BM" contains only the basic module.

TABLE V
ABLATION STUDY ON THE LEARNABLE FRACTIONAL FOURIER TRANSFORM

	PSNR ↑	SSIM ↑
Without Learnable Parameter	21.33±3.55	0.919±0.047
With Learnable Parameter	22.62±3.30	0.934±0.038

F. Ablation Study

1) Effect of Network Modules: To validate the significance of the proposed modules in our method for image restoration tasks, we conduct ablation experiments. As shown in Table IV, excluding any of these modules results in the lowest performance across all metrics. Specifically, without frequency-domain processing, the network struggles to capture frequency-specific details. The spatial domain-focused SPPB improves generalization by extracting spatial domain information. Dynamic parameter adjustment further enhances the network's adaptability. Incorporating SFF significantly boosts performance by integrating spatial and frequency-domain features, leveraging their complementary strengths. The full model, which combines all components, achieves the best results, demonstrating the synergistic effects of these modules.

2) Effect of Learnable FrFT: This section evaluates the dehazing performance of the learnable FrFT. As shown in Table V, the results demonstrate that learnable parameters significantly enhance the dehazing effect on remote sensing images. Combined with adaptive optimization of the learnable parameters, the model dynamically adjusts FdE, effectively capturing the complex characteristics of nonuniform haze. This leads to better separation of interference from target information, improving image quality and preserving structural fidelity.

3) Effect of Loss Function: To evaluate the effectiveness of the proposed hybrid loss function, we conducted a series of ablation experiments by systematically removing its individual components. As presented in Table VI, utilizing only the $\mathcal{L}_{L1}$ loss provides basic pixel-level supervision but yields the lowest performance. Incorporating $\mathcal{L}_{\text{per}}$ alongside $\mathcal{L}_{L1}$ significantly improves texture and edge preservation by introducing semantic guidance. Furthermore, the addition of $\mathcal{L}_{\text{fre}}$ enhances detail recovery and contrast by refining frequency-specific features. The complete loss function, which combines $\mathcal{L}_{L1}$, $\mathcal{L}_{\text{per}}$, and

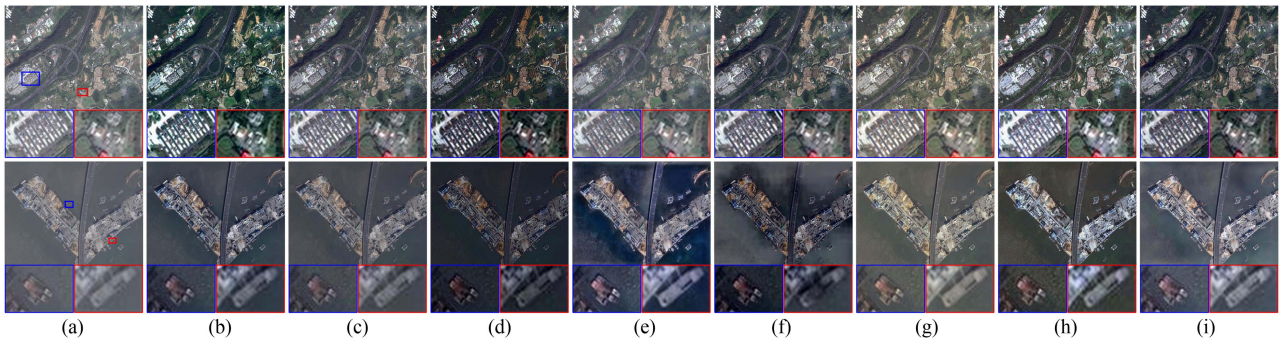

Fig. 8. Visual comparisons of dehazing results on real-world images. From left to right, the images include: (a) degraded images, followed by restored images produced by (b) MSCNN [49], (c) DehazeNet [15], (d) AODNet [16], (e) AirNet [50], (f) DeFormer [51], (g) AoSRNet [52], (h) CASM, and (i) our SFFNet.

