Journal Pre-proof

Bidomain multi-order modeling for image dehazing

Zihao Chen, Hao Shen, Chenxu Wu, Junling Li, Wei Wang, Wenqi Ren

PII: S0031-3203(26)01269-0
DOI: https://doi.org/10.1016/j.patcog.2026.114304
Reference: PR 114304

To appear in: *Pattern Recognition*

Received date : 11 March 2026
Revised date : 25 May 2026
Accepted date : 16 June 2026

Please cite this article as: Z. Chen, H. Shen, C. Wu et al., Bidomain multi-order modeling for image dehazing, *Pattern Recognition* (2026), doi: https://doi.org/10.1016/j.patcog.2026.114304.

Highlights

Bidomain Multi-order Modeling for Image Dehazing

Zihao Chen, Hao Shen, Chenxu Wu, Junling Li, Wei Wang, Wenqi Ren

- A novel spatial-frequency multi-order modeling framework is proposed as an efficient alternative to mainstream global modeling paradigms for image dehazing.
- The functional Spatial Interaction Unit and Frequency Aggregation Unit enable robust spatial and frequency feature learning with the aid of the Spatial-Frequency Uncertainty estimation.
- Experiments on multiple benchmark datasets demonstrate the superiority of our approach in terms of dehazing performance and computational overhead.

Bidomain Multi-order Modeling for Image Dehazing

Zihao Chen[a,1], Hao Shen[b,1], Chenxu Wu[a], Junling Li[a,*], Wei Wang[c,*], Wenqi Ren[c]

[a]*School of Information Science and Engineering, Southeast University, Nanjing 211189, China*
[b]*School of Public Security and Emergency Management, Anhui University of Science and Technology, Hefei 231131, China*
[c]*School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China*

Abstract

Though transformer-dominated global modeling designs have achieved impressive performance gains in single image dehazing task, they usually require a high memory footprint and computation budget compared with computationally efficient convolutional counterparts. In this work, we identify two key ingredients behind existing global modeling paradigms, i.e., high-order spatial interactions and channel evolution, and also take into account the distinguished frequency characteristics associated with haze degradation, thereby presenting a novel yet effective spatial-frequency (i.e., bidomain) multi-order modeling block, termed BMMB. Our BMMB is implemented based on a spatial-frequency multi-order modulation architecture, i.e., spatial multi-order modulation and frequency multi-order modulation units, in which 1) the spatial interaction is achieved through more efficient convolution coupled with element-wise dot-product operations; 2) followed by, a novel Fractional Fourier Transform (FRFT) is introduced to perform information aggregation instead of the commonly used FFN architecture within transformer; 3) moreover, we also introduce the spatial-frequency uncertainty that is capable of incorporating the degradation priors into feature learning within the above two process, achieving robust representation ability. Overall, the proposed BMMB is a general and plug-and-play design that can be easily extended to arbitrary

*Corresponding authors: Junling Li and Wei Wang.
Email addresses: 213233966@seu.edu.cn (Zihao Chen), haoshenhs@gmail.com (Hao Shen), 220250763@seu.edu.cn (Chenxu Wu), junlingli@seu.edu.cn (Junling Li), wangwei29@mail.sysu.edu.cn (Wei Wang), renwq3@mail.sysu.edu.cn (Wenqi Ren)
[1]Equal contribution.

high orders conditioned on the task requirements. Extensive experiments demonstrate that our method yields a favorable performance against other state-of-the art methods on multiple datasets including synthetic and real world ones. The code is available after possible acceptance. Extensive experiments demonstrate that our method achieves favorable performance against state-of-the-art methods on multiple synthetic and real-world dehazing datasets with competitive computational overhead. The code is available at https://anonymous.4open.science/r/BMMB-Dehazing-257A.

Keywords: Image dehazing, Global modeling, Bidomain, Multi-order

1. Introduction

Haze refers to a common meteorological phenomenon that exerts substantial influence on both our daily activities and machine vision systems. For example, haze often significantly diminishes visibility and impairs people's ability to perceive and identify objects in the surroundings, posing a potential hazard to traffic safety. In terms of computer vision tasks, hazy images usually affect model performance in other high-level vision applications, such as scene understanding [1] and object detection [2]. This is attributed to their degraded visual quality, which results in misleading assessments by machine vision systems.In light of the pressing demand for clean images in real-world vision tasks, image dehazing technology has garnered considerable attention from academic and industry communities [3, 4].

Over the past several years, deep learning technology has made a big hit in computer vision, demonstrating impressive performance improvements in low-level and high-level visual applications. In image dehazing, a flood of deep learning techniques, primarily built on convolutional neural networks (CNNs), have been proposed to learn the mapping relationship between clear images and corresponding hazy counterparts in an end-to-end manner [5–9]. Although CNN-based dehazing approaches have made remarkable progress compared with traditional algorithms, they are still bottlenecked by the limited receptive fields inherent in convolution, failing to model the global feature dependencies. Recently, transformer-based global modeling paradigms have become popular in image dehazing tasks, significantly outperforming CNN-based ap-

proaches [10–13]. However, there are some common limitations within these architectures: *1) compared with CNN-based counterparts, they often require substantial computational resources due to the dense self-attention computation in cascaded transformer designs; 2) they struggle to effectively capture local features, leading to inaccuracies in the reconstruction of fine-grained details.*

FSDGN [6] first presents a spatial-frequency dehazing framework, revealing that haze degradation primarily resides in the amplitude spectrum. This observation provides a frequency-domain prior for image dehazing. Following this, many spatial-frequency methods [14–17] have been proposed to improve dehazing performance by leveraging the complementary strengths of both domains: the spatial domain complements and refine local texture details, while the frequency domain captures global structural information. Despite their effectiveness, existing methods are unable to distinguish the reliability of different frequency components, hindering further improvement for frequency-sensitive image dehazing tasks. UDL-SR [18] is the first to incorporate spatial uncertainty into image dehazing through supervised information and feature extraction. It reveals the correlation between uncertainty and degradation, and shows that uncertainty estimation improves performance. However, although this spatial domain uncertainty models pixel-wise reliability by identifying unreliable predictions and severely degraded regions, it cannot directly reflect the restoration reliability of high-frequency components in the frequency domain, which is critical for image dehazing. In light of the aforementioned concerns, we contemplate whether it is possible to offer an efficient global modeling design that embraces the merits of both CNN and transformer, while considering image degradation priors. In light of the aforementioned concerns, we contemplate whether it is possible to offer an efficient global modeling design capable of embracing the merits of both CNN and transformer, while also considering the inherent priors associated with the image degradation process.

In this work, we delve into the key ingredients behind the prevailing global modeling paradigms illustrated in Fig. 1. Specifically, Fig. 1 (a) distills the general global modeling rule into two fundamental components: high-order spatial interaction and channel evolution, while Fig. 1 (b), (c) and (d) instantiates three prevailing alternatives. *However, these paradigms rarely explore higher-order interactions in both spatial and*

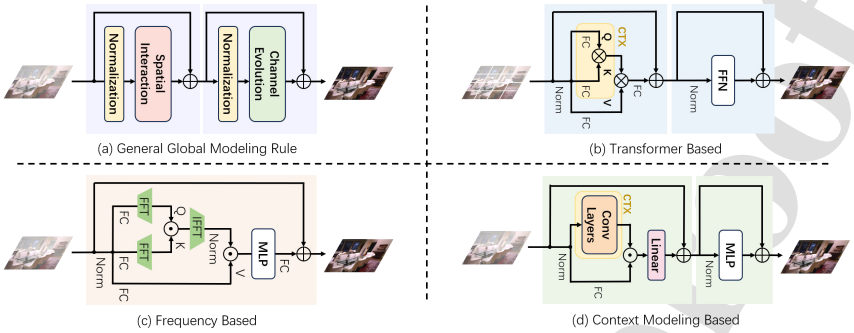


Figure 1: Illustration of (a) the general global modeling rule and three prevailing alternatives: (b) Transformer based [19, 20], (c) Frequency based [21], and (d) Context (CTX) modeling based [22, 23]. $\otimes$ and $\odot$ represent the matrix multiplication and element-wise multiplication, respectively; DFT and IDFT are the Fourier and inverse Fourier transform. These prevailing global modeling paradigms are mainly built upon the high-order spatial interactions in a single domain, simply neglecting the potential of higher-order interactions in both spatial and channel dimensions across diverse domains. FFN denotes the feed-forward network, which is typically implemented by MLP layers.

