VDMamba: Vector Decomposition in Vision Mamba for Image Deraining and Beyond

Kui Jiang, *Member, IEEE*, Junjun Jiang, *Senior Member, IEEE*, Shiqi Wang, *Senior Member, IEEE*, Wenqi Ren, *Member, IEEE*, Chia-Wen Lin, *Fellow, IEEE*, Zhengguo Li, *Fellow, IEEE*

***Abstract*—Image deraining aims to remove rain perturbations from rainy images and restore clear backgrounds. Recent research has employed the Mamba technique for image restoration, achieving exceptional results due to its effectiveness and efficiency in modeling long-range sequence relationships. However, a significant challenge remains: developing a comprehensive framework that considers the intrinsic coupling characteristics between image deraining and the Mamba architecture is largely unexplored. We propose that introducing a 1D sequential representation of Mamba could enhance image deraining by characterizing the direction-aware distribution of rain perturbations. This motivates us to introduce a new vector decomposition-based vision Mamba approach (VDMamba). This method investigates vector decomposition within the context of vision Mamba, addressing the challenging task of image deraining and beyond in the frequency embedding space. The key innovation of VDMamba is the Mamba-based vector decomposition and synthesis module (VDSM). This module derives 1D basic vectors (vertical and horizontal) from the frequency components via vector decomposition and employs the single-direction scanning of Mamba to eliminate the direction-specific degradation perturbation. This transformation allows the incipient Mamba to explore direction-specific global sequential relationships for accurate perturbation distribution learning, without requiring an elaborate design of the Mamba scanning process. Additionally, the vertical and horizontal components in VDSM are encoded jointly in a bidirectional coupling manner, enabling the exploration of complementary and redundant components for refinement. Experiments on various main image enhancement tasks, including image deraining, raindrop removal, rain haze removal, image dehazing, low light image enhancement, and underwater image enhancement, demonstrate that VDMamba delivers competitive performance compared to the state-of-the-art NeRD method. Specifically, it achieves a 0.58 dB improvement in PSNR for the image deraining task while reducing model parameters by 94.3%, computational cost by 88.3%, and inference time by 77.5%.**

***Index Terms*—Image deraining, Mamba, Vector decomposition and synthesis, Mutual representation.**

This research was financially supported by the Natural Science Foundation of Heilongjiang Province of China for Excellent Youth Project (YQ2024F006), the National Natural Science Foundation of China (U23B2009) and Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) (GML-KF-24-09).

K. Jiang and J. Jiang are with the School of Computer Science and Technology, Harbin Institute of Technology, China, 150001 (e-mail: jiangkui@hit.edu.cn, jiangjunjun@hit.edu.cn).

Shiqi Wang is with the Department of Computer Science, City University of Hong Kong, Hong Kong, 999077 (e-mail: shiqwang@cityu.edu.hk).

Wenqi Ren is with the School of Cyber Science and Technology, Sun Yat-sen University, China, 510275 (e-mail: renwq3@mail.sysu.edu.cn).

C.-W. Lin is with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 300044, Taiwan (e-mail: cwlin@ee.nthu.edu.tw).

Z. Li is with VI Department, Institute for Infocomm Research, Agency for Science, Technology and Research (A⋆STAR), Singapore 138632, Republic of Singapore (e-mail: ezgli@i2r.a-star.edu.sg).

I. Introduction

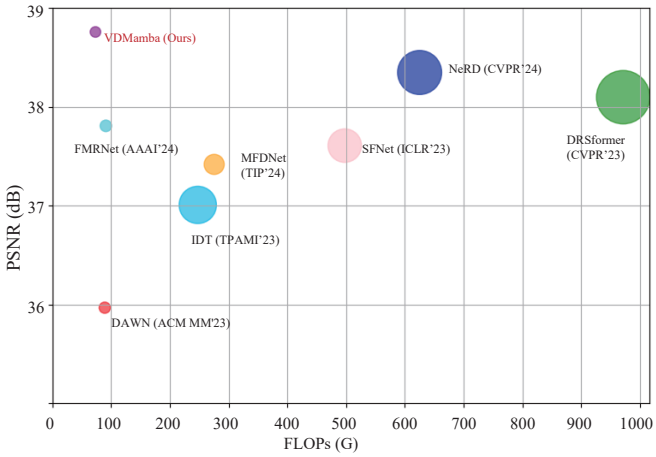

Fig. 1: Model parameters, FLOPs and deraining performance comparison between our VDMamba and other state-of-the-art algorithms. The size of the circle area indicates the number of model parameters. It is noted that our VDMamba enjoys the best deraining performance while with highly competitive computational cost and model parameters.

RAIN perturbation can significantly degrade image quality, resulting in a loss of object details and contrast. These adverse effects lead to unpleasant visual results that substantially affect the performance of advanced visual tasks, such as image understanding [1], [2], object detection [3], and segmentation [4]. Image deraining [5], [6] aims to produce rain-free results from rainy inputs and has garnered widespread attention over the past decade. However, effectively removing rain while recovering clear background content is challenging due to the complex and diverse rain patterns.

Before the advent of deep neural networks, traditional methods primarily relied on statistical analyses to examine physical properties, and enforce handcrafted priors (*e.g.*, sparsity and non-local means filtering) on both the rain and background layers. However, these methods often lack robustness against diverse and complex rain conditions, as the effectiveness of these priors may be unreliable [7], [8].

Deep learning methods have emerged as a preferable alternative due to their ability to learn generalizable priors from large-scale data. However, CNN-based image deraining methods struggle to manage content diversity and model global rain distribution due to their intrinsic local connectivity and translation equivariance. Accurately modeling the global degradation of rainy images is crucial for addressing complex challenges associated with image deraining. To enhance deraining performance, the non-local process [9], [10] and self-attention

(SA) mechanism in Transformers leverage global relation to mitigate degradation perturbations [11], [12]. While SA achieves impressive performance through global calculations, it faces scalability challenges due to its quadratic complexity with respect to token numbers, making it impractical for high-resolution images [13]. For example, the high-accuracy model DRSformer [14] has 971.6 GFLOPs and 33.65 million parameters, requiring 1.328 seconds to derain a 512×512 pixel image using a single 4090 GPU, as shown in Fig. 1. This computational and memory demand can be prohibitive for many real-world applications on resource-constrained devices.

To alleviate the computational burden while capturing global structural information, some efforts have harmonized the merits of Wavelet transform and Transformer into the deraining task, aiming to eliminate rain perturbations in the frequency embedding space [15]. By leveraging coefficient-based (approximation, vertical, horizontal and diagonal) learning [16], [17], these approaches offer greater flexibility and practicality for handling diverse rain patterns, resulting in more distinct representations that are conducive to the removal of real-world rain streaks. However, improved model performance typically requires a larger number of model parameters and increased computational complexity [18], which becomes inevitable to confront issues related to computational complexity and memory limitations on resource-constrained devices.

Due to its efficient ability to capture long-range sequence relationships with linear complexity, there is growing interest in the Mamba model, an enhanced structured state-space sequence model (S4) [19], [20]. However, it is noteworthy that the standard Mamba is specifically tailored for 1D sequential data in NLP tasks, making it less suitable for direct application in 2D image restoration. In particular, Mamba projects the input image into pixel sequences and only aggregates previous tokens, which can result in losses of intrinsic spatial structures and global context. While some efforts [15], [21] have combined Mamba with Fourier transforms to develop multi-branch or -scanning architectures that explore complementary information for frequency dependence learning, rarely attempts have investigated the coincidence between characteristics of rainy images and Mamba architecture.

Recognizing this limitation, we attempt to extend the Mamba-based image deraining model from the perspective of vector decomposition and synthesis. *Specifically, the 1D vertical and horizontal basic vectors can be derived from any arbitrary vector via decomposition and transformation, which is applicable for the distribution of both rain and background textures*. Therefore, it is reasonable to derive and eliminate the corresponding 1D vertical and horizontal basic rain components from the observed rain input.

Based on the analyses above, we propose to explore the potential of vector decomposition learning in vision Mamba, and construct a novel vector decomposition-based vision Mamba approach (VDMamba) for image deraining. Instead of directly addressing rain distribution in the image space, we present a Mamba-based vector decomposition and synthesis module (VDSM) to progressively fit and eliminate the 1D basic vectors of rian perturbation in the frequency space. This provides an effective solution for local-global degradation representation without requiring a specific scanning mechanism design. Specifically, the input to the first VDSM is an initial distribution of background which is encoded via a feature extraction module (FEM), and the output of each VDSM is the input to its subsequent VDSM. In each VDSM, we generate the initial representation of 1D basic vectors from Wavelet coefficients via an elaborate direction-aware projection unit (DPU), followed by two direction-individual residual Mamba blocks (RMB) to capture the direction-specific non-local correlations of degradation. This approach transforms the prediction and removal of rain perturbations or background recovery into a parameterized fitting problem of vector decomposition and synthesis. Since it is hard to completely decouple and separate two orthogonal parts (vertical and horizontal components) from the rainy distribution via the vector decomposition To enhance feature representation and training, the vertical and horizontal components in VDSM are encoded jointly in a bidirectional coupling manner. Specifically, the operations of DPU and Mamba blocks, applied in different directions, extract complementary and redundant components from each other for refinement. Ultimately, the global rain distribution in the frequency space can be modeled and eliminated using the basis vectors and the corresponding transformation and projection parameters via all the progressive VDSMs and the rain-free image can be restored by a Reconstruction Module (RM). In this way, our approach implicitly encodes the physics—orientation (degradation or texture) and and heterogeneous degradation—into the network's architecture. By disentangling directional components, VDSM mimics the physical process of degradation accumulation, enabling interpretable feature recovery. ince the vector decomposition and synthesis, non-local aggregation strategies are task-agnostic, we propose to assess their versatility and effectiveness by extending their application to image deraining and beyond.