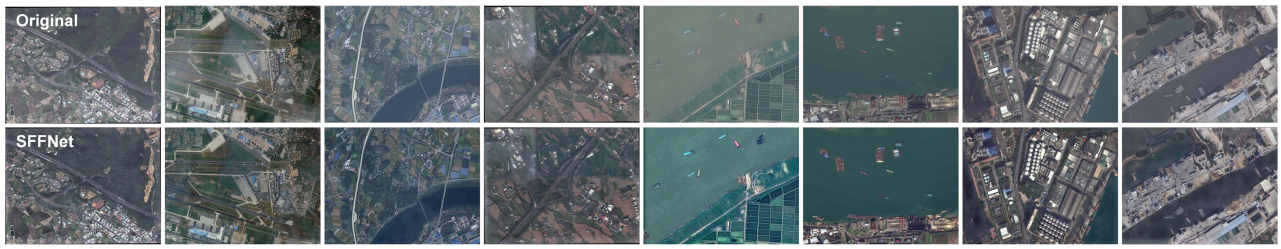

Fig. 9. Visual comparisons of real-world degraded images (top) and our enhancement results (bottom) for more examples from the DOTA [46] dataset.

TABLE VI
ABLATION STUDY ON THE LOSS FUNCTION COMPONENTS

$\mathcal{L}_{L1}$	$\mathcal{L}_{\text{per}}$	$\mathcal{L}_{\text{fre}}$	PSNR ↑	SSIM ↑
✓			19.32±4.24	0.904±0.046
✓	✓		21.47±4.08	0.913±0.043
✓		✓	22.05±3.82	0.919±0.041
✓	✓	✓	22.62±3.30	0.934±0.038

TABLE VII
REMOTE SENSING IMAGE DETECTION PERFORMANCE OF DIFFERENT METHODS UNDER DIFFERENT YOLO11 MODELS, EVALUATED USING MEAN AVERAGE PRECISION

	Nano	Small	Medium	Large	Extra Large
Haze	62.7	64.2	65.4	66.1	66.9
DCP [8]	71.5	72.3	72.9	73.2	73.5
AirNet [50]	73.7	74.2	74.7	75.4	75.9
SFFNet	76.7	77.5	78.0	78.3	78.8
Ground Truth	78.4	79.5	80.9	81.0	81.3

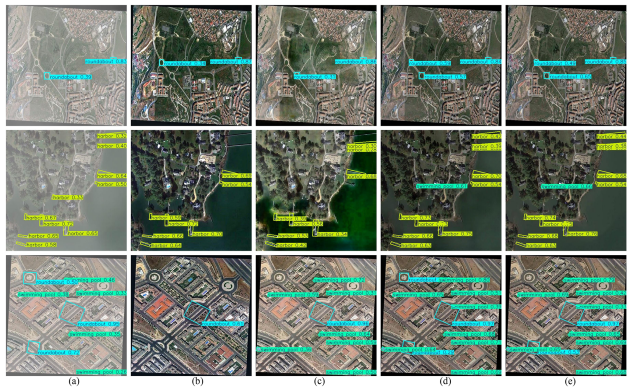

Fig. 10. Visual comparisons of dehazing results are presented for the DOTA dataset [46]. From left to right, the images include: (a) hazy input images, followed by restored images produced by (b) AirNet [50], (c) DeFormer [51], (d) AoSRNet [52], (e) our SFFNet, and (f) the corresponding ground truth.

$\mathcal{L}_{\text{fre}}$, achieves the highest performance, highlighting the complementary nature of these components in achieving superior image restoration results.

G. Improving High-Level Task

To demonstrate the advantages of SFFNet in remote sensing, we evaluate its impact on object detection using YOLOv11 [54] on objects like buildings, vehicles, and ships from the DOTA dataset. The results presented in Table VII demonstrate that the SFFNet consistently outperforms competing methods across all YOLOv11 model sizes, achieving higher mean average precision values even in challenging hazy conditions. Notably, SFFNet's performance approaches that of the ground truth, highlighting its effectiveness in enhancing image quality for robust object detection. As shown in Fig. 10, YOLOv11 struggles with hazy images due to reduced contrast and blurred details. Enhanced images processed by SFFNet significantly improve detection accuracy by revealing clearer features. Competing methods often fail in severe haze, losing fine details or introducing color distortions, which hinder detection performance. In contrast, the SFFNet generates images with superior clarity and structural consistency, enabling robust detection even in dense and nonuniform haze. These results highlight SFFNet's ability to enhance remote sensing imagery for high-accuracy object detection, preserving critical details and improving interpretability.