channel dimensions, as well as their potential across different domains. Herein, we propose an innovative yet effective spatial-frequency (referred to as bidomain) multi-order modeling block, termed BMMB, tailored for the dehazing task. Overall, the key insight of our method is to develop an efficient global modeling alternative, enabling the consideration of haze priors. Our core building design, BMMB, is implemented based on a spatial-frequency modeling architecture, including a Spatial Interaction Unit (SIU) and a Frequency Aggregation Unit (FAU). Inside BMMB, specifically, 1) SIU employs efficient convolutions coupled with element-wise dot-product operations to accomplish high-order spatial interactions, enabling it to capture hierarchical features encompassing global and local information. 2) Following, FAU introduces an innovative Fractional Fourier Transform (FRFT) to perform information aggregation, enabling it to achieve any intermediate feature representation between spatial and frequency domains. 3) Moreover, we also introduce the Spatial-Frequency Uncertainty (SFU), which is capable of incorporating the degradation priors into feature learning within the above two components, achieving robust feature representation. Extensive experiments on multiple widely recognized datasets demonstrate the applicability of our dehazing framework assembled by the proposed BMMB. In summary, the primary contributions of this study can be outlined as follows:

- We explore an innovative yet effective spatial-frequency multi-order modeling block (BMMB), which serves as an efficient alternative to mainstream global modeling paradigms tailored for dehazing learning.
- Inside our core design BMMB, the Spatial Interaction Unit (SIU) is capable of learning both global and local information; the Frequency Aggregation Unit (FAU) offers any intermediate feature representation between spatial and frequency domains, enabling more effective feature integration; the Spatial-Frequency Uncertainty (SFU) is introduced to achieve more robust representation ability and better interpretability within SIU and FAU.
- By incorporating spatial–frequency uncertainty, the network jointly models pixel-wise reliability and high-frequency recovery difficulty, thereby adaptively focusing on high-frequency details while reliably preserving low-frequency structures.
- Extensive experiments over multiple widely recognized datasets demonstrate that our dehazing framework yields favorable performance compared with other cutting-edge approaches with less computational overhead.

2. Related Work

2.1. Efficient Global Modeling.

Very recently, many efforts have been dedicated to alleviating some long-standing and challenging limitations inherent in transformer architecture [20, 24, 25] dominated global modeling designs, such as high memory footprint and computation budget. Currently, there are two promising strategies: the first solution is committed to overcoming the constraint of the Softmax function in self-attention computation, while the second research line focuses on exploring more efficient convolution alternatives. For example, Han *et al.*[26] propose a novel Focused Linear Attention module that replaces the conventional Softmax function with a simple mapping function. Qiu *et al.*[27] leverage the first-order Taylor expansion of self-attention, enabling the matrix multiplicative associate law, to reduce the computational complexity. Despite higher computational efficiency, they suffer from significant performance degradation compared to their Softmax counterparts. FocalNet [22] and HorNet [23] are two admirable techniques that

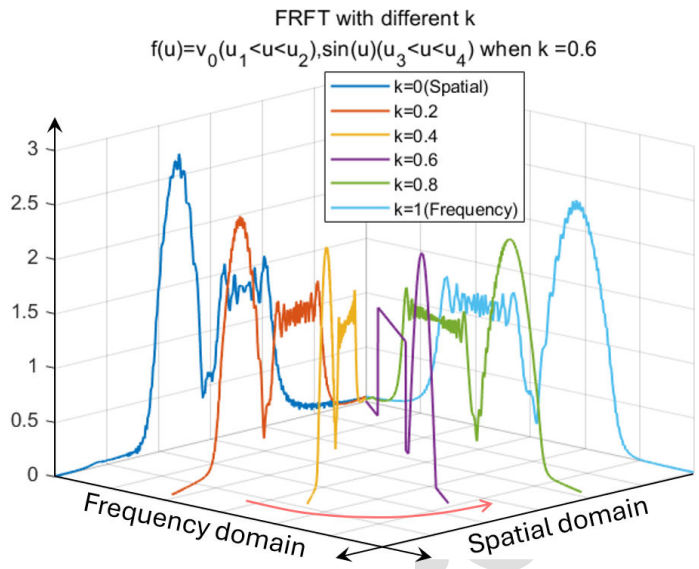


Figure 2: Illustration of the Fractional Fourier Transform (FRFT). We demonstrate the disentangling capability of the FRFT using a composite signal, which can be readily separated into a square wave and a partial sine wave after applying the FRFT with an order of $k = 0.6$ (indicated by the purple line).

customize the efficient convolution modules to mimic the spatial interactions within self-attention, presenting competitive performance with favorable computational overhead in comparison to their self-attention counterparts. By contrast, Zheng *et al.*[28] delve into high-order channel-wise interactions and put forward a novel Rubik's cube convolution operator. Although these high-order based on convolutional operations show promising performance, there is still room for exploration. For instance, the FFN used in transformer architecture is usually used to perform information aggregation, yet the intrinsic hierarchical characteristics of the high-order modeling mechanisms are ignored. *Besides, these higher-order operations are usually implemented within a single domain, with limited exploration of higher-order modeling across diverse domains.*

2.2. *Frequency Domain Image Restoration.*

Frequency information has gained widespread application [6, 14, 29, 30] in image restoration tasks, due to the innate image disentanglement abilities and global relationships of the Fourier transform. Hu *et al.*[6] propose a spatial-frequency net-

work for image dehazing by observing the amplitude-phase relationships between the haze image and the corresponding clear image. Kong *et al.*[21] devise an efficient frequency-based alternative to conventional self-attention based on the convolution theorem. Zhou *et al.*[14] propose an efficient Fourier global modeling framework for image restoration by analyzing the key ingredients behind the transformer architecture. SFIR [29] integrates a multi-scale spatial enhancement module and a frequency amplitude modulation module to jointly model spatial structures and frequency-domain amplitude information, leading to enhanced image restoration performance. Though demonstrating the promising performance of frequency information, existing methods rarely explore the intermediate disentangled feature representations between the spatial and frequency domains. Very recently, Hu *et al.*[30] introduce the Fractional Fourier Transform (FRFT) to attain continuous spatial and frequency representations of images. Given the flexibility of FRFT as illustrated in Fig. 2, we endeavor to integrate it with higher-order convolutions that involve hierarchical feature representation, thereby implementing an efficient multi-order spatial-frequency modeling.

2.3. Applications of Uncertainty in Image Restoration.

Recently, there has been a surge of interest in the deep learning and image restoration fields regarding Bayesian uncertainty. The trend is rooted in the recognition that uncertainty quantification transcends enhancing performance and provides a critical measure for assessing both the reliability and robustness of the restored images. For instance, Uncer2Natural [31] addresses the problem of unsupervised image denoising by estimating the aleatoric uncertainty in the noisy image and reducing its effect, enabling the model to focus on the degraded and textured regions. Ning *et al.*[18] propose a new adaptive weighted loss for image super-resolution, enhancing deep network training by prioritizing regions of high uncertainty. Liu *et al.*[32] propose a novel dual-domain learning framework to quantify spatial and spectral Bayesian uncertainty in image super-resolution. This enables a reliability assessment and reasoning from both the spatial and frequency domain perspective. However, the exploration of uncertainty applications in image dehazing is nascent. As far as we know, the [33] is the first work contributing novel methodologies and insights about the spatial uncertainty

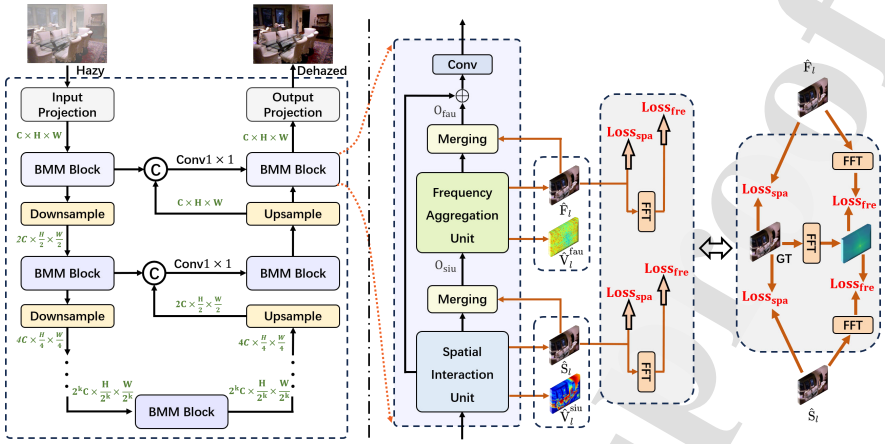


Figure 3: Overview of the proposed image dehazing framework. It is implemented using the key bidomain multi-order modeling (BMM) block within a fundamental U-Net architecture. Each BMM block comprises a spatial interaction unit (SIU) and a frequency aggregation unit (FAU), both integrated with a spatial-frequency constraint. $\hat{\mathrm{V}}_l^{\mathrm{siu}}$ and $\hat{\mathrm{V}}_l^{\mathrm{fau}}$ denote the variance of the features from SIU and FAU of the l-th BMM block, respectively, characterizing their prediction uncertainty.

in image dehazing. It exploits the relationship between the uncertain and confident representations and adaptively enhances the learned features, by estimating the aleatoric and epistemic uncertainty together. While no one has explored the frequency domain uncertainty in image dehazing, we will investigate the bidomain epistemic uncertainty in our work, which can help our module achieve more robust representation ability and better interpretability.

