2025 International Joint Conference on Neural Networks (IJCNN) | 979-8-3315-1042-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/IJCNN64981.2025.11228128

SWR-BIDeN: An Improved BIDeN Model for Severe Weather Removal in Image Processing

Yixin Zhang[†1,2], Tianze Zhang[†2,3], Xingbiao Zhou[1,2], Gang Shi[*1,2]
[1]School of Computer Science and Technology, Xinjiang University, Urumqi, China
[2]Xinjiang Key Laboratory of Signal Detection and Processing, Xinjiang University, Urumqi, China
[3]Faculty of Science, The University of Melbourne, Melbourne, Australia
Emails: 107552204132@stu.xju.edu.cn, zhangtianze.unimelb@gmail.com,
107552204142@stu.xju.edu.cn, shigang@xju.edu.cn

***Abstract*—Image restoration plays a crucial role in improving the visual quality of images under severe weather conditions. However, existing methods often face challenges such as high computational complexity, limited feature representation, and suboptimal perceptual fidelity. To address these issues, we propose SWR-BIDeN, an improved version of the Blind Image Decomposition Network (BIDeN). SWR-BIDeN introduces the lightweight BSConv-Attention Residual Block (BARB), which utilizes blueprint separable convolution (BSConv) to replace standard convolution and incorporates attention mechanisms. This significantly improves feature representation while reducing model size. Additionally, we propose a Channel-Spatial Integration Attention Module (CSIAM), which captures cross-dimensional dependencies to strengthen feature representation and contextual understanding. Furthermore, we adopt a dual perceptual constraint strategy by combining LPIPS and VGG losses. This approach effectively prevents over-smoothing and detail loss, ensuring that the restored images align better with human visual perception. The results show that SWR-BIDeN model achieves an average PSNR of 23.69 and an average SSIM of 0.824 on the WeatherStream dataset. We also conducted extensive experiments on synthetic datasets, where the SWR-BIDeN still maintains excellent performance under the most challenging mixed weather conditions, achieving a PSNR of 22.80 and an SSIM of 0.832. Compared to the BIDeN model, our method reduces the number of parameters and FLOPs by 29.5% and 15.6%, respectively. With its balance of efficiency and robustness, SWR-BIDeN is well-suited for practical real-world vision applications.**

***Index Terms*—severe weather removal, image restoration, attention mechanism, perceptual loss.**

I. Introduction

Computer vision tasks have become a core component of computer applications. However, severe weather conditions, such as rain, snow, and haze, can significantly degrade image quality and pose substantial challenges for computer vision tasks like object detection, segmentation, and autonomous driving. The degradation caused by these weather conditions not only blurs key features but may also obscure important semantic information, thus affecting the performance of systems in real-world applications. Therefore, developing technologies capable of effectively restoring weather-degraded images is of great significance for improving the robustness and reliability of these systems.

Fig. 1. Degraded images from WeatherStream. The lower part is the corresponding clear images.

Traditional image restoration methods typically rely on manually designed prior knowledge or physical models [1]. While these methods excel in specific degradation scenarios, they lack adaptability to complex weather conditions and have obvious limitations in generalization ability and processing efficiency. Furthermore, the weather conditions in real-world scenarios are often highly dynamic and uncertain, further limiting the performance of traditional methods.

In recent years, the rapid development of deep learning technologies has provided new insights into image restoration problems. Deep learning models can learn complex patterns and implicit correlations from large-scale data, significantly enhancing image processing performance. Against this backdrop, task-specific models targeting specific weather conditions, such as ID-CGAN [2], Desnownet [3], and AOD-Net[4], as well as all-in-one frameworks capable of handling multiple weather types, such as TransWeather [5], WGWS [6], and LDR [7], have emerged. While these two strategies have made breakthroughs in terms of accuracy and adaptability, they still face challenges such as high model complexity, insufficient feature representation, and limited perceptual quality. These limitations are particularly evident when processing degraded

This research was supported by the Key Research and Development Project in Xinjiang Uygul Autonomous Region (No.2022B01006) and the National Natural Science Foundation of China(No.62162059).

* Corresponding author.
† These authors contributed equally to this work.

image data from real-world environments.

To address the above issues, this paper proposes an improved Blind Image Decomposition Network (BIDeN) [8] to tackle the image degradation problem under severe weather conditions. By incorporating the BSConv-Attention Residual Block (BARB), Channel-Space Integration Attention Module (CSIAM), and the LPIPS Loss [9], the proposed model significantly improves feature extraction efficiency and perceptual quality, while reducing computational complexity. The main contributions of this paper can be summarized as follows:

1. A lightweight residual block, BSConv-Attention Residual Block(BARB), is designed to replace the residual block in the encoder. This module uses Blueprint Separable Convolution (BSConv) [10] in place of traditional convolution and integrates spatial and channel attention mechanisms, significantly improving feature extraction performance while reducing the number of parameters by 12.5M and computational cost by 100.4 GFLOPs.

2. A Channel-Space Integration Attention Module (CSIAM) is proposed to adjust the extracted feature vectors, enhancing feature representation and contextual understanding, thereby optimizing the subsequent generation tasks.

3. The model's perceptual loss is improved by introducing the LPIPS loss, ensuring that the generated images align more closely with human visual perception, thus enhancing the overall perceptual quality.

4. Extensive experiments were conducted on both the CityScapes-based synthetic dataset and the WeatherStream dataset, with thorough analysis of the results. The findings demonstrate that the proposed network significantly enhances image restoration capability while exhibiting strong robustness in complex scenarios.

II. Related Work

A. Task-Specific Methods

Task-specific methods focus on removing specific types of weather degradations, such as rain, snow, or haze. These methods leverage tailored models and priors to restore visibility by removing the targeted weather effects. While these approaches have shown strong performance in their respective domains, they typically lack flexibility when faced with hybrid weather conditions, making them less versatile in real-world applications.