Overall, the main contributions of our proposed VDMamba are summarized as follows.

- We propose the novel VDMamba framework for image deraining and beyond, which investigates vector decomposition learning within a frequency-based vision Mamba. To the best of our knowledge, this is the first dedicated coupling of the Mamba architecture and vector decomposition and synthesis on the mainstream image restoration tasks, providing a new insight for designing efficient and effective image restoration framework.
- An elaborate Mamba-based vector decomposition and synthesis module (VDSM) is devised to characterize and eliminate the perturbation components associated with the fundamental basis vectors in the frequency space. It is allowed to explore direction-specific global sequential relationship without requiring an elaborate design of the Mamba scanning process, which is easy to extend to tasks with similar decomposition characteristics.
- Extensive experimental results on mainstream image restoration tasks, including image deraining, raindrop removal, rain haze removal, image dehazing, low-light image enhancement, and underwater image enhancement, demonstrate that our proposed VDMamba approach

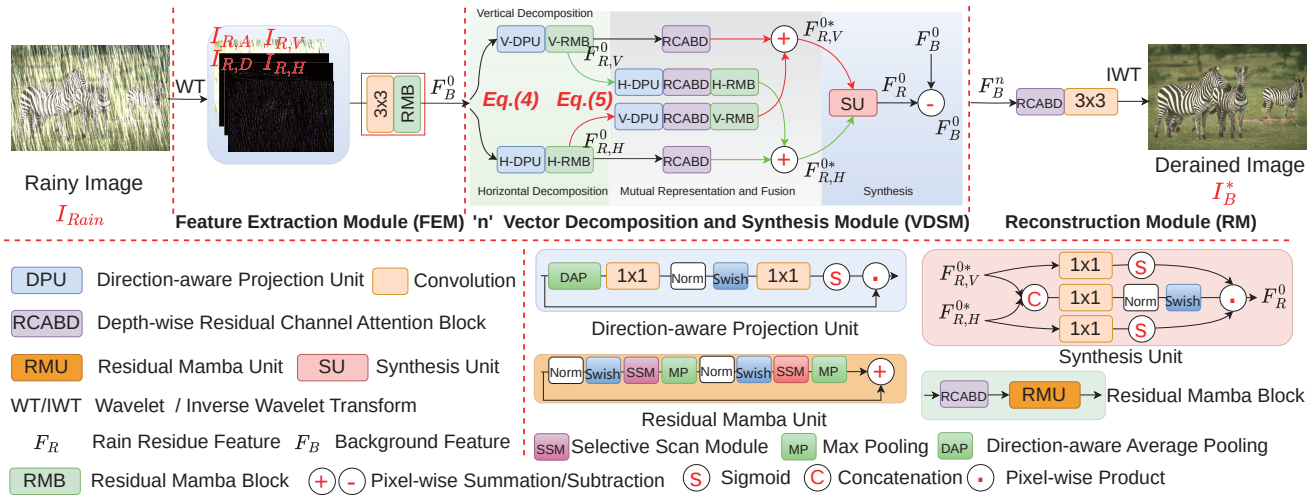

Fig. 2: The architecture of our proposed vector decomposition-based vision Mamba approach (VDMamba) comprises a feature extraction module (FEM), several cascaded vector decomposition and synthesis modules (VDSMs), and a reconstruction module (RM). FEM learns the initial representation of background contents F_B^0 from the Wavelet coefficients. VDSM takes F_B^0 as input, first characterizing and deriving the corresponding direction-specific rain residue (vertical ($F_{R,V}^0$) and horizontal ($F_{R,H}^0$)) via the direction-aware projection unit (DPU) and residual Mamba block (RMB). The rain residues of vertical and horizontal vectors are then encoded jointly to explore complementary information for mutual representation and better decomposition. The refined rain residues ($F_{R,H}^{0*}$ and $F_{R,V}^{0*}$) are used to compose the rain distribution prediction (F_R^0) of the current rain layer via the synthesis unit (SU) and generate the updated background representation F_B^1. Through progressive vector decomposition and synthesis with several VDSMs, the network is encouraged to eliminate rain perturbations in the frequency space, producing rain-free background features. Finally, the reconstruction module transforms the predicted background contents (F_B^n) into the image domain to generate the rain-free image (I_B^*).

quantitatively outperforms the current SOTA technologies while maintaining considerable computational efficiency and reduced model parameters.

The remainder of this paper is organized as follows. In Sec. II, we introduce the development of image deraining, wavelet transform, state space model in the field of computer vision. Sec. III reveals the efficient network designs of our VDMamba and optimization constraints. In Sec. IV, experimental settings and compared results are presented to demonstrate the superiority of the proposed VDMamba. Finally, the conclusion and potential limitation are provided in Sec. V.

II. RELATED WORK

In this section, we first review the recent advances in image deraining and wavelet transform. Then, we introduce the progress of the state space model in vision tasks.

A. Single Image Deraining

Prior to deep learning, conventional methods [22], [23] rely on handcrafted priors for deraining tasks, offering various available schemes. However, since the priors and assumptions are tailored to specific scenarios, these methods [8] struggle with poor generalization to highly complex and varied rainy scenes. Recently, deep learning technologies have been applied to rain removal tasks [24], [25], leading to advancements in innovative architectures, optimization strategies, and training practices, significantly improving visual quality compared to conventional algorithms. For instance, some researchers [26], [27] view the complete rain distribution prediction as the combination of multiple sub-spaces, estimating and aggregating multi-scale features for textural reconstruction. Additionally, increasing efforts also take rain characteristics such as rain density [28], size and the veiling effect [25] into account, and use recurrent neural networks to remove rain streaks via multiple stages [29] or the non-local networks [9], [30] to exploit long-range spatial dependencies for better rain removal. To further promote background recovery, some researchers employ Wavelet transform [17], semantic context [31] and Mamba [15] to encode the rain distribution and explore the complementary information for frequency dependence learning [32], [33]. Despite these advances, current methods still struggle to produce visually pleasing results due to the absence of reasonable design. In particular, few efforts are developed to investigate the intrinsic coincidence between characteristics of rainy images and the Mamba architecture, such as the adaptivety between the scanning mechanism of Mamba and the direction characteristics of rain perturbation.

B. Wavelet Transform

wavelet transforms [34] are widely used in signal processing tasks [35] due to their reversible property and preeminent ability to capture contextual and textural information of an image at various levels. Recent studies have aimed to combine these advantages to enhance image deraining performance [16], [36]. However, many of these methods [37], [38] overlook the coincidence between characteristics of rainy images and Wavelet transform. They often apply the common multi-branch strategy for frequency-based rain degradation learning, ignoring the intrinsic direction characteristics of texture representation. In contrast, we marries the power of Mamba and vector decomposition learning with the Wavelet-based image deraining. Unlike existing approaches [17], [36], our proposed vector decomposition-based vision Mamba approach (VDMamba) transforms the prediction and removal of rain distribution or background recovery into a parameterized fitting problem of vector decomposition and synthesis in the frequency space. By characterizing and eliminating the rain distribution associated

with the fundamental basis vectors, it offers greater flexibility and practicality in the learning process.

C. State Space Model

Recently, state space models (SSMs) [39] have emerged as a promising approach, demonstrating competitive performance in long-range modeling compared to transformers. The success of SSMs is primarily due to their linear scaling with sequence length, which provides a global perspective with linear complexity. By incorporating the scanning strategy [40], SSMs have shown superior performance to transformers in natural language processing tasks. Inspired by this, SSM-based technologies have also been applied to computer vision tasks, achieving impressive results [41], [42], including object detection [43], image classification [44], and biomedical image segmentation [45]. For instance, Gu *et al.* [46] pioneer SSM to tackle long-sequence data modeling, demonstrating promising linear scaling properties. They subsequently devise a variant named Mamba [47], which involves a more effective selective scanning mechanism and network design. More recently, some researchers have integrated Mamba vision into image restoration tasks [48]–[51], such as image super-resolution [52], image dehazing [53] and image deraining [21], [54]. For example, Chen *et al.* [55] propose to dynamically adjust the quantization mapping range, and devise the Q-MambaIR, an accurate, efficient, and flexible Quantized Mamba for IR tasks. Specifically, they introduce a Statistical Dynamic-balancing Learnable Scalar (DLS) to mitigate the peak truncation loss caused by extreme values, while designing a Range-floating Flexible Allocator (RFA) with an adaptive threshold to flexibly round values. To simultaneously perceive image textures and achieve a trade-off between performance and efficiency, the authors in [56] design a novel texture-aware image restoration method, TAMambaIR. Specifically, TAMambaIR involves a novel Texture-Aware State Space Model and a Multi-Directional Perception Block, which enhances texture awareness and improves efficiency by modulating the transition matrix of the state-space equation and focusing on regions with complex textures. To generalize to complex degradation, the authors in [57] devise DPMambaIR, a novel All-in-One image restoration framework with a Degradation-Aware Prompt State Space Model (DP-SSM) and a High-Frequency Enhancement Block (HEB). Specifically, the DP-SSM utilizes a pre-trained degradation extractor to capture fine-grained degradation features and dynamically incorporates them into the state space modeling process, enhancing the model's adaptability to diverse degradation types while mitigating the loss of high-frequency details caused by task competition. However, the exploration for vision Mamba to promote image deraining tasks is still rough and immature, and in particular the investigation of the coincidence between characteristics of rainy images and Mamba architecture remains uncharted. Unlike existing methods designed to encode 2-D channel relationships and long-range dependencies with positional information, our VDSM focuses on vector decomposition and parameter fitting learning to achieve direction-specific representation, mimicking the physical process of rain accumulation, enabling interpretable feature recovery.