3. Methodology

As illustrated in Fig. 3, the proposed BMMB comprises three fundamental components: spatial interaction unit, frequency aggregation unit, and spatial-frequency uncertainty. In this section, we will first present the details of each core building design, then conclude the overall architecture, and finally, the related loss function adopted in our image dehazing experiments will be elaborated.

3.1. Spatial Interaction Unit

As commonly acknowledged, global information is paramount for pixel-level image restoration tasks, but local details should not be underestimated, especially concerning visual effects. In light of this, recent endeavors have been made to mimic

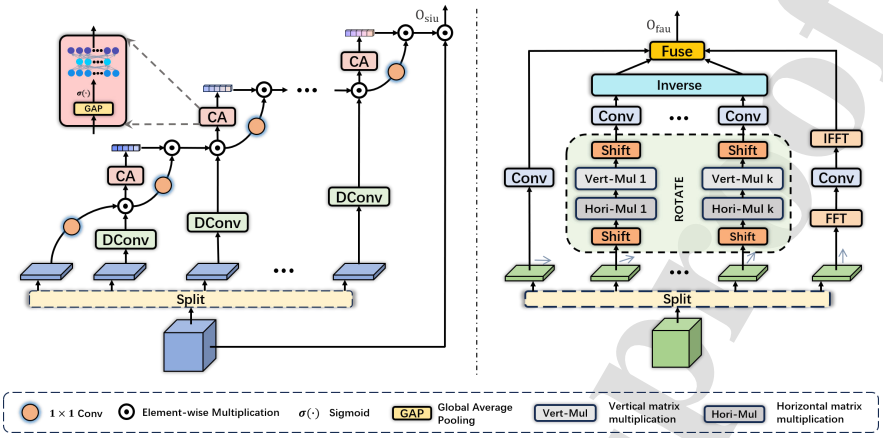


Figure 4: The detailed structure of the proposed spatial interaction unit (SIU) and frequency aggregation unit (FAU), where CA denotes channel attention.

the key designs within transformer architecture using basic convolution operations and element-wise multiplications, enabling it to model both global and local dependencies. Motivated by these promising designs, herein, we devise a novel yet effective convolution-based spatial interaction unit. However, unlike existing high-order convolution operators that are implemented only in spatial or channel dimensions, our design incorporates both the spatial-wise and channel-wise multi-order interactions.

As detailed in Fig. 4, given an input feature $\mathrm{X} \in \mathbb{R}^{H\times W\times C_{in}}$, we first evenly divide it into N groups along the channel dimension, which can be formulated as follows:

$$[\mathrm{g}_0, \mathrm{g}_1, \mathrm{g}_2, \cdots, \mathrm{g}_{\mathrm{N}-1}] = \mathrm{Split}(\mathrm{X}), \tag{1}$$

where $\mathrm{Split}(\cdot)$ denotes the channel split operation, and $\mathrm{g_i} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\frac{\mathrm{C_{in}}}{\mathrm{N}}}$, $\mathrm{i} \in \{0, 1, \cdots, \mathrm{N}-1\}$ is the i-th group. Subsequently, the first group is projected through a point-wise convolution, while the remaining $N-1$ parts are performed by depth-wise convolution (DConv), which can be written as:

$$\begin{aligned} \widetilde{\mathrm{g}}_0 &= \mathrm{Conv}_{1\times1}(\mathrm{g}_0), \\ \widetilde{\mathrm{g}}_\mathrm{k} &= \phi(\mathrm{g}_\mathrm{k}), \mathrm{k} \in \{1, \cdots, \mathrm{N}-1\}, \end{aligned} \tag{2}$$

where $\mathrm{Conv}_{1\times1}(\cdot)$ is the point-wise convolution, and $\phi(\cdot)$ denotes the DConv.

Next, let us introduce the basic operation within our design. Using $\widetilde{\mathrm{g}}_1$ and $\widetilde{\mathrm{g}}_2$ as

inputs, the corresponding 1-order spatial-channel operation can be obtained:

$$\widetilde{y}_1 = \mathrm{CA}(\widetilde{g}_0 \odot \widetilde{g}_1) \odot \mathrm{Conv}_{1\times1}(\widetilde{g}_0 \odot \widetilde{g}_1), \tag{3}$$

where $\odot$ is the element-wise multiplication, $\mathrm{CA}(\cdot)$ denotes the channel attention, and $\widetilde{y}_1$ can be regarded as the output of the 1-order spatial-channel interaction.

After obtaining the 1-order spatial-channel interaction, we can further employ it to achieve multi-order spatial-channel interactions, thereby enhancing the model expressiveness. Formally, the multi-order spatial-channel interactions can be successively expressed as follows:

$$\begin{aligned}
\widetilde{y}_2 &= \mathrm{CA}(\widetilde{y}_1 \odot \widetilde{g}_2) \odot \mathrm{Conv}_{1\times1}(\widetilde{y}_1 \odot \widetilde{g}_2),\\
\widetilde{y}_3 &= \mathrm{CA}(\widetilde{y}_2 \odot \widetilde{g}_3) \odot \mathrm{Conv}_{1\times1}(\widetilde{y}_2 \odot \widetilde{g}_3),\\
&\cdots,\\
\widetilde{y}_k &= \mathrm{CA}(\widetilde{y}_{k-1} \odot \widetilde{g}_k) \odot \mathrm{Conv}_{1\times1}(\widetilde{y}_{k-1} \odot \widetilde{g}_k),
\end{aligned} \tag{4}$$

Finally, we conduct the element-wise multiplication between the advanced feature generated by the multi-order spatial-channel interactions and the original input feature to obtain the output of the proposed spatial interaction unit:

$$\mathrm{O}_{\mathrm{siu}} = \widetilde{y}_k \odot \mathrm{X}, \tag{5}$$

where $\mathrm{O}_{\mathrm{siu}}$ is the output of the SIU module. Through recursive spatial–channel multi-order interactions, the SIU progressively expands its effective receptive field despite relying on local kernels. This enables it to simultaneously capture fine local details and aggregate broader contextual information, thereby endowing the module with strong representation capabilities.

3.2. Frequency Aggregation Unit

The Fractional Fourier Transform (FRFT), recognized as a potent signal processing technique, can achieve continuous spatial and frequency representations of images by manipulating the order k. Given a 2D signal $\mathrm{f}(\mathrm{x}, \mathrm{y}) \in \mathbb{R}^2$, its FRFT can be formulated as follows:

$$\mathrm{F}^{\mathrm{p}_1,\mathrm{p}_2}(\mathrm{u}, \mathrm{v}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{f}(\mathrm{x}, \mathrm{y})\mathrm{K}_{\mathrm{p}_1,\mathrm{p}_2}(\mathrm{x}, \mathrm{y}, \mathrm{u}, \mathrm{v})\mathrm{dxdy}, \tag{6}$$

where $F^{p_1,p_2}(u,v)$ represents the transformed signal, p_1 and p_2 denote the fractional transform orders along the horizontal and vertical directions, respectively, and $K_{p_1,p_2(x,y,u,v)}$ denotes the corresponding FRFT transform kernel. When $p_1, p_2 \neq 0, 1$, the kernel can be written as:

$$\begin{aligned} K_{p_1,p_2}(x,y,u,v) &= A_{\alpha,\beta} e^{\left(\frac{j(x^2+u^2)}{2tan\alpha} - \frac{jxu}{sin\alpha}\right)+\left(\frac{j(y^2+v^2)}{2tan\beta} - \frac{jyv}{sin\beta}\right)}, \\ A_{\alpha,\beta} &= \frac{\sqrt{(1-jcot\alpha)(1-jcot\beta)}}{2\pi}, \alpha = \frac{p_1\pi}{2}, \beta = \frac{p_2\pi}{2}, \end{aligned} \tag{7}$$

where $K_{0,0}, K_{0,1}, K_{1,0}, K_{1,1}$ denote δ functions, indicating that the signal exists in the single spatial domain or frequency domain. By employing the Wigner-Ville distribution, we can characterize the FRFT transformed signal as the rotation of this distribution in the spatial domain. Thanks to this, we adopt an alternative approach for the realization of the FRFT [34]. To be specific, we first shift the input features to interchange the central and edge portions. Following this, tailored matrix multiplication is applied independently to the left and right components. The left component executes vertical transformations, while the right component executes horizontal transformations, which can be expressed as follows:

$$Fr(\cdot) = Mat_V \otimes Shift(\cdot) \otimes Mat_H, \tag{8}$$

where $Fr(\cdot)$ represents the FRFT operator, $Shift(\cdot)$ donates the shift operation applied to the matrix, $\otimes$ indicates the matrix multiplication, Mat_V and Mat_H are the vertical and horizontal transform matrices, respectively.