Early approaches to haze removal, such as the Dark Channel Prior (DCP) [11] and Tarel's visibility restoration method [12], employed atmospheric scattering models and handcrafted priors for image restoration. While these methods laid the groundwork for image dehazing, they often struggled to generalize effectively across varying weather conditions. Building on this foundation, more recent dehazing networks, including AOD-Net [4] and DehazeNet [13], utilized convolutional neural networks (CNNs) to predict transmission maps, significantly improving haze removal performance. However, these approaches still encounter limitations when applied to scenes with complex weather effects.

In the realm of rain removal, several methods have emerged, incorporating multi-scale deep learning architectures to address the challenges posed by rain streak. Notable techniques such as DID-MDN [14] and JORDER [15] have made significant strides in handling rain degradation, with the latter introducing a three-step strategy and context-aware dilated convolution network to effectively remove rain streak and rain accumulation effects. More recently, approaches like RainNet [16] have used end-to-end learning models to enhance rain streak removal performance by incorporating a combination of generative adversarial networks (GANs) and CNNs, further improving results in complex weather scenarios.

For snow removal, DesnowNet [3] proposes a multi-stage deep network to address both translucent and opaque snow particles by modeling chromatic aberration and residual complement learning. Recent advancements, such as CPLFormer [17], have extended this concept by using unsupervised learning to adapt to varying snow intensities and real-world conditions, achieving impressive results in both static and dynamic snow scenes.

B. All-in-One Frameworks

All-in-one frameworks aim to address multiple weather types using a unified model, thereby overcoming the limitations inherent in task-specific approaches.

Transformer-based models have shown significant promise in this domain. For instance, TransWeather [5] employs transformer-based attention mechanisms to model long-range spatial dependencies, achieving excellent results in adaptive weather restoration. MWFormer [18] utilizes hyper-networks and feature-wise linear modulation to adaptively restore images degraded by multiple weather types, achieving enhanced performance in multi-weather image restoration. However, these transformer-based models share a common limitation: they often struggle with high computational costs and may require large datasets for optimal performance, which limits their general applicability and scalability in real-world weather restoration applications.

Generative Adversarial Network (GAN)-based models, such as BIDeN [8] and AU-GAN [19], balance adversarial and reconstruction objectives to generate perceptually superior outputs for weather degradation restoration tasks. The latter introduces an asymmetric generator to decouple weather-specific attributes from scene content, paired with an uncertainty-aware discriminator that prioritizes severely degraded regions. Despite these methods' strengths, they often face challenges related to training instability.

Attention and Feature Fusion-based Models have also made notable contributions. MPRNet [20] and NAFNet [21] introduce progressive attention and nonlinear fusion strategies to enhance feature extraction for multi-weather restoration. Despite these advancements, these models still face challenges related to balancing the complexity of multi-weather tasks with their need for accurate feature extraction, particularly in cases of severe weather degradation.

Other Approaches have also been explored to enhance weather restoration. For example, Zhu et al. [6] introduced a two-stage model that separates weather-general and weather-specific features, thereby improving model generalization. Zhou et al. [22] proposed an approach utilizing codebook priors to improve restoration accuracy across various weather conditions. These advancements further exemplify the ongoing efforts to enhance the robustness and versatility of weather restoration models.

In this paper, we aim to integrate the advantages of both task-specific and all-in-one frameworks, proposing a more powerful BIDeN for Severe Weather Removal.

III. METHOD

A. Overall Architecture

The SWR-BIDeN model is an improved variant of BIDeN and fundamentally operates as a Generative Adversarial Network. As illustrated in Fig.2, its structure is broadly divided into two components: a generator G and a discriminator D. The generator G consists of an encoder and a multi-head module. The encoder employs a three-branch design, with each branch responsible for extracting features from the input image at different scales. This multi-scale feature extraction mechanism ensures the model captures comprehensive image features, ranging from local details to global context. The feature maps from the three scales are concatenated along the channel dimension and fed into the CSIAM module. The CSIAM module generates a 3D attention map by capturing pairwise relationships across three dimensions, thereby refining the feature maps before passing them to the multi-head module. The multi-head module generates various sub-components of the degraded weather image. The discriminator D comprises two branches. The first branch D_S functions as a standard discriminator, classifying whether its input is real or generated, guiding the generator to produce realistic images. The second branch D_P predicts the source components involved in a mixed image z, with a confidence threshold of zero. This architecture, through its multi-scale encoder design, attention mechanism, and dual-branch discriminator, ensures robust performance in severe weather image restoration tasks.

B. Proposed BSConv-Attention Residual Block

In the encoder section, the original model utilizes three branches to extract features from input images at different scales, ensuring comprehensive capture of image characteristics ranging from local details to global information. As illustrated in Fig. 3, each branch of the encoder employs nine consecutive traditional residual blocks. This approach not only significantly increases the number of parameters and computational cost but also restricts the model's potential for lightweight development. To mitigate these issues, this paper introduces a lightweight residual module BARB to replace the traditional residual blocks in the original encoder. The structure of the BARB module is depicted in the lower part of Fig. 3.

The BARB module effectively reduces computational complexity and parameter count by incorporating BSConv in place of conventional convolution operations. The core concept of BSConv is to approximate the weights of convolution kernels by sharing a blueprint. Specifically, BSConv represents convolution kernels as a linear combination of a set of shared base filters (blueprints). This means that convolution kernels for different output channels share the same spatial structure but have distinct combination weights in the channel dimension. Concretely, each filter kernel $F^{(n)}$ can be expressed as a combination of a blueprint $B^{(n)}$ and the weights $w_{n,1}, \ldots, w_{n,M} \in \mathbb{R}$ via

$$F^{(n)}_{m,;,\cdot} = w_{n,m} \cdot B^{(n)} \tag{1}$$

with $m \in \{1, \ldots, M\}$ and $n \in \{1, \ldots, N\}$. Compared to standard convolution, BSConv not only reduces the computational burden but also significantly decreases the number of trainable parameters. While standard convolution requires $M \times N \times K^2$ trainable parameters, BSConv only necessitates $N \times K^2 + M \times N$ parameters. Furthermore, existing research has demonstrated that BSConv often outperforms standard convolution in various scenarios. The BSConv optimizes the model's parameter count and computational efficiency, while maintaining robust feature extraction capabilities and enhancing training speed. This enables the BARB module to achieve a lightweight architecture without compromising performance.