III. Method

This section first offers an overview of the proposed vector decomposition-based vision Mamba approach (VDMamba) and then details the architecture of our vector decomposition and synthesis module (VDSM) and its essential components, including background decomposition, mutual representation and fusion, and synthesis unit.

A. Architecture Overview and Optimization

Architecture Overview. Fig. 2 outlines the framework of VDMamba, which involves a Wavelet transform (WT), a FEM, several VDSMs, and a RM. For a given rain image I_{Rain} and its clean background image I_B with a size of $W \times H \times C$, we first use the Haar Wavelet to decompose the corresponding frequency components, including the low-frequency approximation coefficient $I_{R,A}$ and high-frequency vertical coefficient $I_{R,V}$, horizontal coefficient $I_{R,H}$, and diagonal coefficient $I_{R,D}$. Compared to Fast Fourier Transform (FFT), which is suitable for analyzing the overall spectral characteristics of images, Wavelet transform has time-frequency joint localization characteristics and can analyze signals at different scales and positions while capturing more structural information. Thus it is critical for applications like detecting transient events or analyzing signals with abrupt changes (*e.g.*, rain distribution, machine fault detection).

Then these components have the same size of $W/2 \times H/2 \times C$. The coefficients are then transformed into the feature space through the FEM, involving an initial 3×3 convolution and a RMB to generate the initial background representation (F_B^0). Subsequently, F_B^0 is converted into a deep mutual learning representation via several VDSM. Specifically, VDSM fits and eliminates the direction-specific (vertical and horizontal) basic vectors of rain distribution in the frequency space using the DPU. Finally, in the RM, an inverse Wavelet transform is performed to produce the derained results (I_B^*) with fixed parameters of Haar filters [58].

Model Optimization. Similar to existing studies [59], the Charbonnier penalty loss [60] is used to mediate between the predicted rain-free image (I_B^*) and its ground truth (I_B). The constraint is formulated as

$$\mathcal{L}_{\mathrm{RGB}} = \sqrt{(I_B^* - I_B)^2 + \varepsilon^2}, \tag{1}$$

where the penalty coefficient ε is set to 10^{-3} with the α set to 0.2 to balance loss components. To encourage the fidelity of both structural information and texture, the structural similarity (SSIM) [61] loss is introduced, depicted as

$$\mathcal{L}_{\mathrm{Y}} = \mathrm{SSIM}(I_{B,Y}^*, I_{B,Y}), \tag{2}$$

where Y denotes the luminance channel of the YCbCr space. To this end, the final loss function is defined as $\mathcal{L} = \mathcal{L}_{\mathrm{RGB}} + \lambda \cdot \mathcal{L}_{\mathrm{Y}}$, where λ is used to balance the loss components, and experimentally set as -0.15.

B. Feature Extraction Module

In VDMamba, we have devised a FEM to encode the initial distribution of background. Specifically, a 3×3 convolution

operation is applied to these wavelet coefficients to generate the initial characteristics F_{fea}. Then a RMB is composed of a RCABD and a RMU, which are used to aggregate the global response and refine the feature representation F_B^0. In RCABD, the standard convolution is replaced with depth-wise separable convolutions, which are computationally more efficient while achieving similar or better performance over the incipient RCAB. This process can be formulated as

$$\begin{aligned} F_{fea} &= H_{conv}(I_{R,A}, I_{R,V}, I_{R,H}, I_{R,D}), \\ F_B^0 &= H_{RMB}(F_{fea}), \end{aligned} \tag{3}$$

where H_{conv} denotes a 3×3 convolution layer. H_{RMB} denotes the RMB, as shown in Fig. 2, involving a RCABD, a RMU and a global residual connection to characterize both the local and global dependencies.

C. Vector Decomposition and Synthesis Module

The pivotal design of VDMamba is the VDSM, which is utilized to characterize and eliminate the direction-specific rain degradation. There are n VDSMs in the proposed VDMamba. The input of the ith VDSM is F_B^{i-1} and its output is F_B^i. Each VDSM involves vector decomposition, mutual representation and fusion, and vector synthesis, detailed as follows.

Vector Decomposition. We first characterize and derive the corresponding direction-specific rain perturbation distribution (vertical $F_{R,V}^0$ and horizontal $F_{R,H}^0$) via the DPU and RMB ($H_{RMB}(\cdot)$) from input features (F_B^0). This transformation allows the single-direction scanning of the incipient Mamba to explore direction-specific sequential relations for accurate perturbation distribution learning, requiring no elaborate design of Mamba scanning. This process is represented as

$$\begin{aligned} F_{R,V}^0 &= H_{RMB}(H_{DPU,V}(F_B^0)), \\ F_{R,H}^0 &= H_{RMB}(H_{DPU,H}(F_B^0)), \end{aligned} \tag{4}$$

where $H_{DPU,V}$ and $H_{DPU,H}$ denote the vertical and horizontal DPUs (shown in Fig. 2), respectively.

Mutual Representation and Fusion. Due to the diversity and complexity of rain patterns, it is hard to completely decouple and separate two orthogonal parts (vertical and horizontal components) from the rainy distribution via the vector decomposition. Besides the direction-specific representation via vector decomposition, we present the mutual representation and fusion across the vertical and horizontal components, greatly facilitating the global degradation learning. Specifically, the vertical and horizontal components in VDSM are encoded jointly in a bidirectional coupling manner to utilize the non-local correlation better. This scheme explores the complementary and redundant components from each other for refinement. Specifically, taking the vertical vectors $F_{R,V}^0$ as an example, $F_{R,V}^0$ is processed by a horizontal DPU ($H_{DPU,H}(\cdot)$) to generate descriptions along the horizontal direction via the vertical coordinate information embedding. This is followed by the RCABD and RMU, enabling the network to fully explore the complementary information from the vertical vectors. This enriched output is then supplemented to the horizontal vectors, which are also enhanced by another RCABD. The refined vertical vectors $F_{R,V}^{0*}$ can be generated similarly. The aforementioned process are described as

$$\begin{aligned} F_{R,H}^{0*} &= H_{RMB,H}(E_1(H_{DPU,H}(F_{R,V}^0))) + E_2(F_{R,H}^0), \\ F_{R,V}^{0*} &= H_{RMB,H}(E_1(H_{DPU,V}(F_{R,H}^0))) + E_2(F_{R,V}^0), \end{aligned} \tag{5}$$

where E represents the encoding process achieved by RCABD.

Vector Synthesis. To better characterize the rain perturbation and alleviate training burden, we devise a lightweight yet effective synthesis unit to generate and eliminate the rain distribution of the current rain layer. In particular, Synthesis Unit (H_{SU}) takes the refined rain residues ($F_{R,V}^{0*}$ and $F_{R,H}^{0*}$) as inputs, and learns the inverse projection and transformation parameters of vector decomposition using two 3×3 convolutions, which are used to derive the complete rain distribution (F_R^0), and generate the updated background representation F_B^1. These processes can be summarized as

$$\begin{aligned} F_R^0 &= H_{SU}(F_{R,V}^{0*}, F_{R,H}^{0*}), \\ F_B^1 &= F_B^0 - F_R^0. \end{aligned} \tag{6}$$

Using the progressive VDSMs, our approach enables the network to effectively eliminate the rain perturbation. This is accomplished by characterizing the direction-specific rain distribution while exploring the intrinsic relationships among them, leading to a high-quality representation of rain-free background features. Our innovative scheme not only outperforms traditional separation methods in terms of accuracy, but also preserves intricate textures, requiring no elaborate design of Mamba scanning mechanism.

D. Reconstruction Module

Followed by a reconstruction module, the predicted background contents (F_B^n) from the last VDSM are transformed into the image domain to generate a rain-free image (I_B^*). Specifically, an RCABD followed by a convolution layer 3×3 receives F_B^n, and an inverse wavelet transformation (IWT) converts the frequency signal into image space. The reconstruction process can be summarized as follows:

$$I_B^* = IWT(H_{conv}(H_{RCABD}(F_B^n))). \tag{7}$$

IV. Experimental Results

To validate our proposed VDMamba, we conduct extensive experiments on both synthetic and real rainy datasets, comparing it against a range of representative image deraining methods. These methods include CNN-based DGUNet [62] and MPRNet [27], transformer-based Restormer [13], IDT [63], DRSformer [14], DAWN [17], SFNet [64], FMRNet [65], DPCNet [66] and NeRD [67], and Mamba-based TransMamba [68]. Five commonly used image quality evaluation metrics, such as Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), Feature Similarity (FSIM), Naturalness Image Quality Evaluator (NIQE) and Spatial-Spectral Entropy-based Quality (SSEQ) are employed for comprehensive evaluation.