TABLE VIII
MODEL SIZE (UNIT, KB), RUNTIME (UNIT, SECONDS), AND FLOPS (UNIT, G) COMPARISON OF VARIOUS METHODS ON TWO DIFFERENT SIZE DATASETS OF 1920×1080 AND 2560×1440

Methods	Language	Size (KB)	1920×1080		2560×1440	
			Time	Flops	Time	Flops
DCP [8]	Matlab (C)	—	2.966	—	5.284	—
ROP+ [10]	Matlab (C)	—	1.503	—	2.831	—
LDCP [47]	Matlab (C)	—	2.642	—	4.729	—
CEEF [48]	Matlab (C)	—	2.774	—	5.096	—
MSCNN [49]	Matlab (C)	627	1.545	16.534	2.977	29.390
DehazeNet [15]	Matlab (C)	31	7.351	16.938	13.277	30.170
AODNet [16]	Python (G)	9	0.025	0.900	0.043	1.601
AirNet [50]	Python (G)	35393	0.255	111.315	0.421	197.893
DeFormer [51]	Python (G)	18960	0.155	95.433	0.273	169.658
AoSRNet [52]	Python (G)	4207	0.050	53.738	0.102	95.565
CASM [53]	Python (G)	32	0.313	5.201	0.427	9.246
SFFNet	Python (G)	18894	0.184	215.787	0.447	383.622

H. Computational Efficiency Analysis

Table VIII summarizes the model size, runtime, and computational complexity comparisons of various methods at resolutions of 1920×1080 and 2560×1440. Traditional CPU-based methods show significantly longer processing times compared to GPU-accelerated learning-based approaches. Our proposed SFFNet achieves competitive runtime performance through GPU acceleration. While the SFFNet exhibits higher computational complexity (FLOPs) than other methods, this reflects its sophisticated architecture that prioritizes comprehensive feature extraction for superior dehazing quality over minimal computational cost.

VI. CONCLUSION

In this article, we propose SFFNet, an innovative deep learning framework for addressing the challenges of haze removal in remote sensing images. By integrating spatial-domain enhancement, frequency-domain modeling, and spatial–frequency fusion, the SFFNet effectively overcomes the limitations of traditional and existing dehazing methods. Key contributions include the self-learned fractional Fourier transform module, which adaptively extracts frequency-domain features, and the self-attention-guided fusion mechanism, which enables robust integration of spatial and frequency information to handle nonuniform haze distributions. Extensive experimental results on synthetic and real-world datasets verify the advantages of SFFNet in terms of both quantitative metrics and visual quality.

The SFFNet excels in haze removal for remote sensing images but faces limitations due to its reliance on synthetic datasets, which may hinder generalization to diverse real-world scenarios, particularly regarding the model's sensitivity to different levels or nonuniform distributions of haze. Its computational complexity, though optimized, could challenge real-time use on resource-limited devices.

Future work will focus on improving real-world adaptability through more robust learning methods to address varying haze conditions, optimizing efficiency for edge deployment, and extending SFFNet to other restoration tasks or multimodal data for broader applications. In addition, we will delve deeper into the integration of image enhancement and advanced vision tasks through joint learning frameworks, aiming to systematically quantify the impact of enhancement techniques on the performance of high-level vision models.

REFERENCES

[1] C. Li et al., "Efficient dehazing method for outdoor and remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4516–4528, 2023.

[2] Q. Liu, T. Song, A. Qin, Y. Liu, F. Yang, and C. Gao, "HDSA-Net: Haze density and semantic awareness network for hyperspectral image dehazing," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 3989–4003, 2025.

[3] Y. Li, Q. Hou, Z. Zheng, M.-M. Cheng, J. Yang, and X. Li, "Large selective kernel network for remote sensing object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 16794–16805.

[4] B. Zhang, Y. Chen, W. Dang, S. Xiong, and X. Lu, "A spatial and semantic alignment fusion network for sea-land port segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 7420–7435, 2025.

[5] J. Li, Y. Pei, S. Zhao, R. Xiao, X. Sang, and C. Zhang, "A review of remote sensing for environmental monitoring in China," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1130.

[6] X. Zhang, F. Xie, H. Ding, S. Yan, and Z. Shi, "Proxy and cross-stripes integration transformer for remote sensing image dehazing," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5640315.