To further enhance the model's ability to process image details, the BARB module integrates the Enhanced Spatial Attention (ESA) [23] and the Contrast-aware Channel Attention (CCA) [24] mechanisms. The structure of ESA is shown in Fig. 4(a). Initially, a 1×1 convolution is applied to reduce the number of input feature channels, followed by stride convolution and stride max pooling to decrease spatial dimensions. Subsequently, a series of BSConv layers replace standard convolution to improve the efficiency and effectiveness of feature extraction. Afterward, a 1×1 convolution restores the original number of channels. Finally, a Sigmoid activation function generates an attention matrix to perform weighted adjustments on the input features. Through these steps, the ESA module effectively focuses on important spatial regions within the image, enhancing feature representation and overall model performance.

Following the ESA module, the CCA module is incorporated, as illustrated in Fig. 4(b). The CCA module is a channel attention mechanism specifically designed for low-level image processing. Unlike traditional channel attention modules, CCA replaces the global pooling operation with the sum of the standard deviation and the mean. This modification aids in enhancing image details and texture structure information, thereby improving the model's adaptability to complex image scenes. By integrating both the ESA and CCA modules, BARB not only attends to key regions in the spatial dimension but also strengthens the representation of important features in the channel dimension. This dual attention mechanism significantly enhances the model's ability to capture both texture and detailed information.

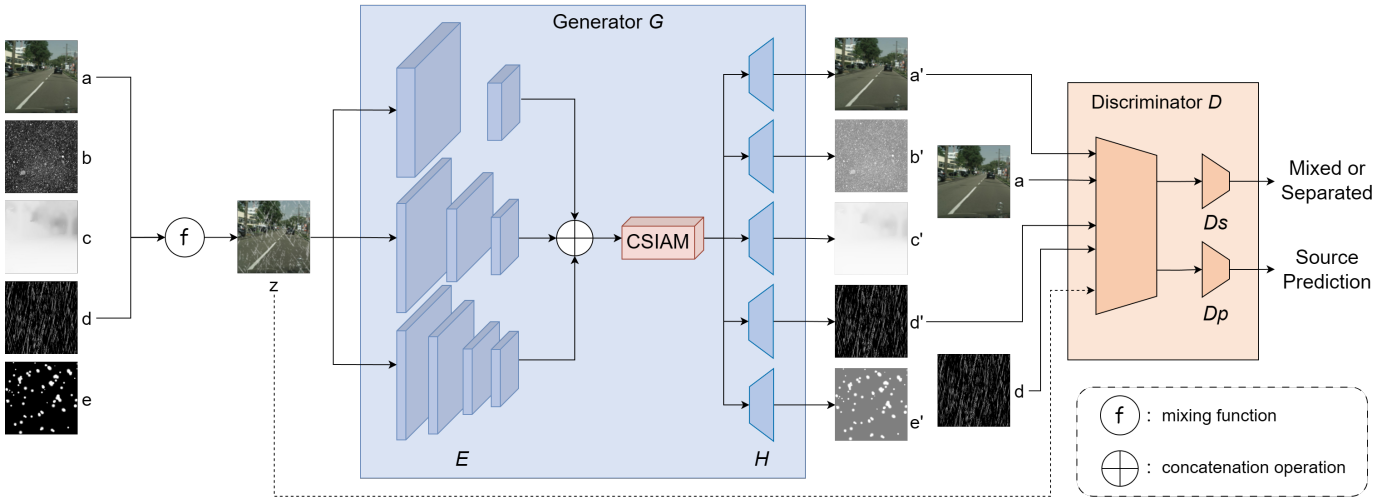

Fig. 2. Diagram of the SWR-BIDeN architecture

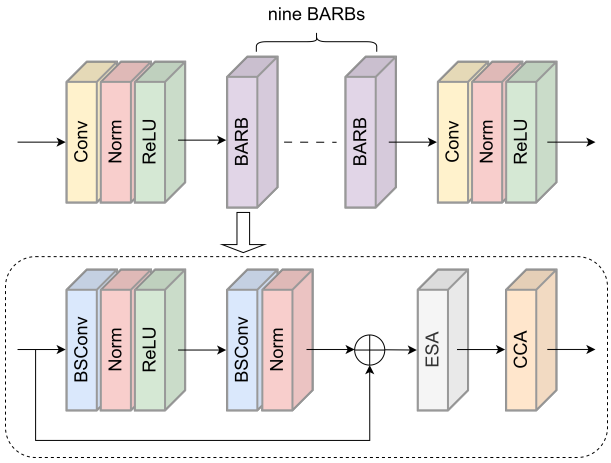

Fig. 3. One branch of the encoder. The bottom section of the diagram illustrates the BARB architecture.