Model	DPU	MRF	SU	RMB	PSNR	SSIM	Par.	Time	GFLOPs
Rain Image	–	–	–	–	22.16	0.732	–	–	–
w/o DPU	×	✓	✓	✓	32.86	0.919	1.324	0.089	75.62
w/o V-DPU	×	✓	✓	✓	33.01	0.920	1.321	0.088	74.56
w/o H-DPU	×	✓	✓	✓	33.12	0.923	1.321	0.088	74.56
w/o MRF	✓	×	✓	✓	33.14	0.921	1.296	0.082	72.35
w/o SU	✓	✓	×	✓	33.21	0.923	1.308	0.085	73.29
w/o RMB	✓	✓	✓	×	33.01	0.919	1.356	0.095	93.46
w/o DPU+MRF	×	×	✓	✓	32.71	0.917	1.319	0.086	73.65
w/o DPU+SU	×	✓	×	✓	32.69	0.916	1.304	0.085	71.96
w/o MRF+SU	✓	×	×	✓	33.02	0.920	1.315	0.083	73.48
w/o all	×	×	×	×	32.37	0.914	1.320	0.091	91.57
VDMamba	✓	✓	✓	✓	33.35	0.926	1.317	0.086	73.51

TABLE I: Ablation study on the Direction-aware projection unit (DPU), mutual representation and fusion (MRF), and the synthesis unit (SU) in the vector decomposition and synthesis module (VDSM), as well as the residual Mamba block (RMB) on the Test1200 dataset. We obtain the model parameters (Million (M)), inference time (Second (S)), and computation complexity (GFLOPs (G)) of deraining on images with 512×512 pixels.

A. Implementation Details

Data Collection. Following [26], we use $13,700$ clean/rain image pairs from [69], [70] for training all compared methods to guarantee fairness since these methods are originally trained with different datasets. In particular, the compared methods are retrained with the publicly released codes by tuning the optimal settings. For testing, four synthetic (Test100 [69], Test1200 [28], R100H, and R100L [71]) and three real-world datasets (Rain in Driving (RID), Rain in Surveillance (RIS) [72] and Real127 [28]) are considered for evaluation.

Experimental Setup. In our baseline model, we empirically set the number of Vector Decomposition and Synthesis Modules (VDSMs) to 10. Each VDSM involves vector decomposition, mutual representation and fusion, and vector synthesis. During the training process, the images are cropped into small patches of size 192×192 for convenience. We utilize the Adam optimizer with an initial learning rate of 2×10^{-4}, which decays at a rate of 0.8 every 80 epochs, continuing up to 500 epochs. The batch size is set to 8, and the model is trained on a single NVIDIA RTX 3090 GPU.

B. Ablation Studies

Validation on Basic Components. We conduct ablation studies to validate the contributions of individual components in vector decomposition and synthesis module (VDSM) to the final deraining performance on the Test1200 dataset. The components examined include the residual Mamba block (RMB), direction-aware projection unit (DPU), mutual representation and fusion (MRF), and synthesis unit (SU). For simplicity, we refer to our final model as VDMamba, which utilizes ten VDSMs. To evaluate the local-global representation of RMB, we design a *w/o RMB* model by replacing RMB with a transformer. Additionally, we devise six models (*w/o DPU*, *w/o MRF*, *w/o SU*, *w/o DPU+MRF*, *w/o DPU+SU*, and *w/o MRF+SU*) to investigate the effect of vector decomposition and synthesis. Two models (*w/o V-DPU* and *w/o H-DPU*) are used to separately evaluate the effects of vertical and horizontal projection units. Notably, these models consume approximately the same number of parameters as VDMamba.

Quantitative results regarding deraining performance and efficiency on the Test1200 dataset are presented in Table I. These results reveal that the complete deraining model, VDMamba, significantly outperforms its incomplete variants. Specifically, the introduction of the residual mamba block (RMB) yields an impressive performance gain of 0.34 dB in restoration quality while reducing computational costs by 21.4% (as seen in the results of VDMamba and *w/o RMB* models). We speculate that the RMB effectively balances local and global feature representation with linear complexity. Similarly, mutual representation and fusion (MRF) module enhances the model's ability to represent global rain distribution by leveraging complementary information between vertical and horizontal components, resulting in a 0.21 dB PSNR gain (referring to VDMamba and *w/o MRF* models). In contrast, replacing the any direction-aware projection unit (DPU) with standard convolution diminishes the model's ability to characterize direction-specific rain distribution, leading to a noticeable performance decline of 0.23 dB, 0.34 dB and 0.49 dB (as indicated by the results of VDMamba, *w/o V-DPU*, *w/o H-DPU* and *w/o DPU* models). It is noted that the vertical-DPU can provide more clues for modeling the rain distribution when compared to the horizontal one. Furthermore, removing synthesis unit (SU) significantly degrades the reconstruction capability of rain distribution, resulting in a noticeable performance decline of 0.14 dB in PSNR (comparing VDMamba with *w/o SU* models). Notably, when compared to the baseline model *w/o all*, VDMamba demonstrates considerable superiority, achieving a 0.98 dB improvement in restoration quality.

C. Comparison with State-of-the-arts

Synthesized Data. We compare the performance of VDMamba with several representative methods across four commonly used rainy image datasets. Additionally, we evaluate inference time, computational cost, and model parameters using an image size of 512×512. Quantitative results are provided in Table II. As anticipated, our proposed VDMamba demonstrates substantial performance improvement across all evaluation metrics compared to its competitors. Notably, VDMamba exhibits significant superiority over current CNN-based methods, surpassing DGUNet [62] and MPRNet [27] in PSNR by 0.84 dB and 1.25 dB, respectively, while greatly reducing the model parameters (47.5% and 47.6%) and inference time (59.9% and 67.7%). In comparison to transformer-based methods such as FMRNet [65], SFNet [64], and DRSformer [14], VDMamba remains highly competitive in deraining performance and demonstrates significant advantages in inference efficiency. On average, VDMamba outperforms the SOTA NeRD [67] by 0.58 dB and DRSformer by 0.60 dB in PSNR, respectively. Furthermore, *NeRD requires 16.38 $\times$ more model parameters, 3.43 $\times$ longer inference time, and 7.50 $\times$ greater computational cost* compared to our VDMamba. In contrast, the advantages of our proposed method in inference speed and storage indicate that VDMamba has greater potential for deployment on resource-constrained devices, such as edge or mobile devices.

To provide further evidence, we present visual comparisons across four datasets are shown in Fig. 3. Spatial domain-based methods, such as DRSformer [14], effectively eliminate

TABLE II: Comparison of average PSNR/SSIM/FSIM scores on Test100/Test1200 and R100H/R100L datasets. We obtain model parameters (Million), average inference time (Second) and computational cost (GFLOPs (G)) of deraining on images with the size of 512× 512.

Methods	MPRNet [27]	IDT [63]	DGUNet [62]	MFDNet [73]	DAWN [17]	DRSformer [14]	SFNet [64]	DPCNet [66]	FMRNet [65]	NeRD [67]	TransMamba [68]	VDMamba (Ours)
Test100/Test1200	30.27/32.91 0.897/0.916	29.69/31.38 0.905/0.908	30.32/33.23 0.899/0.920	30.78/33.01 0.914/0.925	29.86/32.76 0.902/0.919	**31.54**/32.04 0.917/0.918	31.16/32.04 0.917/0.904	30.78/33.23 0.914/**0.928**	30.66/33.21 0.915/0.924	31.13/31.62 0.913/0.916	30.81/33.02 0.911/0.921	31.24/**33.35** **0.918/0.928**
R100H/R100L	30.41/36.40 0.889/0.965	29.95/37.01 0.898/0.971	30.48/36.67 0.891/0.969	30.66/37.42 0.899/0.973	29.89/35.97 0.889/0.963	30.90/38.10 0.897/0.973	31.59/37.61 0.903/0.970	30.48/37.61 0.899/0.973	30.69/37.81 0.900/0.974	31.51/38.35 0.901/0.972	30.16/37.45 0.898/0.974	**31.62/38.76** **0.910/0.977**
Avg-PSNR ↑	32.49	32.00	32.90	32.97	32.12	33.14	33.10	33.02	33.09	33.16	32.86	**33.74**
Par.(M) ↓	3.637	16.41	3.633	4.741	1.489	33.65	13.23	–	1.551	22.89	15.49	**1.317**
Time (S) ↓	0.207	1.126	0.167	0.094	0.087	1.328	**0.082**	–	0.150	0.381	0.445	0.086
GFLOPs (G) ↓	565.8	247.63	–	275.57	89.34	971.6	497.7	–	91.23	625.2	327.5	**73.51**

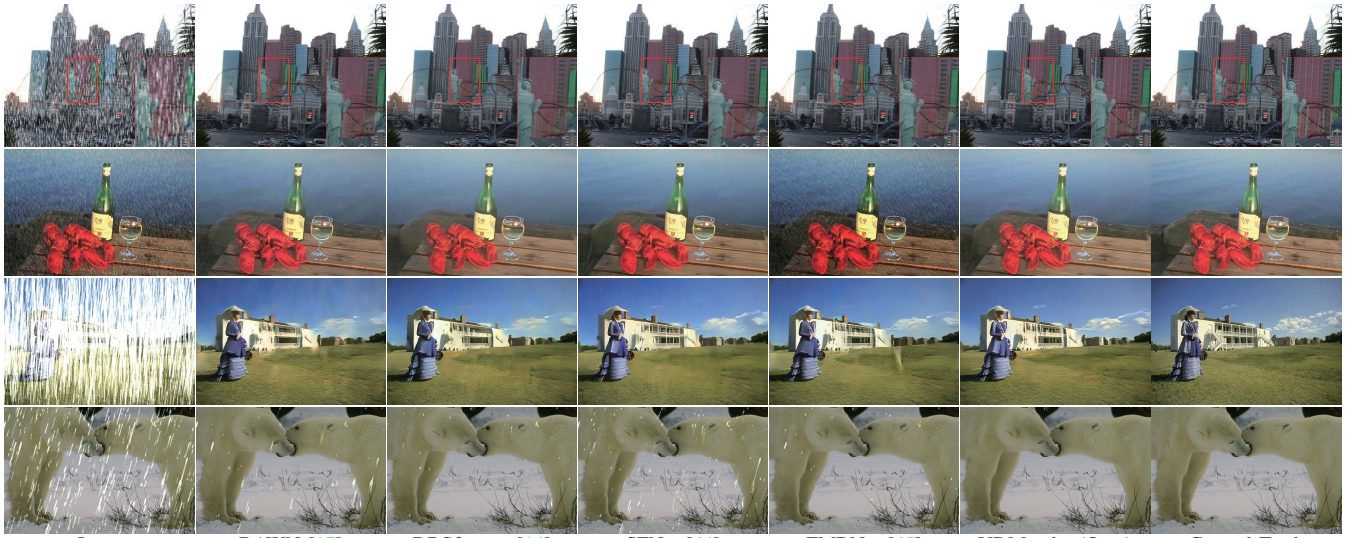

Fig. 3: Visual comparison of derained images obtained by five methods on Test100, Test1200, R100H and R100L datasets.