Considering the flexibility of FRFT in continuous spatial-frequency feature representation, we devise an innovative frequency aggregation unit (FAU) by introducing the FRFT to conduct effective feature refinement. As displayed in Fig. 4, the output of the SIU is firstly separated into M parts along the channel dimension as follows:

$$[z_1, z_2, \ldots, z_M] = Split(O_{siu}), \tag{9}$$

where $z_i \in \mathbb{R}^{H \times W \times \frac{C_{in}}{M}}$, $i \in \{1, 2, \cdots, M\}$ denotes the *i*-th segment. Then, we employ a 3×3 convolution to address the first group in the spatial domain, and perform the Fourier transform on the last segment. The FRFT is applied to the middle segments to obtain the intermediate feature representation between spatial and frequency domains.

The detailed procedure is formulated as follows:

$$\begin{aligned}\widetilde{z}_1 &= z_1 = \mathrm{Fr}^0(z_1),\\ \widetilde{z}_k &= \mathrm{Fr}^{k-1}(z_k), k \in \{2, \cdots, M-1\},\\ \widetilde{z}_M &= \mathcal{F}(z_M) = \mathrm{Fr}^{M-1}(z_M),\end{aligned} \tag{10}$$

where $\mathrm{Fr}^k(\cdot)$ represents the FRFT operator with $p_1 = p_2 = \frac{k}{M-1} \cdot \frac{\pi}{2}$, and $\mathcal{F}(\cdot)$ is Fourier transform.

Following that, the point-wise convolution is employed to refine the disentangled feature information, which can be written as follows:

$$\widetilde{u}'_k = \sigma(\mathrm{Conv}_{1\times1}(\widetilde{z}_k)), k \in \{1, \cdots, M\}, \tag{11}$$

where $\sigma(\cdot)$ denotes the activation function, $\widetilde{u}'_k$ is refined features. Finally, we transfer all features back to spatial domain and concatenate them, with a convolution group to merge all these features before output, formulated as follows:

$$O_{fau} = \mathrm{Conv}(\sum_k (\mathrm{Trans}^{-1}(\widetilde{u}'_k))), k \in \{1, \cdots, M\}, \tag{12}$$

where O_{fau} is the output of the FAU, and $\mathrm{Trans}^{-1}(\cdot)$ represents the corresponding iFRFT (inverse FRFT) of each $\mathrm{Fr}^k(\cdot)$. Based on an adjustable transform order p, FAU flexibly achieves continuous and learnable feature aggregation from the spatial domain to the frequency domain. In synergy with the spatial-channel multi-order interactions of SIU, it effectively enhances the network's dehazing capability.

3.3. Spatial-Frequency Uncertainty

Bayesian deep learning frameworks are adept at capturing two distinct forms of uncertainty: aleatoric uncertainty, which originates from the inherent noise present within the observational data; and epistemic uncertainty, which represents the uncertainty within the model itself and stems from the lack of knowledge or the imperfectness of the designed model [35]. The former is irreducible and arises from data itself, such as sensor noise or variations in the hazing process, which cannot be controlled or reduced by improving the model. Thus in our work, we only consider the latter and further extend our proposed BMMB to the Bayesian framework.

Given training data $\mathbf{X} = \{\mathrm{I}_1, ..., \mathrm{I}_\mathrm{N}\}$, $\mathbf{Y} = \{\mathrm{I}_1^*, ..., \mathrm{I}_\mathrm{N}^*\}$, $\mathbf{X}$ is hazing data, and $\mathbf{Y}$ is the clear ground truth data. Bayesian learning aims to compute the posterior distribution over the plausible model parameters $\mathrm{p}(\theta|\mathbf{X}, \mathbf{Y})$ and find a model $\mathrm{B}(\mathbf{X}; \theta) \rightarrow \mathbf{Y}$:

$$\mathrm{p}(\theta|\mathbf{X}, \mathbf{Y}) = \frac{\mathrm{p}(\mathbf{Y}|\mathbf{X}, \theta)\mathrm{p}(\theta)}{\mathrm{p}(\mathbf{Y}|\mathbf{X})}, \tag{13}$$

Based on this, we can generate the dehazing output for a new input I' by integrating over all possible model parameters:

$$\mathrm{p}(\mathrm{I}'^* \mid \mathrm{I}', \mathbf{X}, \mathbf{Y}) = \int \mathrm{p}(\mathrm{I}'^* \mid \mathrm{I}', \theta)\mathrm{p}(\theta \mid \mathbf{X}, \mathbf{Y})\mathrm{d}\theta. \tag{14}$$

Nonetheless, the integral is intractable, and the posterior distribution cannot be assessed analytically either, leading to various existing approximations. Dropout variational inference is a practical approach for approximating inference in large models [36]. We utilize it to perform multiple stochastic forward computations and generate multiple Monte Carlo (MC) samples both in the training and inference phases. Subsequently, Bayesian uncertainty is deduced from the dispersion in predicted MC samples. While pixel-wise variance functions as the spatial uncertainty indicator, it is inappropriate for frequency uncertainty due to substantial variation in the dynamic range across frequencies. Instead, we adopt the Coefficient of Variation (CV) statistic [32, 37] as the measure of frequency uncertainty.

3.4. Overall Network and Optimization

To demonstrate the effectiveness of the proposed BMMB, our image dehazing framework is implemented by using a common multi-scale U-Net architecture, which possesses a L-level symmetric encoder-decoder structure. In particular, embedding only one proposed spatial-frequency multi-order modeling block at each level of the encoder-decoder empowers our model to attain competitive performance.

Overall, in addition to the final dehazed output $\hat{\mathrm{I}}_{\mathrm{out}}$, our U-Net [38] architecture will have two finite sampling distributions of the prediction at each intermediate level, *i.e.* $\{\hat{\mathrm{S}}_\mathrm{t}^l\}_{\mathrm{t}=1}^{\mathrm{T}}$ and $\{\hat{\mathrm{F}}_\mathrm{t}^l\}_{\mathrm{t}=1}^{\mathrm{T}}$, where $l = 1, ..., \mathrm{L}$. The former represents the outputs from SIU, and the latter is from FAU. For the two intermediate finite predicted distributions, we can easily compute the mean $\hat{\mathrm{S}}_l$ and $\hat{\mathrm{F}}_l$, variance $\hat{\mathrm{V}}_l^{\mathrm{siu}}$ and $\hat{\mathrm{V}}_l^{\mathrm{fau}}$, CV $\hat{\mathbb{V}}_l^{\mathrm{siu}}$ and CV $\hat{\mathbb{V}}_l^{\mathrm{fau}}$.

For better capturing the multi-scale details that are essential for reconstructing a clear image from a hazy one, and enabling the desired uncertainty to reflect the degradation priors, we supervise the intermediate predicted distributions and output simultaneously by the L_1 norm distance. In this manner, uncertainty estimation is explicitly involved in the intermediate prediction and supervision process:

$$L_s = \frac{\lambda_1}{L} \sum_{l=1}^{L} \|\hat{I}_l - I_l^*\|_1 + \|\hat{I}_{out} - I^*\|_1, \tag{15}$$

where $\hat{I}_l = \{\hat{S}_l, \hat{F}_l\}$ and I_l^* is the resized ground truth for each level intermediate output. Similarly, in the frequency domain, we adopt the following loss function:

$$L_f = \frac{\lambda_1}{L} \sum_{l=1}^{L} \|\mathcal{F}(\hat{I}_l) - \mathcal{F}(I_l^*)\|_1 + \|\mathcal{F}(\hat{I}_{out}) - \mathcal{F}(I^*)\|_1. \tag{16}$$

The final loss function is given by $L = L_s + \lambda_2 L_f$. In our experiments, λ_1 and λ_2 are set to 0.01 and 0.1, respectively.