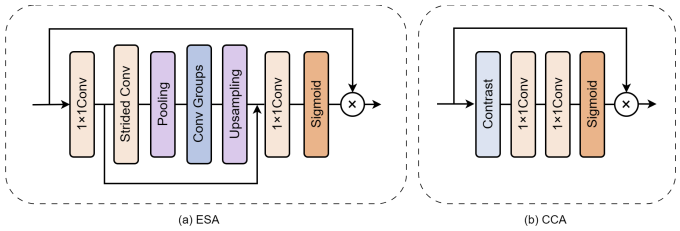

Fig. 4. The Structural Diagram of ESA and CCA

Although traditional residual modules are effective in extracting local features, their high computational complexity, parameter redundancy, and limited capacity for capturing global features result in suboptimal performance in complex scenarios. In contrast, the BARB module substantially reduces computational burden and parameter count by introducing BSConv and attention mechanisms, while simultaneously enhancing feature extraction efficiency and effectiveness. Specifically, the introduction of BSConv not only decreases the model's computational and parameter requirements but also enhances the expressiveness of convolution kernels through shared templates. The combination of the ESA and the CCA modules optimizes feature representation in both spatial and channel dimensions, thereby improving the model's contextual understanding and robustness. These enhancements enable the BARB module to significantly boost the overall performance of image restoration tasks while maintaining efficient feature extraction.

In summary, the proposed BARB module successfully replaces traditional residual blocks through a lightweight design and the integration of attention mechanisms, achieving dual improvements in both model performance and efficiency. The incorporation of this module equips the enhanced BIDeN model with greater competitiveness and application potential in image restoration tasks.

C. Proposed Channel-Spatial Integration Attention Module

The multi-branch encoder in the generator effectively captures the multi-scale information of the image, thereby enhancing both the detail and overall consistency of the generated outputs. However, despite its strong performance in multi-scale feature extraction, the model struggles to capturing the complex relationships among the channel, height, and width dimensions. This limitation is primarily reflected in feature fusion and information interaction. The channel, height, and width dimensions each play distinct roles in image features, and neglecting the intricate interplay among these dimensions can lead to inadequate feature fusion. Consequently, this compromises the quality of the generated images, limiting their fidelity and fine-grained detail.

To address the aforementioned issues, this paper proposes a novel attention module CSIAM to further refine the feature maps obtained from the encoder. The structure of CSIAM is illustrated in Fig. 5. CSIAM is designed to enhance the richness and expressive power of feature representations by

deeply capturing the pairwise relationships among the channel, height, and width dimensions. Furthermore, CSIAM replaces conventional pooling operations with the sum of standard deviation and mean, enabling a more holistic capture of the statistical properties of feature maps. This approach preserves both the central tendencies and variability of feature distributions while enhancing the model's robustness to local variations and dynamic ranges. As a result, CSIAM effectively balances global information and local details when handling scenes with complex textures or lighting variations, thereby significantly improving the capability of feature representation.

CSIAM adopts a three-branch structure corresponding to the channel, height, and width dimensions. Specifically, the workflow of CSIAM involves the following steps:

Upon receiving the input feature map $F \in \mathbb{R}^{C\times H\times W}$, the CSIAM module first applies a 1×1 convolution to reduce the number of channels, resulting in $X \in \mathbb{R}^{C/r\times H\times W}$, where r is the reduction ratio. X is then passed through three branches.

The first branch is designed to learn the relationships between the height and width dimensions. It encodes X along the channel dimension by computing the sum of the standard deviation and the mean, ultimately producing $Y_{h,w} \in \mathbb{R}^{1\times H\times W}$. Specifically:

$$Y_{h,w} = \frac{1}{C/r}\sum_{c=1}^{C/r} X_{c,h,w} + \sqrt{\frac{1}{C/r}\sum_{c=1}^{C/r}\left(X_{c,h,w} - \frac{1}{C/r}\sum_{c=1}^{C/r} X_{c,h,w}\right)^2} \quad (2)$$

Subsequently, a 7×7 convolution kernel is applied to $Y_{h,w}$, generating a tensor of shape $1\times H\times W$. This tensor is then expanded to restore its original dimensions, $C/r\times H\times W$, resulting in Y_1, i.e.,

$$Y_1 = \text{Expand}\left(\text{Conv}_{7\times 7}(Y_{h,w})\right) \in \mathbb{R}^{C/r\times H\times W} \quad (3)$$

The second branch is designed to capture the relationships between the height and channel dimensions. First, X is reshaped into $X_2 \in \mathbb{R}^{W\times H\times C/r}$. Encoding is then performed along the width dimension by computing the sum of the standard deviation and the mean, resulting in $Y_{h,c} \in \mathbb{R}^{1\times H\times C/r}$. Subsequently, a 7×7 convolution kernel is applied, generating a tensor of $1\times H\times C/r$. This tensor is then restored to its original dimensions $C/r\times H\times W$ using reshape and expand operations, producing Y_2.

The third branch is designed to capture the relationships between the channel and width dimensions. First, X is reshaped into $X_3 \in \mathbb{R}^{H\times C/r\times W}$. Encoding is then performed along the width dimension by calculating the sum of the standard deviation and the mean, resulting in $Y_{c,w} \in \mathbb{R}^{1\times C/r\times W}$. A 7×7 convolution is subsequently applied, producing a tensor of shape $1\times C/r\times W$. This tensor is then reshaped and expanded back to $C/r\times H\times W$, forming Y_3.

The three tensors Y_1, Y_2, and Y_3 are fused through element-wise arithmetic averaging. Subsequently, a 1×1 convolution is applied to restore the channel count, followed by activation using a sigmoid function to produce the final attention matrix $\alpha \in \mathbb{R}^{C\times H\times W}$.

$$Y_{\text{fused}} = \frac{Y_1 + Y_2 + Y_3}{3} \in \mathbb{R}^{C/r\times H\times W} \quad (4)$$

$$\alpha = \sigma(\text{Conv}_{1\times 1}(Y_{\text{fused}})) \in \mathbb{R}^{C\times H\times W} \quad (5)$$

where σ means Sigmoid activation function.

The attention matrix α is multiplied element-wise with the input tensor, yielding the refined refined feature map.

By integrating the CSIAM module between the encoder and the multi-head, the model effectively focuses on critical regions within the feature maps, thereby enhancing the performance of subsequent image generation tasks.