TABLE III: Comparison of average NIQE/SSEQ scores by seven deraining methods on three real-world datasets. The bold and underlined are the best and second performances, respectively.

Datasets	DRDNet [74]	MPRNet [27]	ELFormer [27]	DAWN [17]	MFDNet [73]	SFNet [64]	FMRNet [65]	NeRD [67]	VDMamba (Ours)
Real127	4.208/30.34	3.965/30.05	3.735/29.16	3.762/29.19	3.873/29.60	3.990/30.16	**3.714**/29.27	3.886/29.65	3.729/**29.07**
RID (2495)	5.715/39.98	6.452/40.16	**4.318**/37.89	4.785/37.49	6.468/42.51	5.352/39.28	4.431/37.25	4.892/38.74	4.369/**37.16**
RIS (2348)	6.269/45.34	6.610/48.78	5.835/**42.16**	5.536/43.16	**4.475**/45.24	5.934/46.39	5.628/42.29	5.783/43.59	5.107/42.23

rain perturbation in the image domain and thus bring an improvement in visibility. However, they often fail to produce visually appealing results due to residual artifacts in the frequency embedding space. The previous frequency domain-based methods (SFNet [64] and FMRNet [65]) promote spatial and frequency domain representation but exhibit unsatisfactory deraining performance due to the insufficient exploration of the Mamba architecture and Wavelet transform. Besides recovering cleaner and more credible image textures, our VDMamba produces results with better contrast and less color distortion. Specifically, for the second and fifth scenarios in Fig. 3, taking the "wine bottle" and "bear" as examples, only VDMamba infers clearer details with consistent contrast to the ground truth, while other methods produce results with oversmooth content or noticeable color distortion. Moreover, when dealing with heavy rain conditions, methods like DRSformer and SFNet struggle to eliminate rain streaks and often introduce visible artifacts and unnatural color appearance. In contrast, our VDMamba excels in completing the deraining task with the highest visual quality, producing more natural and credible details, as demonstrated in the third and fourth scenarios.

Real-world Data. We further conduct experiments on three real-world datasets: Real127 [28], Rain in Driving (RID), and Rain in Surveillance (RIS) [72]. RID and RIS respectively contain 2,495 and 2,348 real-world rain samples collected from car-mounted cameras and networked traffic surveillance cameras on rainy days. These datasets vary in rain types, image quality, object size, angle, and other factors, representing real application scenarios where deraining is desirable. Quantitative results for NIQE [76] and SSEQ [77] are listed in Table III, where smaller scores indicate better perceptual quality and clearer contents. When compared to SFNet [64], FMRNet [65] and NeRD [67], our VDMamba achieves competitive performance, recording the lowest average SSEQ scores in the Real127 and RID datasets, as well as the second best average scores on other evaluation metrics. Visual comparisons in Fig. 4 provide additional evidence to demonstrate that VDMamba surpasses other competing methods in effectively removing rain disturbances and preserving fine background details in the derained images.

Fig. 4: Visual comparison of derained images obtained by six methods on six real-world scenarios, covering heavy rain (3^{st}), light rain (1^{st}, 2^{st} and 6^{st}) and rain veiling effect (4^{st} and 5^{st}). Please refer to the highlighted regions for a close-up comparison.

TABLE IV: Comparison of average PSNR/SSIM scores on the UAV-Rain1k dataset [78]. We obtain the model parameter (Million) and inference time (Second) of deraining on images with the size of 512×512.

Methods	Input	IDT [63]	Restormer [13]	SFNet [79]	MFDNet [73]	DRSformer [14]	DAWN [17]	FMRNet [65]	VDMamba (Ours)
PSNR	16.80	22.47	24.78	24.69	22.47	<u>24.93</u>	24.32	23.28	**25.46**
SSIM	0.719	0.895	0.905	0.911	0.881	<u>0.915</u>	0.914	0.903	**0.918**
Par.(M)	–	16.41	26.12	13.23	4.741	33.65	<u>1.487</u>	1.551	**1.317**
Time (S)	–	1.126	1.124	0.082	0.094	1.328	<u>0.087</u>	0.150	**0.086**

TABLE V: Comparison of average PSNR/SSIM scores on the RainCityscapes dataset [80]. We obtain the model parameter (Million) and inference time (Second) of deraining on images with the size of 512×512.

Methods	Input	PCNet [59]	Restormer [13]	SCDformer [75]	SFNet [79]	MFDNet [73]	DAWN [17]	FMRNet [65]	VDMamba (Ours)
PSNR	16.87	24.49	<u>28.42</u>	27.52	27.44	26.91	26.72	28.24	**30.35**
SSIM	0.851	0.946	<u>0.959</u>	0.928	0.951	0.948	0.948	0.950	**0.962**
Par.(M)	–	**0.655**	26.12	2.086	13.23	4.741	1.489	1.551	<u>1.317</u>
Time (S)	–	**0.062**	1.124	2.252	0.082	0.094	0.087	0.150	<u>0.086</u>

D. Generality for Other Image Restoration Tasks

To assess the versatility and effectiveness of our vector decomposition and synthesis strategy in VDMamba model, we have extended its application to address various image restoration tasks, encompassing raindrop removal [78], rain-haze removal [80], underwater image enhancement [81], image dehazing [82], and low-light image enhancement [83].

1) Raindrop Removal: Raindrops on a camera lens can cause undesired artifacts, such as blurriness, distortions, and color aberrations in the captured images. Effective raindrop removal is crucial for various applications, including surveillance, autonomous vehicles, and outdoor photography. To evaluate the versatility, we compared VDMamba to the typical image restoration methods, such as Restormer [13], IDT [63], SFNet [79], MFDNet [73], DAWN [17], DRSformer [14] and FMRNet [65] on the UAV-Rain1k dataset [78]. The UAV-Rain1k contains 800 and 220 synthetic images for training and testing, respectively. The average resolution of all images is 1500×1000, covering four rain density labels (*e.g.*, light, moderate, heavy, and irregular). For fair comparison, we retrain the aforementioned methods with the UAV-Rain1k dataset following their released source codes and public settings. Quantitative results in Table IV indicate that our proposed VDMamba method performs favorably, achieving the highest scores across

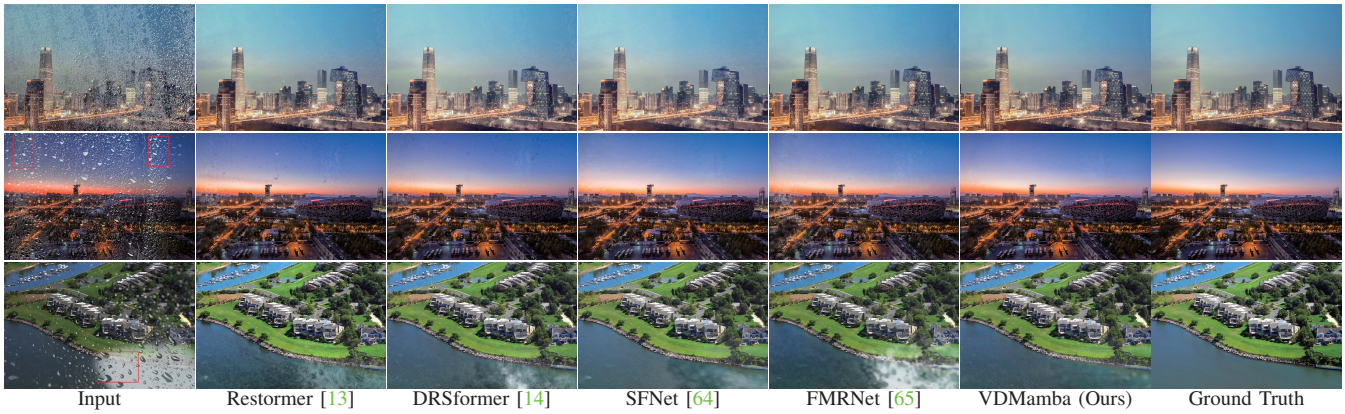

Fig. 5: Visual comparisons of derained images obtained by five methods on the UAV-Rain1k dataset [78].

Fig. 6: Visual comparison of derained images obtained by five methods on the RainCityscapes dataset [80]. Please refer to the detailed comparison in the red regions.

all metrics, particularly when compared to Restormer and SFNet, tailored for image restoration tasks. It is noted that our VDMamba surpasses the SOTA method DRSformer by 0.53 dB in terms of PSNR, while significantly saving 96.1% and 93.6% of the model parameters and inference time. Fig. 5 presents visual comparisons, demonstrating that our VDMamba outperforms other methods in effectively removing rain disturbances and inferring fine background details in the degraded regions, more faithful to the ground truth. In contrast, the compared methods are inclined to produce results with either color distortion or obvious artifacts residue (Please refer to the highlighted regions).