4. Experiments and Analysis

We perform extensive experiments to showcase the efficacy of our proposed BMMB. In this section, the results, compared to other existing methods, will offer a more profound understanding of the performance achieved by our proposed module. The ablation studies will demonstrate the effectiveness of our customized units.

4.1. Experimental Setup

Datasets and Evaluation. In our study, the SOTS-indoor (ITS) and SOTS-outdoor (OTS) subsets from the RESIDE dataset [39] are employed as the training dataset, ensuring equitable comparisons with existing dehazing approaches. The assessment of synthetic image dehazing performance is carried out on the SOTS subset. For evaluating the robustness of our BMMB in real-world scenarios, we utilize two widely employed real-world datasets, specifically NH-HAZE [40] and Dense-Haze [41].

Metrics. We adopt widely used image quality assessment indicators, including peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [42], to quantitatively evaluate dehazing performance. For a fair comparison, all metrics are calculated on the full RGB color images.

Table 1: **Quantitative comparisons of various methods on dehazing benchmarks**. The best and second-best results are highlighted in **bold** and underlined, respectively

Method	Venue	SOTS-indoor		SOTS-outdoor		NH-Haze		Dense-Haze		Overhead	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	#Param	MACs
DCP [45]	TPAMI'10	16.62	0.818	19.13	0.815	10.57	0.520	12.72	0.442	-	0.6G
DehazeNet [5]	TIP'16	19.82	0.821	24.75	0.927	16.62	0.524	13.84	0.430	0.01M	0.6G
GFN [46]	CVPR'18	22.30	0.880	21.55	0.844	18.16	0.671	-	-	0.50M	14.9G
GDN [47]	ICCV'19	32.16	0.984	30.86	0.982	13.80	0.537	14.96	0.536	0.96M	21.5G
MSBDN [48]	CVPR'20	33.67	0.985	33.48	0.982	19.23	0.706	15.13	0.555	31.35M	41.54G
FFA-Net [49]	AAAI'20	36.39	0.989	33.57	0.984	19.87	0.692	15.70	0.549	4.46M	287.8G
AECR-Net [50]	CVPR'21	37.17	0.990	-	-	19.92	0.672	14.88	0.505	2.61M	52.2G
MAXIM-2S [51]	CVPR'22	38.11	0.991	34.19	0.985	-	-	-	-	14.10M	216.0G
SGID-PFF [52]	TIP'22	38.52	0.991	30.20	0.975	-	-	12.49	0.517	13.87M	156.4G
Restormer [24]	CVPR'22	38.88	0.991	-	-	-	-	15.78	0.548	26.10M	141.0G
FSDGN [6]	ECCV'22	38.63	0.990	-	-	19.99	**0.708**	16.91	0.581	2.73M	19.57G
Dehamer [11]	CVPR'22	36.63	0.988	35.18	0.986	**20.66**	0.684	16.62	0.560	132.5M	60.3G
DehazeFormer-M [10]	TIP'23	38.46	**0.994**	34.29	0.983	-	-	-	-	4.63M	48.64G
MBTFormer [13]	ICCV'23	40.71	0.992	**37.42**	**0.989**	-	-	16.66	0.560	2.68M	38.50G
DEA-Net [53]	TIP'24	40.20	0.993	36.03	**0.989**	17.59	0.654	16.54	0.586	3.65M	32.23G
SGDN [54]	AAAI'25	40.41	0.889	37.25	0.985	20.02	0.691	16.56	**0.593**	5.41M	29.50G
Ours	-	**41.04**	**0.994**	36.47	**0.989**	20.52	0.686	**17.49**	0.589	5.69M	26.74G

Training Setting. In our methodology, the ADAM optimizer [43] is utilized with specific parameter settings: $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 10^{-8}$. Furthermore, each training iteration involves the random cropping of 8 patches, with each sized at 3×256×256, serving as input images. Within each mini-batch, these patches are further enhanced and augmented, broadening the diversity of the training samples through the application of horizontal or vertical flips and 90-degree rotations. The model implementation is carried out using PyTorch [44] on an NVIDIA 4090 RTX GPU. For the ablation studies, we utilize the ITS to train all models, where the input patch is 3×256×256, and the iterations are 200K. Additionally, we calculate the model size and MACs, with the latter calculated on the 3×256×256 image patch.

4.2. Comparison with State-of-the-Art Methods

We compare our proposed model with fifteen image dehazing models: DCP [45], DehazeNet [5], GFN [46], GDN [47], MSBDN [48], FFA-Net [49], AECR-Net [50], MAXIM-2S [51], SGID-PFF [52], Restormer [24], FSDGN [6], Dehamer [11], DehazeFormer [10], MBTFormer [13], and DEA-Net [53]. All of these models are trained and tested on the same datasets for fair comparisons.

Evaluation on Synthetic Dataset. Table 1 provides a quantitative comparison of our method with other cutting-edge techniques on the widely-used synthetic datasets. It is clear that our BMM achieves the highest PSNR on the SOTS-indoor dataset, ranks

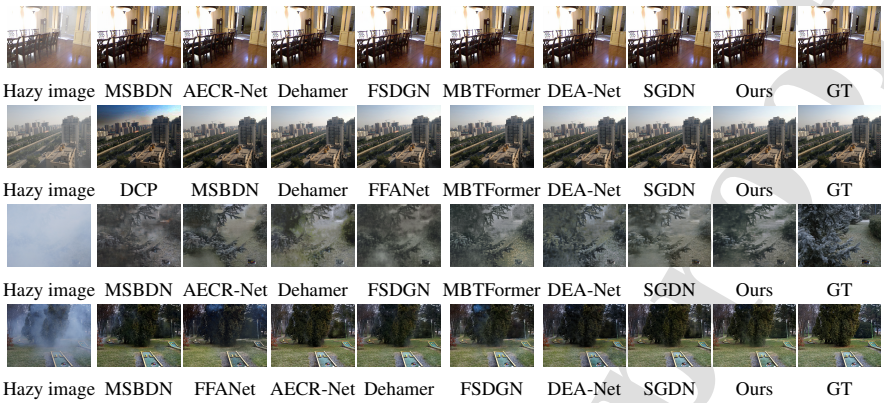


Figure 5: **The first two rows**: Visual comparisons of various methods on synthetic SOTS-indoor and SOTS-outdoor datasets [39]. **The last two rows**: Visual comparisons of various methods on real-world Dense-Haze [41] and NH-HAZE [40] datasets. Please zoom in on the screen for a better view.

second for PSNR on the SOTS-outdoor dataset, and surpasses all other methods in terms of the SSIM metric. More precisely, our model attains 41.04 dB and 36.47 dB PSNR values on synthetic datasets ITS and OTS, respectively, which surpasses most of the compared methods. Moreover, it also achieves SSIM scores of 0.994 and 0.989 on these two datasets, demonstrating the promising dehazing performance. It is worth noting that our model demonstrates the PSNR improvements of 2.58 and 2.41 dB PSNR on SOTS-indoor dataset compared to the representative Transformer-based method DehazeFormer and the bidomain modeling technique FSDGN. We further illustrate the dehazed outputs from several representative methods on the SOTS-indoor and SOTS-outdoor datasets, as visualized in the first two rows of Fig. 5. Our method effectively enhances image details and minimizes color distortions compared to other techniques, demonstrating the closest resemblance to the original clean images and achieving desirable haze removal.

Evaluation on Real-world Datasets. Furthermore, we evaluate our method against previous SOTA methods on two real image datasets, NH-HAZE and Dense-Haze. Table 1 shows quantitative evaluation results of our approach and other state-of-the-art methods. As we can observe, on the Dense-Haze, our model outperforms the MBTFormer [13] by a 0.83 dB margin while utilizing only 69.5% of its MACs and beats all other SOTA approaches. The result indicates the efficiency and less complexity of

Table 2: Component-wise breakdown of the computational cost of the proposed framework. The merge operations are counted into the backbone part.