D. Loss Function

To further improve the perceptual quality of the generated images, we introduce LPIPS loss into the perceptual loss to improve visual detail reconstruction while preserving semantic consistency. This improvement focuses on strengthening the perception of local image features through learnable weights, thereby increasing the perceptual similarity between generated and real images, particularly in recovering intricate details in complex scenes.

In the original BIDeN model, the perceptual loss used a pre-trained VGG network as a feature extractor and computes the L1 norm difference between the feature maps of the generated and real images across multiple VGG network layers. Specifically, the VGG loss attempts to maintain the consistency of high-level semantic features in the generated image by comparing feature maps at different layers of the network, thereby improving the visual effect. However, one limitation of the VGG loss is that it relies on fixed pre-trained network parameters and does not consider the specific perceptual quality requirements of the image generation task. This means that the VGG loss might not be able to adaptively optimize the image details, especially in complex image generation tasks, where it may struggle to fully capture the differences in low-level details between the generated image and the target image, thus limiting its performance in detail recovery and visual quality improvement.

To address these shortcomings, we introduced the LPIPS loss. Unlike the VGG loss, the LPIPS loss employs a pre-trained AlexNet to compute perceptual differences between images and incorporates learnable weights w_l, which adaptively modulate the contributions of different feature channels. This mechanism enables LPIPS to better align feature channel contributions with human visual perception, dynamically emphasizing features critical for perceptual similarity. Consequently, LPIPS provides a more flexible and task-specific approach to enhancing image generation quality, particularly excelling in detail recovery and the refinement of complex image features.

The improved perceptual loss is expressed as:

$$\mathcal{L}_{\text{perceptual}}(G) = \lambda_{\text{VGG}}\mathcal{L}_{\text{VGG}}(G) + \lambda_{\text{LPIPS}}\mathcal{L}_{\text{LPIPS}}(G) \quad (6)$$

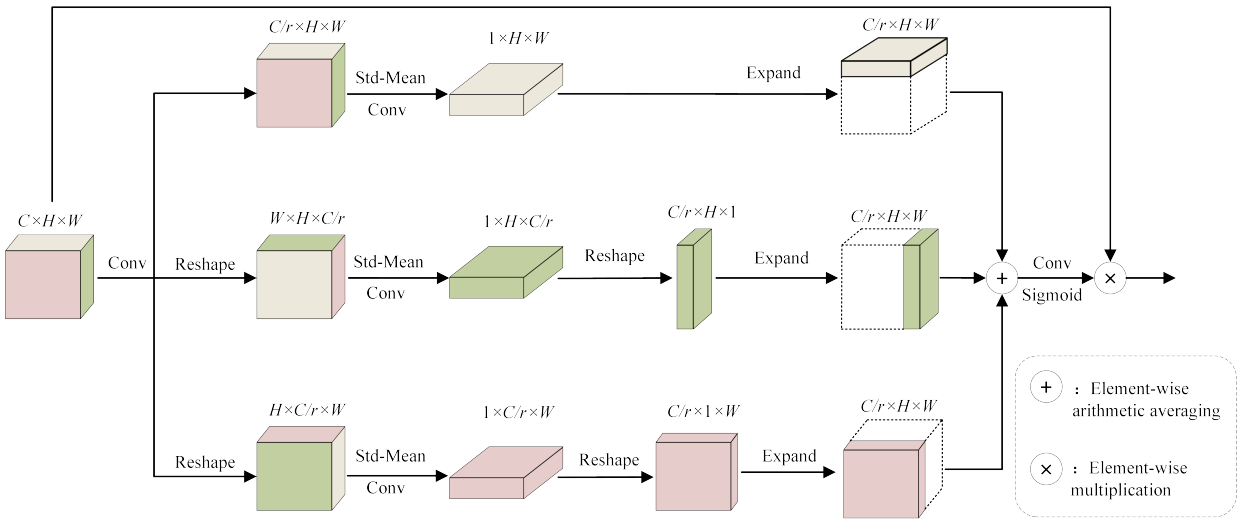

Fig. 5. Graphical representation of CSIAM

$$\mathcal{L}_{\text{VGG}}(G) = \mathbb{E}_{x,z}\left[\sum_{l} \lambda_l \|\Phi_l(x) - \Phi_l(G(z))\|_1\right] \quad (7)$$

$$\mathcal{L}_{\text{LPIPS}}(G) = \mathbb{E}_{x,z}\left[\sum_{l} \|w_l \odot (\Psi_l(x) - \Psi_l(G(z)))\|_2^2\right] \quad (8)$$

In which, Φ_l and Ψ_l represent the feature extractors at layer l for VGG and LPIPS, respectively. While w_l denotes the weight parameters at layer l. The terms λ_{VGG} and λ_{LPIPS} are the weight coefficients for the loss functions.

Our final objective function is:

$$\mathcal{L}(G, D_S, D_P) = \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}}(G, D_S) + \lambda_{\text{perceptual}}\mathcal{L}_{\text{perceptual}}(G) + \lambda_{\text{L}}\mathcal{L}_{\text{L}}(G) + \lambda_{\text{BCE}}\mathcal{L}_{\text{BCE}}(D_P). \quad (9)$$

where G denotes the generator, D_S represents the standard discriminator, D_P is the source component predictor, $\mathcal{L}_{\text{GAN}}$ stands for the adversarial loss, $\mathcal{L}_{\text{L}}$ denotes the L1/L2 loss, and $\mathcal{L}_{\text{BCE}}$ corresponds to the binary cross-entropy loss. $\lambda_{\text{GAN}}, \lambda_{\text{perceptual}}, \lambda_{\text{L}}, \lambda_{\text{BCE}}$ are their respective weighting coefficients.

The incorporation of LPIPS loss enhances both detail recovery and the model's adaptability in image generation tasks. By using learnable channel weights, LPIPS more effectively captures local differences, improving detail preservation and visual consistency, particularly in challenging scenarios such as severe weather removal with high noise or low contrast.