2) Rainhaze Removal: Due to the symbiotic nature of rain and fog in natural scenes, we further compare our proposed VDMamba method with several deraining methods on the rain-haze dataset (RainCityscapes [80]), involving PCNet [59], Restormer [13], SCDformer [75], MFDNet [73], DAWN [17], SFNet [79] and FMRNet [65]. The RainCityscapes dataset [80] is synthesized from the Cityscapes dataset [85]. The training set consists of 9432 images created from 262 Cityscapes images, while the test set includes 1188 images synthesized from 33 Cityscapes images. All Cityscapes images selected for synthesis depict overcast conditions without prominent shadows. Rain streaks and haze are added using various intensity maps, allowing for the generation of 36 different synthesized images from each original image by adjusting the intensity of these effects. To ensure a fair comparison, we trained the previously mentioned methods on the RainCityscapes following their public settings. The quantitative results in Table V indicate that our proposed VDMamba method achieves the highest scores in terms of PSNR and SSIM, outperforming the top-performing methods Restormer and FMRNet by 2.11 dB and 1.93 dB, respectively. Visual comparisons in Fig. 6 demonstrate that our VDMamba is more effective in removing rain and haze disturbances while preserving fine background details in the predicted images. In contrast, the compared

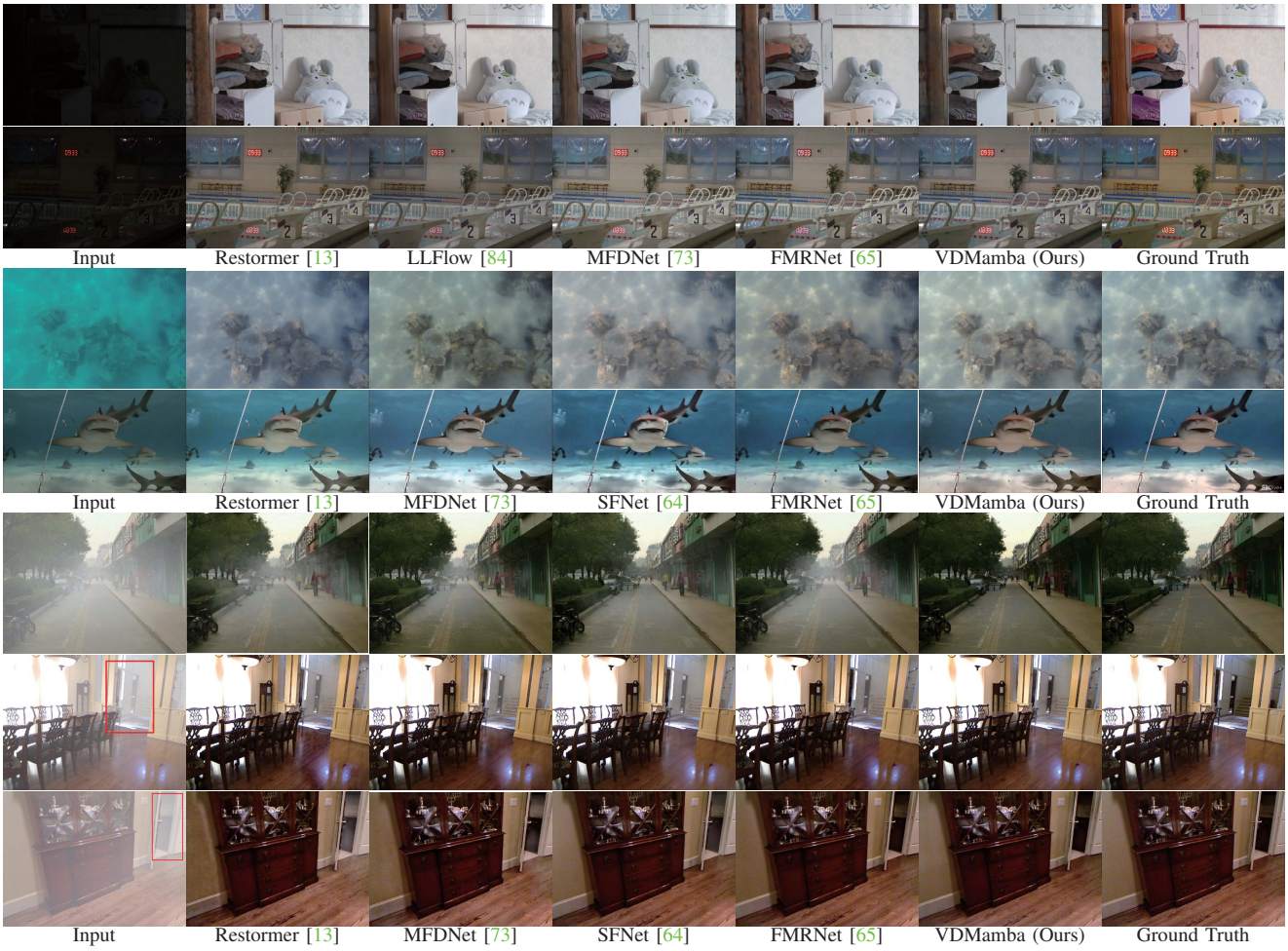

Fig. 7: Visual comparisons of low-light image enhancement, underwater image enhancement and image dehazing.

TABLE VI: Quantitative performance comparison of underwater image enhancement in terms of PSNR on the S1000 [86] and R90 [87] datasets.

Methods	Ucolor [88]	NAFNet [89]	Uformer [12]	MLLE [90]	Restormer [13]	UTUIE [81]	MFDNet [73]	DAWN [17]	FMRNet [65]	VDMamba (Ours)
S1000	23.05	22.08	21.54	21.38	28.77	23.42	27.79	25.31	<u>28.97</u>	**29.74**
R90	20.63	21.15	20.76	20.47	23.18	22.91	<u>23.97</u>	21.49	23.93	**24.38**
Ave-PSNR	21.84	21.61	21.15	20.92	25.97	23.16	25.88	23.40	<u>26.45</u>	**27.06**

methods struggle to restore clear background contents and exhibit obvious artifacts and contrast distortion (please refer to the results of SCDformer, Restormer, and MFDNet).

3) Underwater Image Enhancement: Improving the quality of underwater images via underwater image enhancement technologies is essential for various underwater applications. In accordance with [91], our VDMamba model is trained on a dataset of 2050 pairs of underwater images, covering a wide range of scenes. Among these pairs, 800 are randomly selected from the UIEB dataset [87], while the remaining 1250 pairs are from a synthetic underwater image dataset [86]. To assess the model's performance, we test our VDMamba and all compared methods on two datasets: R90 and S1000. The R90 dataset is a subset of the UIEB dataset [87], consisting of 90 pairs of real underwater images and their corresponding reference images. While the S1000 dataset contains 1000 pairs of synthetic underwater images from [86].

Table VI presents the quantitative results in terms of PSNR, comparing VDMamba to five commonly used underwater image enhancement methods and four image restoration technologies. As expected, our VDMamba consistently outperforms all comparison models by a significant margin across two testing datasets, on average surpassing the state-of-the-art methods Restormer [13] and FMRNet [65] by 0.97 dB and 0.77 dB, respectively. This performance can be attributed to the effectiveness of the vector decomposition and mutual representation among vertical and horizontal components, which excel in learning the direction-specific background textures. In particular, VDMamba effectively eliminates degradation in underwater images without relying on specific knowledge of underwater imaging, such as the reverse medium transmission map. Visual comparisons presented in Fig. 7 demonstrate that our VDMamba produces sharper contents with more accurate color tones while successfully removing the adverse

TABLE VII: Quantitative performance comparison of image dehazing in terms of PSNR/SSIM of seven image dehazing methods on the SOTS dataset.

Methods	FFANet [92]	DANet [29]	Restormer [13]	MixDehazeNet [93]	DehazeFormer [94]	MFDNet [73]	FMRNet [65]	VDMamba (Ours)
PSNR ↑	26.34	30.16	28.30	30.18	30.36	29.45	<u>30.82</u>	**30.94**
SSIM ↑	0.953	0.967	0.956	0.973	0.973	<u>0.976</u>	0.968	**0.978**
Par. (M) ↓	4.456	1.547	26.12	3.16	**0.686**	4.741	1.551	<u>1.317</u>

TABLE VIII: Quantitative performance comparison of several low-light image enhancement methods on the LOL [95] dataset. We obtain the model parameter (Million) and average inference time (Second) of relighting on images with the size of **384**× **384**.

Methods	SNRNet [83]	Restormer [13]	DAWN [17]	LLFormer [84]	IAGC [96]	SFNet [64]	MFDNet [73]	FMRNet [65]	VDMamba (Ours)
PSNR ↑	24.61	26.63	24.77	23.64	24.53	24.95	25.01	<u>25.97</u>	**26.78**
SSIM ↑	0.842	0.909	0.911	0.816	0.842	0.901	<u>0.912</u>	0.908	**0.915**
Par.(M) ↓	39.12	26.12	<u>1.487</u>	24.52	–	13.23	4.741	1.551	**1.317**
Time (S) ↓	0.320	0.568	0.049	0.568	–	**0.048**	0.053	0.055	**0.048**

underwater effects, resulting in outputs more faithful to the corresponding ground truth. In contrast, Restormer [13] and MFDNet [73] fail to eliminate the greenish color, resulting in under-enhanced outputs. Although SFNet [64] and FMR-Net [65] perform better than the aforementioned methods, they still introduce color deviations.

4) Image dehazing: Fog and haze present significant challenges in computer vision tasks, and image dehazing aims to recover clear images from hazy input [97]. As indicated in [9], achieving the global representation is important for image dehazing [82]. Thus, the proposed VDMamba can also be applied to this task. In line with [59], we utilize 6000 pairs of synthesized hazy images from [98] to train our VDMamba and the compared methods. For evaluation, we use the commonly used SOTS (synthetic objective testing set) dataset [98], which consists of 1000 pairs of indoor and outdoor hazy images selected from the RESIDE dataset.