[7] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, pp. 233–254, 2002.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[9] J. Long, Z. Shi, and W. Tang, "Fast haze removal for a single remote sensing image using dark channel prior," in *Proc. IEEE Int. Conf. Comput. Vis. Remote Sens.*, 2012, pp. 132–135.

[10] J. Liu, R. W. Liu, J. Sun, and T. Zeng, "Rank-one prior: Real-time scene recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8845–8860, Jul. 2023.

[11] R. Zhang, Z. Cao, Y. Huang, S. Yang, L. Xu, and M. Xu, "Visible-infrared person re-identification with real-world label noise," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 5, pp. 4857–4869, May 2025.

[12] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[13] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE /CVF Int. Conf. Comput. Vis.*, 2013, pp. 617–624.

[14] D. Berman et al., "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1674–1682.

[15] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[16] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE /CVF Int. Conf. Comput. Vis.*, 2017, pp. 4770–4778.

[17] B. Ding et al., "U2D2Net: Unsupervised unified image dehazing and denoising network for single hazy image enhancement," *IEEE Trans. Multimedia*, vol. 26, pp. 202–217, 2023.

[18] S. Tian, Y. Wang, T. Zeng, and W. Zou, "Multi-level guided discrepancy learning for source-free object detection in hazy conditions," *IEEE Trans. Intell. Transp. Syst.*, early access, doi: 10.1109/TITS.2025.3559384.

[19] R. Zhang et al., "A benchmark and frequency compression method for infrared few-shot object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5001711.

[20] Y. Wang et al., "UCL-Dehaze: Toward real-world image dehazing via unsupervised contrastive learning," *IEEE Trans. Image Process.*, vol. 33, pp. 1361–1374, 2024.

[21] D. Engin, A. Genç, and H.Kemal Ekenel, "Cycle-Dehaze: Enhanced CycleGAN for single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 825–833.

[22] J. Yu and Q. Liao, "Fast single image fog removal using edge-preserving smoothing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 1245–1248.

[23] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 1–14, 2014.

[24] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[25] S. Liang, T. Gao, T. Chen, and P. Cheng, "A remote sensing image dehazing method based on heterogeneous priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5619513.
[26] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.
[27] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE /CVF Int. Conf. Comput. Vis.*, 2019, pp. 7314–7323.
[28] Z. Lihe, J. He, Q. Yuan, X. Jin, Y. Xiao, and L. Zhang, "PhDnet: A novel physic-aware dehazing network for remote sensing images," *Inf. Fusion*, vol. 106, 2024, Art. no. 102277.
[29] Y. Lu, Y. Guo, and M. Liang, "CNN-enabled visibility enhancement framework for vessel detection under haze environment," *J. Adv. Transp.*, vol. 2021, no. 1, 2021, Art. no. 5598390.
[30] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2808–2817.
[31] Y. Liang, B. Wang, W. Zuo, J. Liu, and W. Ren, "Self-supervised learning and adaptation for single image dehazing," in *Proc. Int. Joint Conf. Artif. Intell.*, 2022, pp. 1137–1143.
[32] Y. Cui, W. Ren, X. Cao, and A. Knoll, "Image restoration via frequency selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 2, pp. 1093–1108, Feb. 2024.
[33] Y. Cui, W. Ren, X. Cao, and A. Knoll, "Revitalizing convolutional network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, pp. 9423–9438, Dec. 2024.
[34] X. Jiang, X. Zhang, N. Gao, and Y. Deng, "When fast Fourier transform meets transformer for image restoration," in *Proc. IEEE Eur. Conf. Comput. Vis.*, 2025, pp. 381–402.
[35] H. Liu, J. Huang, J. Nie, J. Xie, L. Chen, and X. Zhou, "Density guided and frequency modulation dehazing network for remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 9533–9545, 2025.
[36] M. Zhou, J. Huang, C.-L. Guo, and C. Li, "Fourmer: An efficient global modeling paradigm for image restoration," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 42589–42601.
[37] Y. Zheng, J. Su, S. Zhang, M. Tao, and L. Wang, "Dehaze-TGGAN: Transformer-guide generative adversarial networks with spatial-spectrum attention for unpaired remote sensing dehazing," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5634320.
[38] H. Sun et al., "Bidirectional-modulation frequency-heterogeneous network for remote sensing image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, early access: doi: 10.1109/TCSVT.2025.3570998.
[39] Z. Wu, W. Liu, J. Wang, J. Li, and D. Huang, "FrePrompter: Frequency self-prompt for all-in-one image restoration," *Pattern Recognit.*, vol. 161, 2025, Art. no. 111223.
[40] Y. Gao, W. Xu, and Y. Lu, "Let you see in haze and sandstorm: Two-in-one low-visibility enhancement network," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5023712.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[42] R. Zhang et al., "Detail-aware network for infrared image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2024, Art. no. 5000314.
[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[44] Y. Guo, Y. Gao, Y. Lu, H. Zhu, R. W. Liu, and S. He, "OneRestore: A universal restoration framework for composite degradation," in *Proc. IEEE Eur. Conf. Comput. Vis.*, 2024, pp. 255–272.
[45] B. Li et al., "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
[46] G.-S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.
[47] Y. Zhu, G. Tang, X. Zhang, J. Jiang, and Q. Tian, "Haze removal method for natural restoration of images with sky," *Neurocomputing*, vol. 275, pp. 499–510, 2018.
[48] X. Liu, H. Li, and C. Zhu, "Joint contrast enhancement and exposure fusion for real-world image dehazing," *IEEE Trans. Multimedia*, vol. 24, pp. 3934–3946, 2021.
[49] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. IEEE Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.
[50] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17452–17462.
[51] Y. Song, Z. He, H. Qian, and X. Du, "Vision transformers for single image dehazing," *IEEE Trans. Image Process.*, vol. 32, pp. 1927–1941, 2023.
[52] Y. Lu, D. Yang, Y. Gao, R. W. Liu, J. Liu, and Y. Guo, "AoSRNet: All-in-one scene recovery networks via multi-knowledge integration," *Knowl.-Based Syst.*, vol. 294, 2024, Art. no. 111786.
[53] X. Wang et al., "Compensation atmospheric scattering model and two-branch network for single image dehazing," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 8, no. 4, pp. 2880–2896, Aug. 2024.
[54] G. Jocher and J. Qiu, "Ultralytics YOLO11," 2024. [Online]. Available: https://github.com/ultralytics/ultralytics