IV. EXPERIMENTS

In this section, we compare the SWR-BIDeN model with the current state-of-the-art models and conduct ablation experiments to demonstrate the effectiveness of each improvement.

A. Datasets

The experiments in this paper were conducted on both synthetic and real-world datasets. The synthetic dataset [8] was selected to align with the one used by the baseline model and was constructed based on the CityScapes dataset. Specifically, the Cityscapes test set was repurposed as the training set (2975) for the synthetic dataset, while its validation set was allocated as the test set (500) of the synthetic dataset. The test set for all source components consists of a fixed number of 500 images. Three different types of masks were used in the experiments: rain streak (1620), raindrop (3500), and snow (3500), covering different intensities of weather conditions. Additionally, haze conditions were simulated using transmission maps(2975×3) of three different intensities, obtained from the Foggy CityScape dataset. For the real-world dataset, the WeatherStream dataset was chosen, which includes three weather conditions: rain, snow, and haze. It contains 176100 training images and 11400 test images.

B. Implementation Details

The experiments in this paper were conducted on a computer equipped with an NVIDIA A40 GPU (48GB GPU memory). The operating system used was Ubuntu 20.04, and the deep learning framework selected was PyTorch 2.4.1. The programming language used was Python 3.8, and the CUDA version was 12.1. In all experiments, the Adam optimizer was used, with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ for both G and D. The initial learning rate for the SWR-BIDeN model was set to 0.0003, with the number of training epochs set to 200, and linear decay started after 50% of the total iterations. During training, the batch size was set to 16, and instance normalization was applied. All training images were cropped into 256×256 image patches, with random horizontal flipping applied for data augmentation. During testing, we loaded test images with a resolution of 256×256.

C. Evaluation Metrics

We used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as evaluation metrics (for the RGB channels), with higher values indicating better performance.

D. Comparison With the State-of-the-Art Models

To evaluate the effectiveness of SWR-BIDeN, we conducted a comprehensive comparison with state-of-the-art image restoration models on both synthetic datasets and the real-world WeatherStream dataset. The comparison models included general image restoration approaches, namely MPRNet [20], NAFNet [21], Uformer [25], Restormer [26], and GRL [27], as well as all-in-one weather restoration models, specifically AirNet [28], TUM [29], Transweather [5], BIDeN [8], WGWS [6], and LDR [7].

The comparison experimental results on the WeatherStream dataset are shown in TABLE I. The experimental results indicate that SWR-BIDeN achieved the best performance under most weather conditions. For example, in the Rain category, the PSNR and SSIM of SWR-BIDeN reached 24.57 and 0.821, outperforming the second-best LDR method, which had PSNR and SSIM values of 24.42 and 0.818. Similarly, in the Snow category, the PSNR and SSIM of SWR-BIDeN were 23.69 and 0.824, also outperforming all other comparison models. In terms of overall performance, the average PSNR and SSIM of SWR-BIDeN reached 23.69 and 0.824, achieving the best results among all methods. The qualitative comparison results on the WeatherStream dataset are shown in Fig. 6.

TABLE I
QUANTITATIVE COMPARISON ON THE WEATHERSTREAM DATASET

Type	Method	Rain		Haze		Snow		Average	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
General	MPRNet [20]	21.50	0.791	21.73	0.763	20.74	0.801	21.32	0.785
	NAFNet [21]	23.01	0.803	22.20	0.803	22.11	0.826	22.44	0.811
	Uformer [25]	22.25	0.791	18.81	0.763	20.94	0.801	20.67	0.785
	Restormer [26]	23.67	0.804	22.90	0.803	22.51	0.828	22.86	0.812
	GRL [27]	23.75	0.805	22.88	0.802	22.59	0.829	23.07	0.812
All-in-One	AirNet [28]	22.52	0.797	21.56	0.770	21.44	0.812	21.84	0.793
	TUM [29]	23.22	0.795	22.38	0.805	22.25	0.827	22.62	0.809
	Transweather [5]	22.21	0.772	22.55	0.774	21.79	0.792	22.18	0.779
	BIDeN [8]	23.29	0.781	22.72	0.785	22.91	0.814	22.97	0.793
	WGWS [6]	23.80	0.807	22.78	0.800	22.72	0.831	23.10	0.813
	LDR [7]	24.42	0.818	**23.11**	**0.809**	23.12	0.838	23.55	0.822
	Ours	**24.57**	**0.821**	22.95	0.798	**23.54**	**0.853**	**23.69**	**0.824**

However, in the haze category, the performance of SWR-BIDeN was inferior to some comparative models. Specifically, the LDR model achieved a PSNR of 23.11 and an SSIM of 0.809 under hazy conditions, which slightly outperformed SWR-BIDeN's PSNR of 23.54 and SSIM of 0.853. This observation suggests that, although SWR-BIDeN demonstrates superior overall performance, its effectiveness in haze removal remains suboptimal. In contrast, the LDR method exhibited enhanced robustness in handling haze removal tasks.

The experimental results on the synthetic dataset are presented in TABLE II. The experiment simulated six complex weather conditions, with Case 1-6 representing: (1) rain streak, (2) rain streak + snow, (3) rain streak + light haze, (4) rain streak + heavy haze, (5) rain streak + moderate haze + raindrop, and (6) rain streak + snow + moderate haze + raindrop. SWR-BIDeN demonstrated superior performance in both PSNR and SSIM metrics. For instance, in Case 1 (rain streak only), SWR-BIDeN achieved a PSNR of 35.30 and an SSIM of 0.957, surpassing the second-best method, LDR, which attained a PSNR of 35.05 and an SSIM of 0.955. Even in the most complex scenario (Case 6), SWR-BIDeN maintained its lead with a PSNR of 22.80 and an SSIM of 0.832, outperforming all comparative models. These results underscore SWR-BIDeN's robust restoration capabilities, particularly in addressing complex scenes characterized by multiple overlapping weather conditions.