The quantitative results presented in Table VII demonstrate that our method outperforms existing dehazing technologies by a significant margin. Specifically, our VDMamba achieves higher PSNR and SSIM scores on the SOTS dataset, surpassing DehazeFormer [94] and FMRNet [65] by 0.58 dB and 0.12 dB, respectively. Visual comparisons in Fig. 7 illustrate that other models produce under-dehazed outputs or introduce noticeable color distortion, while our VDMamba method effectively removes the disruptive effects of haze, generating clearer and visually plausible results with minimal residue. This substantial improvement in both visual quality and evaluation scores highlights the effectiveness and versatility of our proposed vector decomposition-based vision Mamba framework.

5) Low-Light Image Enhancement: Low-light image enhancement techniques are designed to improve the brightness and reduce noise of images captured in low-light conditions. Since exploring global correlations is very helpful for noise reduction [99], to evaluate the versatility of our VDMamba, we compare it to conventional low-light image enhancement methods using the widely used LOL [95] dataset. This dataset comprises a training set of 485 pairs of low-light and normal-light images, along with a test set containing 15 pairs.

The quantitative results, presented in Table VIII, indicate that our VDMamba demonstrates impressive competitiveness. It achieves the highest PSNR and SSIM scores, along with notable superiority in inference speed and model parameters. When compared to the state-of-the-art methods Restormer [13] and FMRNet [65], VDMamba surpasses them by 0.15 dB and 0.81 dB, respectively. This showcases its ability to preserve detailed structural and textural information while effectively suppressing artifacts. Visual comparisons shown in Fig. 7 reveal that the competing methods tend to produce results with color distortions, either through under- or over-relighting, or artifacts. In contrast, our proposed VDMamba effectively relights dark regions, yielding more credible content that aligns closely with the ground truth. It can be asserted that our mamba-based vector decomposition and synthesis strategy favors suppressing the artifacts and inferring structure details.

V. Conclusion

In this paper, we present VDMamba, a novel image deraining framework leverages vector decomposition and mamba-based wavelet transform approach (VDMamba). Specifically, we investigate the potential relationships among rain distribution representation and Mamba characteristics in the wavelet domain, constructing a vector decomposition and synthesis module (VDSM) to progressively separate rain perturbations from the clean background while preserving the structural textures of degraded images. VDSM provides an alternative solution for effective local-global degradation representation while requiring no specific scanning design. In particular for the image restoration and enhancement tasks, VDMamba proposes to transform the prediction and removal of arbitrary perturbations (ignoring pattern and intensity) into a parameterized fitting problem in the frequency space, which is of better robustness and practicality. Extensive experiments on synthetic and real-world rain-image datasets demonstrate the superior performance of VDMamba compared to several state-of-the-art image deraining methods. Beyond image deraining, we showcase the versatility of VDMamba by extending its application to various image restoration tasks, including raindrop removal, rain-haze removal, underwater image enhancement, image dehazing, and low-light image enhancement. This work opens new avenues for future research in Mamba-based image restoration, particularly in deploying efficient and high-performing models on resource-constrained devices.

Limitations and Future Work. In this work, we propose a vector decomposition-based vision Mamba approach (VD-Mamba) to investigate the coincidence between characteristics of rainy images and Mamba architecture. Although

VDMamba specializes in eliminating the aliasing artifact and preserving structure, it is important to acknowledge its limitations and avenues for future exploration. One limitation is that it behaves with significant heterogeneity in different degradation patterns or scenarios. For example, the low-light image enhancement task shows lower PSNR than other tasks. The specific framework and decomposition pattern may be required to investigate and explore the compatibility of Mamba architecture to these tasks. Also, the prompt learning with a large model (LM) may narrow the knowledge gap derived from the degradation diversity. In addition, the absence of a comprehensive analysis for the physics degradation model is evident. Although VDMamba shows better robustness and practicality for arbitrary perturbations (ignoring pattern and intensity), leveraging physics models to facilitate degradation learning could provide more robust and generalizable priors.

REFERENCES

[1] R. Zhang, J. Yu, J. Chen, G. Li, L. Lin, and D. Wang, "A prior guided wavelet-spatial dual attention transformer framework for heavy rain image restoration," *IEEE Transactions on Multimedia*, 2024.

[2] Y. Wang, D. Gong, J. Yang, Q. Shi, D. Xie, B. Zeng *et al.*, "Deep single image deraining via modeling haze-like effect," *IEEE Transactions on Multimedia*, vol. 23, pp. 2481–2492, 2020.

[3] Y. Que, S. Li, and H. J. Lee, "Attentive composite residual network for robust rain removal from single images," *IEEE Transactions on Multimedia*, vol. 23, pp. 3059–3072, 2020.

[4] L. Zhao, H. Zhou, X. Zhu, X. Song, H. Li, and W. Tao, "Lif-seg: Lidar and camera image fusion for 3d lidar semantic segmentation," *IEEE Transactions on Multimedia*, vol. 26, pp. 1158–1168, 2023.

[5] X. Lin, L. Ma, B. Sheng, Z.-J. Wang, and W. Chen, "Utilizing two-phase processing with fbls for single image deraining," *IEEE Transactions on Multimedia*, vol. 23, pp. 664–676, 2020.

[6] L. Yu, B. Wang, J. He, G.-S. Xia, and W. Yang, "Single image deraining with continuous rain density estimation," *IEEE Transactions on Multimedia*, vol. 25, pp. 443–456, 2021.

[7] J. Bossu, N. Hautière, and J.-P. Tarel, "Rain or snow detection in image sequences through use of a histogram of orientation of streaks," *IJCV*, vol. 93, no. 3, pp. 348–367, 2011.

[8] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *CVPR*, 2013, pp. 1968–1975.

[9] C. Zheng, Z. Li, Y. Li, and S. Wu, "Non-local single image de-raining without decomposition," in *ICASSP*. IEEE, 2021, pp. 1425–1429.

[10] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *CVPR*, 2016, pp. 1674–1682.

[11] S. Sun, W. Ren, X. Gao, R. Wang, and X. Cao, "Restoring images in adverse weather conditions via histogram transformer," in *European Conference on Computer Vision*. Springer, 2025, pp. 111–129.

[12] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *CVPR*, 2022, pp. 17 683–17 693.

[13] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *CVPR*, 2022, pp. 5718–5729.

[14] X. Chen, H. Li, M. Li, and J. Pan, "Learning a sparse transformer network for effective image deraining," in *CVPR*, 2023, pp. 5896–5905.

[15] Z. Zou, H. Yu, J. Huang, and F. Zhao, "Freqmamba: Viewing mamba from a frequency perspective for image deraining," in *ACM MM*, 2024, pp. 1905–1914.

[16] H. Huang, A. Yu, Z. Chai, R. He, and T. Tan, "Selective wavelet attention learning for single image deraining," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1282–1300, 2021.

[17] K. Jiang, W. Liu, Z. Wang, X. Zhong, J. Jiang, and C.-W. Lin, "Dawn: Direction-aware attention wavelet network for image deraining," in *ACM MM*, 2023, pp. 7065–7074.

[18] X. Li, Z. Wang, and C. Xie, "An inverse scaling law for clip training," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36, 2023, pp. 49 068–49 087.

[19] A. Gu, I. Johnson, K. Goel, K. Saab, T. Dao, A. Rudra, and C. Ré, "Combining recurrent, convolutional, and continuous-time models with linear state space layers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 572–585, 2021.

[20] A. Gupta, A. Gu, and J. Berant, "Diagonal state spaces are as effective as structured state spaces," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22 982–22 994, 2022.

[21] D. Li, Y. Liu, X. Fu, S. Xu, and Z.-J. Zha, "Fouriermamba: Fourier learning integration with state space models for image deraining," *arXiv preprint arXiv:2405.19450*, 2024.

[22] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-frame-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 3888–3901, 2012.

[23] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[24] K. Jiang, Z. Wang, P. Yi, C. Chen, X. Wang, J. Jiang, and Z. Xiong, "Multi-level memory compensation network for rain removal via divide-and-conquer strategy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 216–228, 2021.

[25] R. Li, L.-F. Cheong, and R. T. Tan, "Single image deraining using scale-aware multi-stage recurrent network," *arXiv preprint arXiv:1712.06830*, 2017.

[26] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang, "Multi-scale progressive fusion network for single image deraining," in *CVPR*, 2020, pp. 8343–8352.

[27] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *CVPR*, 2021.

[28] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *CVPR*, 2018, pp. 695–704.

[29] J. Kui, W. Zhongyuan, W. Zheng, Y. Peng, J. Junjun, X. Jinsheng, and L. Chia-Wen, "Danet: Image deraining via dynamic association learning," in *IJCAI*, 2022.

[30] G. Li, X. He, W. Zhang, H. Chang, L. Dong, and L. Lin, "Non-locally enhanced encoder-decoder network for single image de-raining," in *ACM MM*, 2018, pp. 1056–1064.

[31] Y. Nanba, H. Miyata, and X.-H. Han, "Dual heterogeneous complementary networks for single image deraining," in *CVPR*, 2022, pp. 568–577.

[32] N. Gao, X. Jiang, X. Zhang, and Y. Deng, "Efficient frequency-domain image deraining with contrastive regularization," in *European Conference on Computer Vision*. Springer, 2025, pp. 240–257.

[33] Z. Bao, M. Shao, Y. Wan, and Y. Qiao, "Recursive residual fourier transformation for single image deraining," *International Journal of Machine Learning and Cybernetics*, vol. 15, no. 5, pp. 1743–1754, 2024.

[34] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[35] Y. Huang, J. Huang, J. Liu, M. Yan, Y. Dong, J. Lv, C. Chen, and S. Chen, "Wavedm: Wavelet-based diffusion models for image restoration," *IEEE Transactions on Multimedia*, vol. 26, pp. 7058–7073, 2024.

[36] W. Yang, J. Liu, S. Yang, and Z. Guo, "Scale-free single image deraining via visibility-enhanced recurrent wavelet learning," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2948–2961, 2019.

[37] H.-H. Yang, C.-H. H. Yang, and Y.-C. F. Wang, "Wavelet channel attention module with a fusion network for single image deraining," in *ICIP*. IEEE, 2020, pp. 883–887.

[38] X. Jiang, X. Zhang, N. Gao, and Y. Deng, "When fast fourier transform meets transformer for image restoration," in *European Conference on Computer Vision*. Springer, 2025, pp. 381–402.

[39] J. T. Smith, A. Warrington, and S. Linderman, "Simplified state space layers for sequence modeling," in *International Conference on Learning Representations*, 2022.

[40] X. Pei, T. Huang, and C. Xu, "Efficientvmamba: Atrous selective scan for light weight visual mamba," *arXiv preprint arXiv:2403.09977*, 2024.

[41] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," in *International Conference on Machine Learning*, 2024.

[42] K. Li, X. Li, Y. Wang, Y. He, Y. Wang, L. Wang, and Y. Qiao, "Videomamba: State space model for efficient video understanding," in *European Conference on Computer Vision*. Springer, 2025, pp. 237–255.

[43] W. Dong, H. Zhu, S. Lin, X. Luo, Y. Shen, X. Liu, J. Zhang, G. Guo, and B. Zhang, "Fusion-mamba for cross-modality object detection," *arXiv preprint arXiv:2404.09146*, 2024.

[44] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, "Vmamba: Visual state space model," *arXiv preprint arXiv:2401.10166*, 2024.

[45] J. Ma, F. Li, and B. Wang, "U-mamba: Enhancing long-range dependency for biomedical image segmentation," *arXiv preprint arXiv:2401.04722*, 2024.

[46] A. Gu, K. Goel, and C. Re, "Efficiently modeling long sequences with structured state spaces," in *International Conference on Learning Representations*, 2022.

[47] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[48] A. Jiang, H. Chen, Z. Chen, J. Ye, and M. Wang, "Multi-dimensional visual prompt enhanced image restoration via mamba-transformer aggregation," *arXiv preprint arXiv:2412.15845*, 2024.

[49] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia, "Mambair: A simple baseline for image restoration with state-space model," *arXiv preprint arXiv:2402.15648*, 2024.

[50] H. Wu, Y. Yang, A. I. Aviles-Rivero, J. Ren, S. Chen, H. Chen, and L. Zhu, "Semi-supervised video desnowing network via temporal decoupling experts and distribution-driven contrastive regularization," in *European Conference on Computer Vision*. Springer, 2025, pp. 70–89.

[51] M. V. Conde, G. Geigle, and R. Timofte, "Instructir: High-quality image restoration following human instructions," in *European Conference on Computer Vision*. Springer, 2025, pp. 1–21.

[52] Y. Xiao, Q. Yuan, K. Jiang, Y. Chen, Q. Zhang, and C.-W. Lin, "Frequency-assisted mamba for remote sensing image super-resolution," *arXiv preprint arXiv:2405.04964*, 2024.

[53] Z. Zheng and C. Wu, "U-shaped vision mamba for single image dehazing," *arXiv preprint arXiv:2402.04139*, 2024.

[54] Z. Zhen, Y. Hu, and Z. Feng, "Freqmamba: Viewing mamba from a frequency perspective for image deraining," *arXiv preprint arXiv:2404.09476*, 2024.

[55] Y. Chen, H. Qin, Z. Zhang, M. Magno, L. Benini, and Y. Li, "Q-mambair: Accurate quantized mamba for efficient image restoration," *arXiv preprint arXiv:2503.21970*, 2025.

[56] L. Peng, X. Di, Z. Feng, W. Li, R. Pei, Y. Wang, X. Fu, Y. Cao, and Z.-J. Zha, "Directing mamba to complex textures: An efficient texture-aware state space model for image restoration," *arXiv preprint arXiv:2501.16583*, 2025.

[57] Z. Liu, S. Zhou, Y. Dai, Y. Wang, Y. An, and X. Zhao, "Dpmambair: All-in-one image restoration via degradation-aware prompt state space model," *arXiv preprint arXiv:2504.17732*, 2025.

[58] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution," in *ICCV*, 2017, pp. 1689–1697.

[59] K. Jiang, Z. Wang, P. Yi, C. Chen, Z. Wang, X. Wang, J. Jiang, and C.-W. Lin, "Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining," *IEEE Trans. Image Process.*, vol. 30, pp. 7404–7418, 2021.

[60] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *CVPR*, 2017, pp. 624–632.

[61] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[62] C. Mou, Q. Wang, and J. Zhang, "Deep generalized unfolding networks for image restoration," in *CVPR*, 2022, pp. 17 399–17 410.

[63] J. Xiao, X. Fu, A. Liu, F. Wu, and Z.-J. Zha, "Image de-raining transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 12 978–12 995, 2023.

[64] Y. Cui, Y. Tao, Z. Bing, W. Ren, X. Gao, X. Cao, K. Huang, and A. Knoll, "Selective frequency network for image restoration." in *The Eleventh International Conference on Learning Representations*, 2023.

[65] K. Jiang, J. Jiang, X. Liu, X. Xu, and X. Ma, "Fmrnet: Image deraining via frequency mutual revision," in *AAAI*, vol. 38, no. 11, 2024, pp. 12 892–12 900.

[66] Y. He, A. Jiang, L. Jiang, Z. Wang, and L. Wang, "Dual-path coupled image deraining network via spatial-frequency interaction," *arXiv preprint arXiv:2402.04855*, 2024.

[67] X. Chen, J. Pan, and J. Dong, "Bidirectional multi-scale implicit neural representations for image deraining," in *CVPR*, 2024, pp. 25 627–25 636.

[68] S. Sun, W. Ren, J. Zhou, J. Gan, R. Wang, and X. Cao, "A hybrid transformer-mamba network for single image deraining," *arXiv preprint arXiv:2409.00410*, 2024.

[69] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, 2020.

[70] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *CVPR*, 2017, pp. 3855–3863.

[71] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *CVPR*, 2017, pp. 1357–1366.

[72] S. Li, I. B. Araujo, W. Ren, Z. Wang, E. K. Tokuda, R. H. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao, "Single image deraining: A comprehensive benchmark analysis," in *CVPR*, 2019, pp. 3838–3847.

[73] Q. Wang, K. Jiang, Z. Wang, W. Ren, J. Zhang, and C.-W. Lin, "Multi-scale fusion and decomposition network for single image deraining." *IEEE Trans. Image Process.*, vol. 33, pp. 191–204, 2024.

[74] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. L. Wang, and M. Wang, "Detail-recovery image deraining via context aggregation networks," in *CVPR*, 2020, pp. 14 548–14 557.

[75] Y. Guo, X. Xiao, Y. Chang, S. Deng, and L. Yan, "From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal," in *ICCV*, 2023, pp. 12 063–12 073.

[76] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.

[77] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *SPIC*, vol. 29, no. 8, pp. 856–863, 2014.

[78] W. Chang, H. Chen, X. He, X. Chen, and L. Shen, "Uav-rain1k: A benchmark for raindrop removal from UAV aerial imagery," *arxiv*, vol. abs/2402.05773, 2024.

[79] Y. Cui, Y. Tao, Z. Bing, W. Ren, X. Gao, X. Cao, K. Huang, and A. Knoll, "Selective frequency network for image restoration," in *The Eleventh International Conference on Learning Representations ICLR*, 2023.

[80] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, "Depth-attentional features for single-image rain removal," in *CVPR*, 2019, pp. 8014–8023.

[81] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023.

[82] Z. Li, C. Zheng, H. Shu, and S. Wu, "Dual-scale single image dehazing via neural augmentation," *IEEE Trans. Image Process.*, vol. 31, pp. 6213–6223, 2022.

[83] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *CVPR*, 2022, pp. 17 714–17 724.

[84] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *AAAI*, vol. 37, no. 3, 2023, pp. 2654–2662.

[85] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *CVPR*, 2016, pp. 3213–3223.

[86] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, p. 107038, 2020.

[87] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2019.

[88] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, 2021.

[89] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *ECCV*. Springer, 2022, pp. 17–33.

[90] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.

[91] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.

[92] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *AAAI*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[93] L. Lu, Q. Xiong, B. Xu, and D. Chu, "Mixdehazenet: Mix structure block for image dehazing network," in *IJCNN*. IEEE, 2024, pp. 1–10.

[94] 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.

[95] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *BMVC*, 2018, p. 155.

[96] Y. Wang, Z. Liu, J. Liu, S. Xu, and S. Liu, "Low-light image enhancement with illumination-aware gamma correction and complete image modelling network," in *ICCV*, 2023, pp. 13 128–13 137.
[97] B. Ding, R. Zhang, L. Xu, G. Liu, S. Yang, Y. Liu, and Q. Zhang, "U 2 d 2 net: Unsupervised unified image dehazing and denoising network for single hazy image enhancement," *IEEE Transactions on Multimedia*, vol. 26, pp. 202–217, 2023.
[98] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, 2018.
[99] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *CVPR*, vol. 2, 2005, pp. 60–65.