Wenyu Xu received the B.Sc. degree in information engineering from the School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing, China, in 2020. She is currently working toward the master's degree with the Hong Kong Polytechnic University, Hong Kong.

She is also a Research Assistant with the Hong Kong Polytechnic University. Her research interests include low-visibility image enhancement and machine learning.

Maohan Liang received the M.S. and Ph.D. degrees from the School of Navigation, Wuhan University of Technology, Wuhan, China, in 2018 and 2023, respectively.

He was an Exchange Student with Nanyang Technological University, Singapore, in 2021. He is currently a Research Fellow with the Department of Civil and Environmental Engineering, National University of Singapore, Singapore. His research interests include trajectory data mining, deep learning, and intelligent transportation systems.

Yuxu Lu received the B.E. degree in navigation technology and the M.S. degree in navigation and information engineering from the School of Navigation, Wuhan University of Technology, Wuhan, China, in 2020 and 2023, respectively. He is currently working toward the Ph.D. degree with the Hong Kong Polytechnic University, Hong Kong.

His research interests include computer vision, multimodal data fusion, and intelligent transportation systems.

Ruobin Gao (Member, IEEE) received the B.Eng. degree from Jilin University, Changchun, China, in 2017, the M.Sc. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2018, and the Ph.D. degree from the School of Civil and Environmental Engineering, Nanyang Technological University, in 2022.

He is currently with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China. His research interests include machine learning, randomized neural networks, and time-series forecasting.

Dr. Gao is an Associate Editor for *Computers and Electrical Engineering*.

Dong Yang received the Ph.D. degree in maritime logistics science from Kobe University, Kobe, Japan, in 2008.

He is currently an Associate Professor and an Associate Head of the Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong. He has held various academic positions, including Assistant Professor with the Southern University of Denmark, Odense, Denmark, and a Research Fellow with the Centre for Maritime Studies, National University of Singapore, Singapore. He has authored or coauthored more than 70 SCI/SSCI papers in more than 20 international academic journals. His current research interests include empirical study, big data analysis, multimodal data modeling and fusion, technology innovation in transportation, and trade and regional economics.

Dr. Yang is an Associate Editor for *International Journal of Shipping and Transport Logistics* and *Maritime Business Review* and a Guest Editor for several prestigious journals.