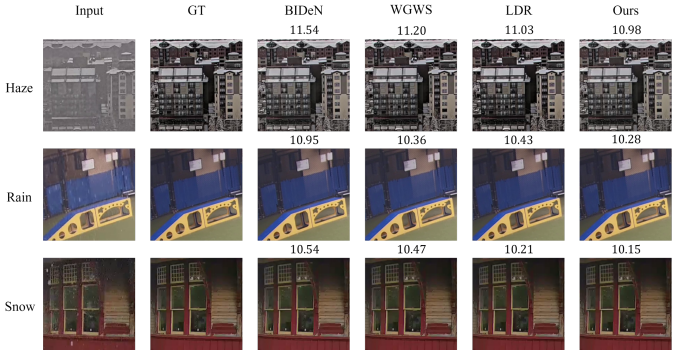

Fig. 6. Qualitative comparison on the WeatherStream dataset. The top four performing models demonstrated excellent restoration effectiveness under haze, rain, and snow conditions. The numbers above the images represent the Root Mean Square Error (RMSE), where smaller values indicate better performance.

In summary, SWR-BIDeN demonstrates superior performance across a wide range of weather conditions, achieving state-of-the-art results in both synthetic and real-world datasets. It excels particularly in handling complex scenes with multiple overlapping weather conditions, as evidenced by its leading PSNR and SSIM metrics in most cases. While its performance in haze removal is slightly inferior to some comparative models, SWR-BIDeN's overall robustness and restoration capabilities make it a highly effective solution for comprehensive severe weather removal tasks.

E. Ablation Experiment

To validate the effectiveness of the proposed modules and loss function designs, we conducted ablation experiments under the weather condition of Case 6 (rain streak, snow, moderate haze, and raindrop) in the synthetic dataset. The contribution of each component to the model's performance was systematically evaluated. The experimental results are presented in TABLE III.

First, the baseline model BIDeN exhibited a slight improvement in performance after integrating the BARB module. Specifically, the PSNR for CityScape image restoration increased from 26.44 to 26.56, and the SSIM improved from 0.87 to 0.872. Furthermore, the model demonstrated enhanced capability in restoring weather mask images, with the PSNR and SSIM of the generated rain streak image increasing from 28.31 and 0.823 to 28.98 and 0.843, respectively. These results indicate that the BARB module, by enhancing feature extraction capabilities, significantly improved the image

TABLE II
QUANTITATIVE COMPARISON ON THE SYNTHETIC DATASET

Type	Method	Case1		Case2		Case3		Case4		Case5		Case6	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
General	MPRNet [20]	33.39	0.945	30.52	0.909	23.98	0.900	18.54	0.829	21.18	0.846	20.76	0.812
	NAFNet [21]	34.12	0.947	31.25	0.912	24.20	0.904	19.05	0.836	22.45	0.854	21.85	0.818
	Uformer [25]	33.89	0.943	30.75	0.910	24.05	0.902	18.75	0.832	22.30	0.852	21.70	0.816
	Restormer [26]	34.29	0.951	30.60	0.917	23.74	0.905	20.33	0.853	22.17	0.859	21.24	0.821
	GRL [27]	34.35	0.949	30.95	0.916	24.50	0.907	20.40	0.856	22.80	0.861	22.05	0.823
All-in-One	AirNet [28]	33.45	0.944	30.80	0.911	23.90	0.901	19.15	0.837	22.15	0.857	21.55	0.819
	TUM [29]	34.05	0.948	31.10	0.914	24.10	0.903	19.45	0.841	22.55	0.860	21.95	0.822
	Transweather [5]	33.80	0.947	30.95	0.913	23.85	0.899	19.05	0.835	22.00	0.856	21.40	0.817
	BIDeN [8]	34.55	0.950	31.20	0.915	24.35	0.905	20.50	0.858	22.85	0.864	22.15	0.825
	WGWS [6]	34.70	0.953	31.50	0.917	24.65	0.908	20.65	0.861	23.10	0.867	22.30	0.828
	LDR [7]	35.05	0.955	31.85	0.919	24.85	0.910	20.95	0.864	23.35	0.870	22.55	0.830
	Ours	**35.30**	**0.957**	**32.15**	**0.921**	**25.05**	**0.912**	**21.15**	**0.866**	**23.65**	**0.873**	**22.80**	**0.832**

TABLE III
ABLATION EXPERIMENT RESULTS

Method	CityScape		Rain Streak		Snow		Haze		Raindrop		Params(M)	FLOPs(G)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
BIDeN	26.44	0.870	28.31	0.823	24.79	0.658	**29.83**	0.948	21.47	0.893	41.8	446.8
BIDeN+BARB	26.56	0.872	28.98	0.843	25.68	0.696	28.85	0.942	22.12	0.912	**29.3**	**346.4**
BIDeN+BARB+CSIAM	27.93	0.881	29.17	0.837	26.18	0.71	28.98	0.943	**23.54**	**0.915**	29.4	346.6
BIDeN+BARB+CSIAM+LPIPS LOSS	**28.37**	**0.896**	**29.65**	**0.867**	**27.54**	**0.754**	29.45	**0.954**	23.47	0.912	29.4	377.3

restoration performance. Additionally, the model's parameter count decreased from 41.8M to 29.3M, and the FLOPs were reduced from 446.8G to 346.4G. This demonstrates that the introduction of the BARB module not only improved performance but also reduced computational overhead.

Building on this, the further addition of the CSIAM module led to additional performance improvements. Specifically, the PSNR and SSIM for the generated CityScape image increased from 26.56 and 0.872 to 27.93 and 0.881, respectively. Similarly, the PSNR for the raindrop mask image improved from 22.12 to 23.54, and the SSIM increased from 0.912 to 0.915. These results demonstrate that the CSIAM module effectively enhanced the model's ability to perceive and restore complex weather degradation features by integrating channel and spatial attention mechanisms.

Finally, the incorporation of the LPIPS Loss into the perceptual loss function achieved optimal overall performance. Specifically, the PSNR for the CityScape image increased to 28.37, and the SSIM improved to 0.896. Additionally, the generation of rain streak and snow mask images reached their peak performance levels. These results indicate that the LPIPS Loss not only effectively optimizes the perceptual quality of the restored images but also enhances their visual consistency.

However, the proposed improvements introduced some instability in the model's ability to restore haze mask images, resulting in a slight decrease in PSNR and a marginal increase in SSIM for the haze mask. Despite this, each enhancement contributes meaningfully to the overall performance of the severe weather removal task.

It is worth noting that the introduction of the BARB module significantly reduced both the parameter count and computational complexity. Specifically, the final model's parameter count decreased by 12.4M, and the computational complexity was reduced by 69.5G. These results demonstrate that SWR-BIDeN not only offers performance advantages but also achieves superior resource efficiency.

V. CONCLUSION

This paper proposes SWR-BIDeN, an efficient architecture specifically designed for image restoration under adverse weather conditions, which achieves significant improvements in restoration performance and efficiency through several key innovations. Firstly, we incorporate a lightweight residual module (BARB) that enhances feature extraction capability while reducing parameters and computational complexity through BSConv implementation. Secondly, we develop a Channel-Spatial Integration Attention Module (CSIAM) that strengthens the model's contextual comprehension by effectively integrating channel and spatial information, enabling robust handling of complex degradation patterns. Thirdly, we introduce an enhanced perceptual loss function combining LPIPS and VGG losses to optimize both structural fidelity and perceptual quality of reconstructed images, ensuring restoration results align with human visual perception standards. Experimental results demonstrate that SWR-BIDeN outperforms existing models in restoration performance while maintaining fewer parameters. Future research will focus on enhancing the model's practical effectiveness in real-world scenarios.

REFERENCES

[1] J. Su, B. Xu, and H. Yin, "A survey of deep learning approaches to image restoration," *Neurocomputing*, vol. 487, pp. 46–65, 2022. I

[2] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE transactions on circuits and systems for video technology*, vol. 30, no. 11, pp. 3943–3956, 2019. I

[3] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "Desnownet: Context-aware deep network for snow removal," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3064–3073, 2018. I, II-A

[4] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778. I, II-A

[5] J. M. J. Valanarasu, R. Yasarla, and V. M. Patel, "Transweather: Transformer-based restoration of images degraded by adverse weather conditions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2353–2363. I, II-B, IV-D, I, II

[6] Y. Zhu, T. Wang, X. Fu, X. Yang, X. Guo, J. Dai, Y. Qiao, and X. Hu, "Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 21 747–21 758. I, II-B, IV-D, I, II

[7] H. Yang, L. Pan, Y. Yang, and W. Liang, "Language-driven all-in-one adverse weather removal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 24 902–24 912. I, IV-D, I, II

[8] J. Han, W. Li, P. Fang, C. Sun, J. Hong, M. A. Armin, L. Petersson, and H. Li, "Blind image decomposition," in *European Conference on Computer Vision*. Springer, 2022, pp. 218–237. I, II-B, IV-A, IV-D, I, II

[9] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595. I

[10] D. Haase and M. Amthor, "Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14 600–14 609. I

[11] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010. II-A

[12] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, pp. 2201–2208. II-A

[13] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 154–169. II-A

[14] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 695–704. II-A

[15] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Joint rain detection and removal via iterative region dependent multi-task learning," *CoRR, abs/1609.07769*, vol. 2, no. 3, pp. 1–12, 2016. II-A

[16] X. Chen, K. Feng, N. Liu, B. Ni, Y. Lu, Z. Tong, and Z. Liu, "Rainnet: A large-scale imagery dataset and benchmark for spatial precipitation downscaling," *Advances in Neural Information Processing Systems*, vol. 35, pp. 9797–9812, 2022. II-A

[17] S. Chen, T. Ye, Y. Liu, J. Bai, H. Chen, Y. Lin, J. Shi, and E. Chen, "Cplformer: Cross-scale prototype learning transformer for image snow removal," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 4228–4239. II-A

[18] R. Zhu, Z. Tu, J. Liu, A. C. Bovik, and Y. Fan, "Mwformer: Multi-weather image restoration using degradation-aware transformers," *IEEE Transactions on Image Processing*, 2024. II-B

[19] J.-g. Kwak, Y. Jin, Y. Li, D. Yoon, D. Kim, and H. Ko, "Adverse weather image translation with asymmetric and uncertainty-aware gan," *arXiv preprint arXiv:2112.04283*, 2021. II-B

[20] A. Mehri, P. B. Ardakani, and A. D. Sappa, "Mprnet: Multi-path residual network for lightweight image super resolution," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 2704–2713. II-B, IV-D, I, II

[21] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European conference on computer vision*. Springer, 2022, pp. 17–33. II-B, IV-D, I, II

[22] T. Ye, S. Chen, J. Bai, J. Shi, C. Xue, J. Jiang, J. Yin, E. Chen, and Y. Liu, "Adverse weather removal with codebook priors," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 653–12 664. II-B

[23] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2359–2368. III-B

[24] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proceedings of the 27th acm international conference on multimedia*, 2019, pp. 2024–2032. III-B

[25] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 683–17 693. IV-D, I, II

[26] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739. IV-D, I, II

[27] Y. Li, Y. Fan, X. Xiang, D. Demandolx, R. Ranjan, R. Timofte, and L. Van Gool, "Efficient and explicit modelling of image hierarchies for image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 278–18 289. IV-D, I, II

[28] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 452–17 462. IV-D, I, II

[29] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 653–17 662. IV-D, I, II