Component	Params (M)	MACs (G)
Backbone	1.877	9.145
SIU	1.441	7.103
FAU	0.844	3.606
SFU	1.533	6.891
Total	**5.696**	**26.745**

our approach. The visual comparison results on these two datasets are depicted in the last two rows of Fig. 5. It can be observed that our model yields pleasing dehazing outcomes on the tree branches, indicating the proficiency of our method in effectively handling multi-texture features. Overall, the quantitative metrics and visual effects showcase the efficacy of our method over both the synthetic and real-world scenes.

Component-wise Computational Cost. To further analyze the efficiency of the proposed framework, we provide a component-wise breakdown of the computational cost. As shown in Table 2, The overhead is primarily dominated by SIU (due to multi-order modeling) and SFU (due to Monte Carlo sampling). In contrast, FAU introduces marginal computational cost.

4.3. Ablation Study

For a fair comparison, the model is trained for 200 epochs in all ablation experiments.

Component-wise Ablation of SIU, FAU, and SFU. To quantitatively evaluate the individual contributions of SIU, FAU, and SFU, we conduct a drop-one-component ablation study on SOTS-indoor. Following the same setting as the order ablation, all variants are trained on ITS and evaluated on SOTS-indoor. Specifically, we remove SIU, FAU, and SFU from the full model respectively, while keeping the remaining framework and training settings unchanged.

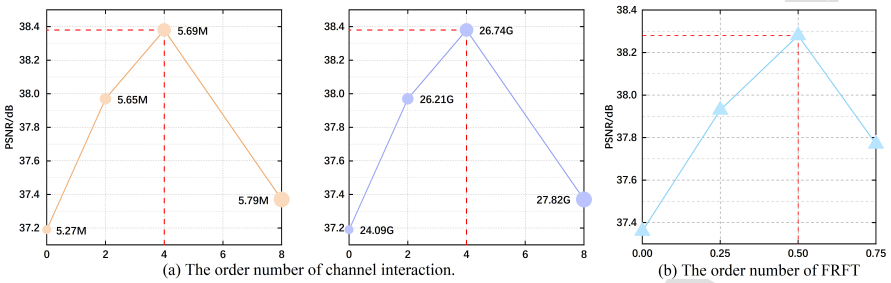
As shown in Table 3, removing SIU causes the largest performance drop, from 38.28 dB to 29.88 dB, demonstrating the importance of spatial-channel multi-order interaction. Removing FAU decreases the PSNR to 35.41 dB, indicating that the intermediate spatial-frequency representation is beneficial for image dehazing. Removing

Table 3: Component-wise ablation study of SIU, FAU, and SFU on SOTS-indoor. All variants are trained on ITS and evaluated on SOTS-indoor, following the same setting as the order ablation.

Config.	SIU	FAU	SFU	PSNR (dB)↑	ΔPSNR
Ours	✓	✓	✓	**38.28**	–
I	✗	✓	✓	29.88	-8.40
II	✓	✗	✓	35.41	-2.87
III	✓	✓	✗	35.60	-2.68

SFU also leads to a performance drop to 35.60 dB, showing that spatial-frequency uncertainty estimation further contributes to robust feature representation. Overall, these quantitative results verify the individual effectiveness of SIU, FAU, and SFU, as well as the necessity of the proposed bidomain modeling architecture. As reported in Table 3, removing SIU causes the most significant performance degradation (ΔPSNR=-8.40 dB), demonstrating the critical role of the proposed multi-order spatial-channel interaction mechanism in capturing informative feature representations for haze removal. This substantial performance degradation is reasonable, since SIU performs hierarchical multi-order spatial-channel interactions to capture multi-granularity feature representations, thereby enabling effective global feature modeling within the proposed BMMB. Moreover, removing FAU (ΔPSNR=-2.87 dB) and SFU (ΔPSNR=-2.68 dB) also results in a noticeable performance drops, further demonstrating their effectiveness in spatial-frequency feature aggregation and robust feature learning, respectively. Overall, these quantitative ablation results further demonstrate the effectiveness and necessity of SIU, FAU, and SFU within the proposed BMM framework.

Effect of the Order k of Channel Interaction. We first probe the effect of the number of orders of channel interactions in SIU on model performance. Specifically, we train and test our model with different times of channel attention, which is represented by $k = 0, 2, 4, 8$, respectively. The minimum value of k is 0, which means there is no channel attention, and the maximum number of channel attentions we explored is set as 8. Then starting from 2, the channel order k of next network doubles. As illustrated in Fig. 6 (a), it is evident that the PSNR has a peak value when the order k equals 4. As we expected, after channel attention is introduced, the model gains a better PSNR score than the model without channel attention. This indicates that our progressive channel



(a) The order number of channel interaction. (b) The order number of FRFT

Figure 6: The ablation studies of the order k of channel interaction and the order p of FRFT.

Table 4: The study of different combinations of c and k, where c denotes the number of channels of the overall model and k denotes the order of channel interaction in SIU.

Method	c=20			c=32		
	k=2	k=4	k=8	k=2	k=4	k=8
PSNR (dB)	37.97	38.28	37.37	38.73	38.91	**39.08**
#Param (M)	**5.65**	5.69	5.79	14.34	14.41	14.53
MACs (G)	**26.21**	26.74	27.82	65.53	66.25	67.69

attention design is effective in feature extraction. However, when k exceeds 4, the result falls back. There may be two reasons: 1) The increase in the number of interactions causes the gradient to become unstable, eventually leading to gradient explosion or gradient vanishing. 2) Excessive channel interaction order results in overly complex extracted features, hindering subsequent modules from effectively utilizing these high-order interaction-derived features.

To further investigate the related reasons, we conduct another experiment with increased the number of channels c of the overall model, as shown in Table 4. Intriguingly, the results reveal that employing larger-scale models yields additional performance enhancements with larger k values. However, the significant escalation in parameters and computational load is a drawback we seek to avoid, thus $c = 20, k = 4$ represents a well-balanced choice when considering the trade-off between various factors. Nonetheless, this suggests the potential of the proposed high-order spatial interaction in scenarios where computational resources are ample.

Effect of the Order p of FRFT. We investigate the effect of the order p of FRFT within our FAU. Specifically, we select four different orders including 0 (spatial), 0.25,

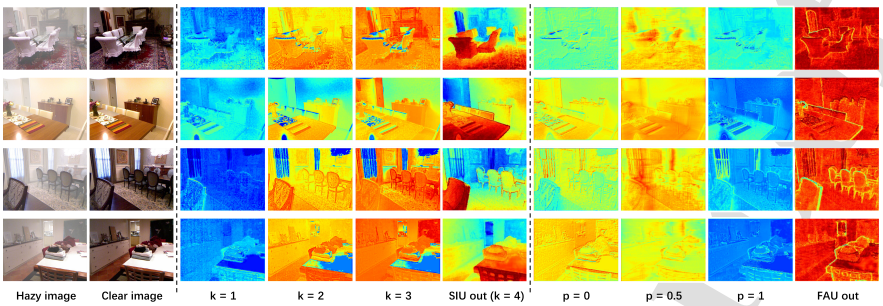


Figure 7: The visualization of feature maps within the proposed BMMB. The left group displays hazy and clear images; the middle group displays the corresponding feature maps obtained by different orders of SIU; the right group displays the feature maps from different orders of FAU. These visualizations illustrate the efficacy of our design from two perspectives: (1) within two multi-order units, feature maps become more distinguishable as the order increases, emphasizing the escalating responses; (2) distinct high-order operations showcase unique responses, demonstrating the diversity in feature representations.

Table 5: The comparison of different loss functions.

Model	(a)	(b)	(c)	(d)
PSNR (dB)	**38.28**	35.91	37.82	35.60
#Params (M)	5.694	5.694	5.694	4.902
MACs (G)	26.74	26.74	26.52	22.90

0.5, and 0.75 for comparison, with the results reported in Fig. 6 (b). Based on the outcomes in the table, the optimal result is observed when the order p is set to 0.5, which means a rotation angle of $\frac{\pi}{4}$. This suggests that a balanced contribution between frequency domain and spatial domain components leads to an increased ability for decoupling, enhancing overall performance. Furthermore, the model with $p = 0.5$ achieves a PSNR improvement of around 1 dB compared to the model that operates solely within the spatial domain. This indicates that the information separation capabilities depending exclusively on either the spatial or frequency domain are typically limited, and a richer feature set leaning towards an intermediate domain can lead to more thorough separation.

Effect of Intermediate Loss Function. We further conduct ablation studies to investigate the impact of extending the BMMB into a Bayesian framework and the related intermediate loss. There are four variants considered: model **(a)**: the BMMB is extended to the Bayesian framework, and the spatial and frequency loss supervises all the

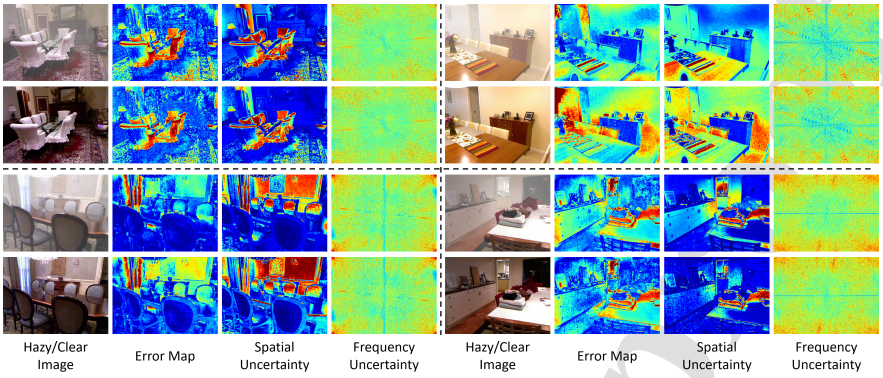


Figure 8: The visualization of spatial and frequency uncertainty. **The first row of each group:** the results from the SIU. **The second row of each group:** the results from the FAU. The network's spatial uncertainty maps align with error maps, showing the awareness of errors and intractable regions. The higher uncertainty in high-frequency components also suggests the network prioritizes low-frequency features over high-frequency ones during training.

intermediate outputs; model **(b)**: the intermediate prediction still consists of multiple MC samples, but the mean of the distribution is only supervised by the spatial loss; model **(c)**: the intermediate prediction is the point estimation and is supervised by the spatial and frequency loss; model **(d)**: there is no intermediate output and corresponding supervision. The results are shown in Table 5. Note that since the prediction from the intermediate layer requires additional convolutional layers (and MC sampling), the number of parameters and MACs will increase slightly.

As we can observe, compared to model (c), model (a) achieves the best performance and maintains the uncertainty output that indicates confidence in the prediction, with slight parameters and MACs increase. The inferior performance of models (b) and (d) shows that the intermediate loss and supervision in the frequency domain are essential for better performance.

Feature Map Visualization within the BMMB. To further demonstrate the feature representational abilities of the proposed BMMB, we present the visualization of feature maps within the BMMB. The input feature map of BMMB will sequentially pass through SIU and FAU. As displayed in Fig. 7, feature maps of the two multi-order units become more distinguishable as the order increases, emphasizing the escalating responses; moreover, distinct multi-order operations showcase unique responses,

demonstrating the diversity in feature representations. Overall, with these two core designs, our BMMB is capable of effectively capturing more informative and discriminative features, benefiting the haze removal.

Spatial Frequency Uncertainty Visualization. By extending our BMMB into the Bayesian framework, spatial and frequency uncertainty can be obtained. Fig. 8 visualizes one image from the ITS test data. The first column represents hazy and clear images, respectively. From the second column to the last one, the error map, spatial, and frequency uncertainty (all from the last block) are shown in order. The first row represents the results from the SIU, and the second row represents the results from the FAU. We can observe an obvious correlation between spatial uncertainty maps and error maps, indicating that our proposed network is aware of errors and intractable degradation regions during the testing phase. Notably, the frequency uncertainty of the high-frequency components is larger than that of the low-frequency components, which indicates that the dehazing network is more inclined to fit the low-frequency part rather than the high-frequency details during the training.

To investigate whether there are specific characteristics and tendencies in the frequency domain in other blocks, we also visualize the output mean/variance of magnitude in the frequency domain for each FAU block of the decoder (the 4-7-th blocks), as shown in Fig. 9. As we delve deeper into the decoder, the range of high-frequency details expands, while their variance decreases. This suggests that deep blocks are progressively more adept at refining hazed high-frequency details and we may need more high-frequency supervision information in the deeper layers of the network.

5. Discussion

Although the proposed BMMB achieves competitive dehazing performance on both synthetic and real-world benchmarks, the current experiments mainly focus on image dehazing. Its applicability to other image restoration tasks, such as image desnowing and deraining, remains to be explored. In addition, this work is limited to single-image dehazing, while video dehazing involves additional challenges, such as maintaining spatial-temporal consistency and handling real-world dynamic scenes. Therefore, extending BMMB to video dehazing is an important future direction.

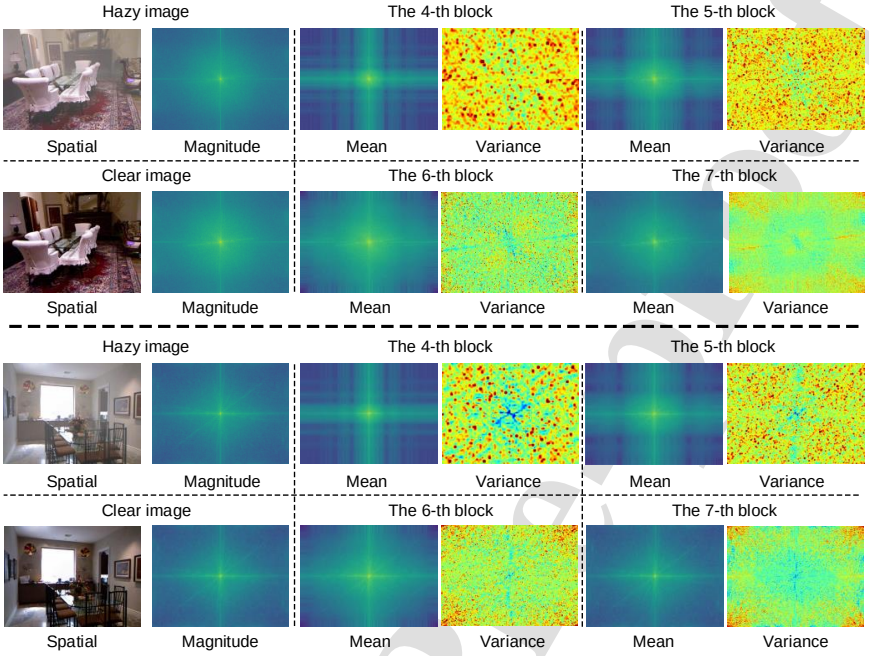


Figure 9: The visualization in the frequency domain. Along with the hazy and clear images and their magnitude in the frequency domain, the mean and variance of magnitude in the frequency domain for 4-7th blocks' FAU are displayed respectively. As the decoder progresses, the range of high-frequency details increases, while their variance decreases. This indicates that deeper blocks are more effective at refining high-frequency details in hazy images, suggesting that additional high-frequency supervision may be required in the deeper layers of the network for improved performance.

In the future, 1) we will further attempt to investigate the potential of adaptive high-order interactions in our core BMMB; 2) we will extend our framework to other image restoration tasks like image desnowing and image deraining; 3) we will investigate the extension of BMMB to video dehazing by incorporating temporal modeling into the current spatial-frequency framework; and 4) we will apply the proposed multi-order modeling design to the currently popular generative diffusion framework, offering a promising avenue for balancing efficiency and expressiveness.

6. Conclusion

In this paper, we propose a spatial-frequency (*i.e.,* bidomain) dehazing framework, which is built upon a novel yet efficient bidomain multi-order modeling block termed BMMB. To be specific, our proposed BMMB consists of a spatial interaction unit (SIU)

and a frequency aggregation unit (FAU). The SIU aims to learn local and global dependencies through spatial-wise and channel-wise multi-order modeling design. Subsequently, the FAU applies the Fractional Fourier Transform (FRFT) with flexible spatial-frequency representation capabilities to process the hierarchical feature information extracted from the SIU, enabling more effective information aggregation. In addition, we introduce the spatial-frequency uncertainty capable of incorporating the degradation priors into feature learning within the above two components, achieving robust representation. Experiments on multiple benchmark datasets demonstrate the applicability of our dehazing framework assembled by the proposed BMMB.

CRediT authorship contribution statement

Zihao Chen: Conceptualization, Methodology, Software, Writing-original draft, Visualization. **Hao Shen:** Methodology, Software, Writing-original draft, Visualization. **Chenxu Wu:** Writing-review & editing, Visualization, Validation. **Junling Li:** Writing-review & editing, Validation. **Wei Wang:** Writing-review & editing, Validation, Supervision. **Wenqi Ren:** Writing-review & editing, Validation, Supervision.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 62301151 and 62306343).

References

[1] C. Sakaridis, D. Dai, S. Hecker, L. Van Gool, Model adaptation with synthetic and real data for semantic dense foggy scene understanding, in: European Conference on Computer Vision, 2018, pp. 687–704.

[2] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, End-to-end united video dehazing and detection, in: AAAI Conference on Artificial Intelligence, Vol. 32, 2018.

[3] J. Gui, X. Cong, Y. Cao, W. Ren, J. Zhang, J. Zhang, J. Cao, D. Tao, A comprehensive survey and taxonomy on single image dehazing based on deep learning, ACM Computing Surveys 55 (13s) (2023) 1–37.

[4] Y. Feng, L. Ma, X. Meng, F. Zhou, R. Liu, Z. Su, Advancing real-world image dehazing: Perspective, modules, and training, IEEE Transactions on Pattern Analysis and Machine Intelligence 46 (12) (2024) 9303–9320. doi:10.1109/TPAMI.2024.3416731.

[5] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, IEEE Transactions on Image Processing 25 (11) (2016) 5187–5198.

[6] H. Yu, N. Zheng, M. Zhou, J. Huang, Z. Xiao, F. Zhao, Frequency and spatial dual guidance for image dehazing, in: European Conference on Computer Vision, 2022, pp. 181–198.

[7] Y. Z. Su, C. He, Z. G. Cui, A. H. Li, N. Wang, Physical model and image translation fused network for single-image dehazing, Pattern Recognition 142 (2023) 109700.

[8] H. Shen, Z.-Q. Zhao, Y. Zhang, Z. Zhang, Mutual information-driven triple interaction network for efficient image dehazing, in: ACM International Conference on Multimedia, 2023, p. 7–16.

[9] Y. Cui, Q. Wang, C. Li, W. Ren, A. Knoll, Eenet: An effective and efficient network for single image dehazing, pattern recognition 158 (2025) 111074.

[10] Y. Song, Z. He, H. Qian, X. Du, Vision transformers for single image dehazing, IEEE Transactions on Image Processing 32 (2023) 1927–1941.

[11] C.-L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, C. Li, Image dehazing transformer with transmission-aware 3d position embedding, in: IEEE Conference on Computer Vision and Pattern Recognition, 2022, pp. 5812–5820.

[12] P. Saxena, A. K. Tiwari, M. Narwaria, Adaptive self-attention enhanced conditional gan for image dehazing, Pattern Recognition (2026) 113277.

[13] Y. Qiu, K. Zhang, C. Wang, W. Luo, H. Li, Z. Jin, Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing, in: IEEE International Conference on Computer Vision, 2023, pp. 12802–12813.

[14] M. Zhou, J. Huang, C.-L. Guo, C. Li, Fourmer: an efficient global modeling paradigm for image restoration, in: International Conference on Machine Learning, PMLR, 2023, pp. 42589–42601.

[15] H. Shen, Z.-Q. Zhao, Y. Zhang, Z. Zhang, Mutual information-driven triple interaction network for efficient image dehazing, in: Proceedings of the 31st ACM international conference on multimedia, 2023, pp. 7–16.

[16] Y. Feng, J. Li, T. Huang, F. Wu, Y. Ju, C. Li, W. Dong, A. C. Kot, Cross-frequency attention and color contrast constraint for remote sensing dehazing, IEEE Transactions on Image Processing (2025) 1–1doi:10.1109/TIP.2025.3644167.

[17] C. Liu, L. Qi, J. Pan, X. Qian, M.-H. Yang, Frequency domain-based diffusion model for unpaired image dehazing, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 7538–7547.

[18] Q. Ning, W. Dong, X. Li, J. Wu, G. Shi, Uncertainty-driven loss for single image super-resolution, Advances in Neural Information Processing Systems 34 (2021) 16398–16409.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in Neural Information Processing Systems 30 (2017).

[20] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: IEEE International Conference on Computer Vision, 2021, pp. 10012–10022.

[21] L. Kong, J. Dong, J. Ge, M. Li, J. Pan, Efficient frequency domain-based transformers for high-quality image deblurring, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5886–5895.

[22] J. Yang, C. Li, X. Dai, J. Gao, Focal modulation networks, Advances in Neural Information Processing Systems 35 (2022) 4203–4217.

[23] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S. N. Lim, J. Lu, Hornet: Efficient high-order spatial interactions with recursive gated convolutions, Advances in Neural Information Processing Systems 35 (2022) 10353–10366.

[24] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, Restormer: Efficient transformer for high-resolution image restoration, in: IEEE Conference on Computer Vision and Pattern Recognition, 2022, pp. 5728–5739.

[25] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Image restoration using swin transformer, in: International Conference on Computer Vision Workshops, 2021, pp. 1833–1844.

[26] D. Han, X. Pan, Y. Han, S. Song, G. Huang, Flatten transformer: Vision transformer using focused linear attention, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 5961–5971.

[27] Y. Qiu, K. Zhang, C. Wang, W. Luo, H. Li, Z. Jin, Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 12802–12813.

[28] N. Zheng, M. Zhou, C. Zhou, C. C. Loy, Rubik's cube: High-order channel interactions with a hierarchical receptive field, in: Thirty-seventh Conference on Neural Information Processing Systems, 2023.

[29] Y. Gu, Y. Meng, S. Chen, J. Ji, X. Sun, W. Ruan, R. Ji, Sfir: Optimizing spatial and frequency domains for image restoration, Pattern Recognition (2025) 112188.

[30] H. Yu, J. Huang, L. Li, M. Zhou, F. Zhao, Deep fractional fourier transform, in: Thirty-seventh Conference on Neural Information Processing Systems, 2023.

[31] C. Huang, W. Tan, J. Shi, Z. Xing, B. Yan, Uncer2natural: Uncertainty-aware unsupervised image denoising, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2023, pp. 1–5.

[32] T. Liu, J. Cheng, S. Tan, Spectral bayesian uncertainty for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 18166–18175.

[33] M. Hong, J. Liu, C. Li, Y. Qu, Uncertainty-driven dehazing network, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 906–913.

[34] H. Ozaktas, O. Arikan, M. Kutay, G. Bozdagt, Digital computation of the fractional fourier transform, IEEE Transactions on Signal Processing 44 (9) (1996) 2141–2150. doi:10.1109/78.536672.

[35] A. Kendall, Y. Gal, What uncertainties do we need in bayesian deep learning for computer vision?, Advances in neural information processing systems 30 (2017).

[36] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, PMLR, 2016, pp. 1050–1059.

[37] C. E. Brown, Coefficient of variation, in: Applied multivariate statistics in geohydrology and related sciences, Springer, 1998, pp. 155–157.

[38] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.

[39] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, Z. Wang, Benchmarking single-image dehazing and beyond, IEEE Transactions on Image Processing 28 (1) (2018) 492–505.

[40] C. O. Ancuti, C. Ancuti, R. Timofte, Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 444–445.

[41] C. O. Ancuti, C. Ancuti, M. Sbert, R. Timofte, Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images, in: IEEE International Conference on Image Processing, 2019, pp. 1014–1018.

[42] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.

[43] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[44] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in pytorch (2017).

[45] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (12) (2010) 2341–2353.

[46] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, M.-H. Yang, Gated fusion network for single image dehazing, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3253–3261.

[47] X. Liu, Y. Ma, Z. Shi, J. Chen, Griddehazenet: Attention-based multi-scale network for image dehazing, in: IEEE International Conference on Computer Vision, 2019, pp. 7314–7323.

[48] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang, Multi-scale boosted dehazing network with dense feature fusion, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 2157–2167.

[49] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Jia, Ffa-net: Feature fusion attention network for single image dehazing, in: AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 11908–11915.

[50] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, L. Ma, Contrastive learning for compact single image dehazing, in: IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 10551–10560.

[51] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, Y. Li, Maxim: Multi-axis mlp for image processing, in: IEEE Conference on Computer Vision and Pattern Recognition, 2022, pp. 5769–5780.

[52] H. Bai, J. Pan, X. Xiang, J. Tang, Self-guided image dehazing using progressive feature fusion, IEEE Transactions on Image Processing 31 (2022) 1217–1229.

[53] Z. Chen, Z. He, Z.-M. Lu, Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention, IEEE Transactions on Image Processing (2024).

[54] W. Fang, J. Fan, Y. Zheng, J. Weng, Y. Tai, J. Li, Guided real image dehazing using ycbcr color space, Proceedings of the AAAI Conference on Artificial Intelligence 39 (3) (2025) 2906–2914.

Declaration of interests

√The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: