Adverse Weather Removal with Codebook Priors

Tian Ye[1,3*] Sixiang Chen[1,3*] Jinbin Bai[2 *] Jun Shi[4] Chenghao Xue[3]
Jingxia Jiang[3] Junjie Yin[3] Erkang Chen[3] Yun Liu[5†]

[1]The Hong Kong University of Science and Technology (Guangzhou) [2]National University of Singapore
[3]School of Ocean Information Engineering, Jimei University
[4]Xinjiang University [5]College of Artificial Intelligence, Southwest University
{*owentianye*, *sixiangchen*} @hkust-gz.edu.cn jinbin.bai@u.nus.edu, junshi2022@gmail.com,
{202021115097, 202021114006, 202021112006, *ekchen*} @jmu.edu.cn, yunliu@swu.edu.cn

Abstract

Despite recent advancements in unified adverse weather removal methods, there remains a significant challenge of achieving realistic fine-grained texture and reliable background reconstruction to mitigate serious distortions.

Inspired by recent advancements in codebook and vector quantization (VQ) techniques, we present a novel Adverse Weather Removal network with Codebook Priors (AWRCP) to address the problem of unified adverse weather removal. AWRCP leverages high-quality codebook priors derived from undistorted images to recover vivid texture details and faithful background structures. However, simply utilizing high-quality features from the codebook does not guarantee good results in terms of fine-grained details and structural fidelity. Therefore, we develop a deformable cross-attention with sparse sampling mechanism for flexible perform feature interaction between degraded features and high-quality features from codebook priors. In order to effectively incorporate high-quality texture features while maintaining the realism of the details generated by codebook priors, we propose a hierarchical texture warping head that gradually fuses hierarchical codebook prior features into high-resolution features at final restoring stage.

With the utilization of the VQ codebook as a feature dictionary of high quality and the proposed designs, AWRCP can largely improve the restored quality of texture details, achieving the state-of-the-art performance across multiple adverse weather removal benchmark.

1. Introduction

The restoration of images under adverse weather conditions, such as heavy haze or rain, is a major topic of research in the field of computer vision. The original images may suffer from severe weather-induced distortions, like intense rain streaks or dense hazing effects, which obscure the true background and deteriorate the performance of high-level vision tasks. Therefore, a difficult inverse problem arises, whereby degraded images are likely to experience significant losses of detail and structure, requiring restoration.

With the advent of deep learning techniques, adverse weather removal methods [33, 49, 41, 10] has achieved remarkable progress. An increasing number of studies are focusing on achieving all-in-one adverse weather removal in one go as a primary objective. This entails the elimination of all weather-related degradation through the utilization of a single, unified model. Classical adverse weather removal methods employ neural architecture search to find an optimal network design [33], use advanced decoders with learnable weather queries to decode clean features in solving this task [49], explore diffusion model for adverse weather removal [41]. However, these methods are still limited in their performance due to *their inability to robustly capture high-quality clear background features from seriously degraded images*. The only feature source that them can employed is only from degraded images, which obviously is a huge drawback in structure rebuilding and realist texture restoring by these models. Additionally, it should be noted that *the irreversible nature of severe texture loss poses a significant challenge for these methods*. Previous methods almost all pay more attention on reconstruct clean features in multi-scale feature stage, ignoring the benefits of restoring fine-grained details in the final high-resolution feature level. In conclusion, recent methods for mitigating adverse weather conditions have demonstrated impressive perfor-

*Equal contributions.
†Yun Liu is the corresponding author.

mance across various types of weather-induced degradation. However, there is still a substantial room for improvement that requires further exploration.

When presented with an image that is degraded due to unfavorable weather conditions, the task of restoring it to a pristine state becomes an incredibly difficult problem because of the various sources of degradation present. Consequently, it can be difficult to find a reliable statistical prior that can address the issue effectively. For deep networks, these challenges are magnified, as they strive to capture ideal, noise-free features as latent variables in a network that only possesses a lone encoder.

The aforementioned deliberations and inferences impel us to contemplate deploying *codebook priors* as a potential solution to the current challenge.Training a VQGAN [12] on high-quality, noise-free images has the potential to produce a Vector-Quantized (VQ) codebook that possesses high-quality feature priors. A well-trained VQ codebook holds the potential to offer a comprehensive high-quality feature set, aiding in the handling of various types of weather-induced degradation. However, due to the feature misalignment between degraded features from degraded images and high-quality features from codebook priors, the straightforward fusion of high-quality features from the vector quantization (VQ) codebook may not yield adequate results in the context of adverse weather removal [53]. Moreover, undesirable effects from various sorts of noise observed in the degraded features could additionally impact the quality of reconstructed superior-quality characteristics.

To address these issues, we introduce two special designs, which allow AWRCP to effectively explore robust codebook priors and keep fidelity in restored results, surpassing previous state-of-the-art methods in restoration performance. For facilitate pliable feature interaction and fusion, we propose a Parallel Decoder Design that integrates Deformable Cross-Attention. This design effectively introduces high-quality features and maintains structural consistency between rstored results and clean background. Our Deformable Cross-Attention (**DCA**) utilizes paired high and low-quality features to guided sparse sampling phase, flexibly adapt features with distinct quality. With the help of sparse sampling, DCA efficiently perform context modeling between two source of features and fuse them adaptively. Its aim is to ameliorate the issue of feature misalignment that arises between degraded and high-quality features. For the restoration of fine-grained details, Hierarchical Texture Warping Head is proposed by us to explore hierarchical high-quality features in high-resolution feature level, restoring fine-grained details step-by-step. We are the first work that successfully exploit codebook priors for adverse weather removal and achieve a state-of-the-art performance across all standard weather task benchmarks.

Our contributions can be summarized as follows:

- We propose a novel framework AWRCP for adverse weather removal using high-quality codebook priors learned by a pre-trained VQGAN. Compared with previous works, the AWRCP introduces codebook priors to formulate adverse weather removal task as a feature matching and fusion problem between degraded and high-quality features, enabling the leading performance.
- We propose a Deformabel Cross-Attention with Sparse Sampling for high/low-quality feature fusion. Such a manner bridge the misalignment between heterogeneous features, avoiding, effectively avoids drawbacks from codebook priors.
- A novel Hierarchical Texture Warping Head is designed to refine textures in high-resolution feature stage by introducing hierarchical prior features.

2. Related Works

2.1. Adverse Weather Removal

Over the past decade, researchers have shown great interest in developing algorithms to eliminate the effects of adverse weather [59, 38, 37, 7, 5, 6, 57, 29, 28] and image/video restoration [47, 58, 22, 23, 27, 30, 26, 18, 65, 19, 16, 17, 67, 20, 15, 68, 40, 60, 61]. These algorithms include single-image deraining [34, 46, 24], single-image dehazing [11, 44, 54], single-image desnowing [39, 8, 9], multiple degradation removal [62, 64], and all-in-one adverse weather removal.

2.1.1 Single Weather Removal

(i) Dehazing: Single image dehazing has made remarkable progress in the past few years. AECR-Net [54] designed a contrastive learning framework to help the network learn discriminative knowledge from negative samples. **(ii) Deraining:** For single image deraining, AttenGAN [43], leverages the combination of generative adversarial networks (GANs) and attention mechanisms to restore images degraded by raindrops. IDT [55] proposed a transformer-based approach that integrates a complementary window-based transformer and a spatial transformer, facilitating the accurate modeling of short- and long-range dependencies in rainy scenes.**(iii) Desnowing:** Compared to dehazing and deraining, single-image snow removal is a more challenging image restoration task. Desnow-Net [39] developed a two-stage framework to remove snowy degradation progressively. JSTASR [8] proposed a framework to remove haze and snow simultaneously. However, as with dehazing and deraining, these methods also face difficulties achieving satisfactory results on other types of adverse weather.

2.1.2 Multiple Degradation Removal

With the recent advancements in CNNs and ViTs, it has become possible to develop a generic network that can focus on multiple image restoration tasks. One such network is MPRNet [62], which proposes a multi-stage approach that utilizes multi-patch and multi-level strategies to progressively restore degraded images. Another network is NAFNet [3], which devises a simple nonlinear activation-free network to efficiently restore degraded images by adopting a simplified channel attention mechanism to refine features in each block. Additionally, Restormer [64] proposes a channel-wise self-attention mechanism that performs global modeling in the channel dimension, leading to excellent performance on multiple image restoration tasks.

Previous multiple degradation removal methods achieve encouraging performance in different degradation scenes with a unified architecture. However, in order to handle multiple degrading effects, the above approaches have to manually choose and load the specific pre-trained weights to match the degradation type, which suffers from limitations for real-world applications.

2.1.3 All-in-one Adverse Weather Removal:

The goal of all-in-one adverse weather removal is to create a unified network with a fixed pre-trained weight that can effectively address the issue of adverse weather. The first solution proposed is All-in-One [33], which employs AutoML (NAS) in an end-to-end approach to extract clean features from various weather-type encoders. However, due to its high number of parameters, it may not be suitable for edge applications in real-world scenarios. TransWeather [49] is the first transformer designed specifically for adverse weather removal, with a weather type decoder that can decode different features from various weather degradations. While it represents an improvement over All-in-One, there is still room for further improvement in its performance. Another approach is TKL&MR [10], which is a distillation framework that transfers knowledge from multiple teacher models to a single student model. This enables the unified model to cover multiple weather tasks with a single pre-trained weight. However, its complex distillation paradigm and use of multiple pre-trained teacher models may require a longer training time to achieve leading performance in the student model. WeatherDiffusion [41] is a novel method based on diffusion for mitigating adverse weather conditions. This technique effectively harnesses the potential of diffusion models for weather removal, and has achieved state-of-the-art results across multiple benchmarks. However, the slow inference speed severely limits its practical applications.

2.1.4 VQ Codebook Learning

The VQ autoencoder was originally introduced in the VQ-VAE [50] framework. It leverages a vector quantization codebook to address the problem of posterior collapse [14]. VQGAN further improves the visual quality of reconstructed images by introducing perceptual and adversarial loss items for better codebook learning. The well-trained codebook could help many image restoration tasks like face restoration [13] and image super-resolution [2]. FeMaSR [2] expands the technique of discrete codebook learning to facilitate blind super-resolution. VQFR [13] frame the problem of blind face restoration as a code prediction task. Inspired by the promising performance of these approaches, we are the first to leverage the high-quality codebook prior for adverse weather removal.

3. Methodology

3.1. Vector-Quantized Codebook for Priors

VQ Codebook. We first make a brief review of VQGAN and its codebook. Given a high-quality clean image patch $\boldsymbol{x} \in \mathbb{R}^{H\times W\times 3}$ is first passed through the encoder E to produce its output feature $\hat{z} = E(\boldsymbol{x}) \in \mathbb{R}^{h\times w\times n_z}$, where n_z is the dimension of latent vectors. Then the vector-quantized representation of z_q is calculated by finding the nearest neighbours of each element $\hat{z}_i \in \mathbb{R}^{n_z}$, in the codebook $\mathcal{Z} \in \mathbb{R}^{K\times n_z}$ with K discrete codes as follows:

$$z_{\mathbf{q}} = \mathbf{q}(\hat{z}) := \left(\arg\min_{z_i \in \mathcal{Z}} \|\hat{z} - \hat{z}_i\| \right) \in \mathbb{R}^{h\times w\times n_z}. \quad (1)$$

where $\mathbf{q}(\cdot)$ denotes the element-wise quantization. The Decoder G maps the quantized representation z_q back to the RGB space. The overall reconstruction mechanism can be formulated as:$\hat{x} = G(z_{\mathbf{q}}) = G(\mathbf{q}(E(x)) \approx x$.

Since the feature quantization operation of Eq. 1 is non-differentiable, we follow VQGAN and simply copy the gradients from G to E for backpropagation. For further improving the quality of codebook, we introduce perceptual loss and adversarial loss as training objective of VQGAN and its codebook.

Analysis. To gain a more comprehensive insight into the potential and limitations of using well-trained VQ codebook for adverse weather removal, we conducted several preliminary experiments and derived the following observations.

Observation I. As illustrated in Fig. 2, the well-trained VQGAN produces vivid realistic details and textures into the reconstructed image. However, due to the information loss by the vector-quantization, the object structure of reconstruction has also been distorted to certain extent. Thus,

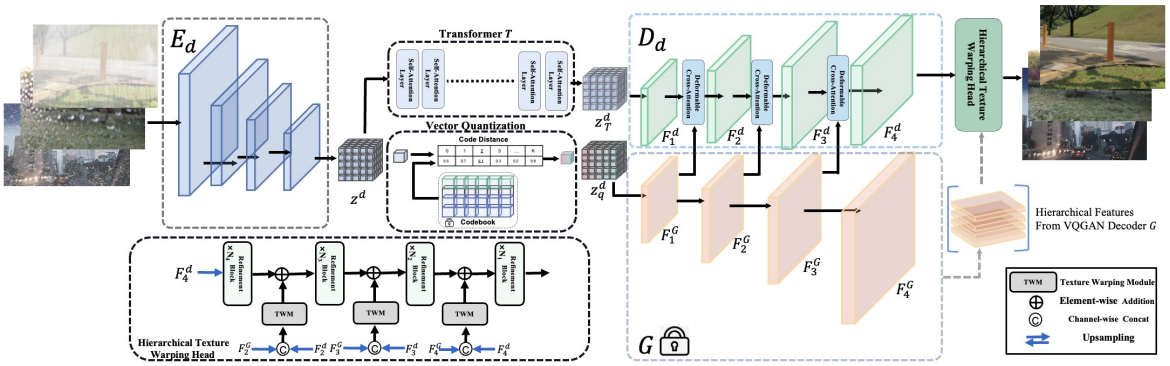

Figure 1. **Overview of AWRCP framework.** It comprises an encoder, which maps degraded images to a latent space, and a parallel decoder that is equipped with deformable cross-attention, enabling the flexible utilization of high-quality codebook priors.

Figure 2. (a) Input clean image. (b) Image reconstructed by the VQGAN that only trained on high-quality images. More realistic fine details textures have been introduced into the reconstruction results of VQGAN, but the object structure has also been distorted and warped.

simply fusing the features from the pre-trained codebook is obviously suboptimal for us. It is intuitive that the significant step in our research is to devise an appropriate solution that can effectively utilize the high-quality features from the codebook while circumventing the structural distortion of the original background. Therefore, we choose the parallel architecture as the design style of our Decoder, which could keep the structure features of original background and introduce high-quality codebook features.

Observation II. We present an investigation into the reconstruction results produced by the highly proficient VQGAN model without any fine-tuning on paired data, with a particular focus on degraded images. As depicted in Figure 3, our findings illustrate that VQGAN is able to mitigate the effects of adverse weather conditions to some degree. However, some areas of extreme degradation still exceed its ability to fully restore the patches. Moreover, VQGAN's ability to restore images trained solely on high-quality sources is limited as it encounters difficulty in properly matching the relevant codebook entry. Therefore, it is evidently inadequate to directly reuse features from the Encoder and Decoder of VQGAN. To further employ hierarchical high-quality features to boost texture restoration, we propose introducing a novel deformable cross-attention technique as a high-quality prior feature fusion mechanism, as well as a texture warping head.

Our novel modules can be constructed as a new Decoder D_d, which learns from scratch with paired data to balance the texture restoration while still preserving background structures.

3.2. Adverse Weather Removal with Codebook Priors

Based on the aforementioned observations and analyses, we formulate the problem of removing adverse weather conditions into three phases: code matching, prior feature fusion, and texture refinement. The overall framework of the proposed method is illustrated in Fig. 1. The training phase of our solution can be divided into two stages. In the first stage, we pre-train a VQGAN on high-quality data to obtain a latent discrete codebook with high-quality code entries and a well-trained Decoder G. In the second stage, AWRCP based on the pre-trained codebook and G, and learns from a widely-used *Allweather* [49, 33, 41] training dataset.

Encoder for Code Matching. We develop an efficient CNN-based encoder with robust feature extraction ability as our new degraded image encoder E_d for accurate code matching. The encoder E_d mainly consists of several efficient convolutional encoder blocks, downsampling the features to $\frac{1}{16}$ resolution step by step.

Latent Transformer for Contextual Modeling. We employ the transformer to excavate the disrupted background structure and perform effective global modeling of contextual content on the background structure. As shown in Fig.1, in the latent layer of our network, we utilize 6 self-attention layers to model global dependency relationship of

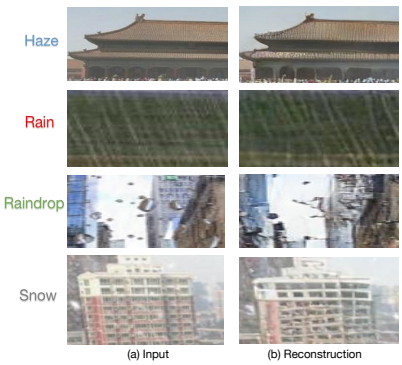

Figure 3. (a) Degraded images by adverse weather. (b) Reconstructed results by the well-trained VQGAN that only learns from high-quality images. The vector-quantized phase of features results in a loss of information, leading to the introduction of distorted structures and textures.

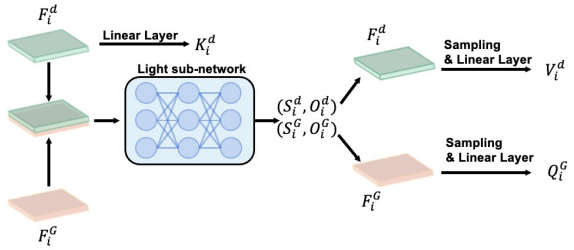

Figure 4. The figure illustrates how our DCA performs sparse sampling on two feature sets with different qualities. Our sampling mechanism does not independently predict the offsets and scales of different features. Instead, it utilizes both feature sets simultaneously to obtain four sets of parameters for conducting sparse sampling.

z^d for next step feature restoration, and z_T^d is the finally globally attended features.

Deformable Cross-Attention with Sparse Sampling. As mentioned and analyzed earlier, the reconstructed results from VQGAN reveal that high-quality features from the codebook are not flawless. Structural deformation and texture distortion lead to a more severe misalignment between high-quality and degraded features. There are face restoration methods that suggest a simple fusion of high-quality and degraded features in a pixel-wise manner. However, such an approach overlooks contextual structure, resulting in suboptimal performance. Taking into account above issues surrounding codebook prior features and feature fusion, we present a novel **Deformable Cross-Attention** (DCA) with sparse sampling, that aims to flexibly model global contextual structures and local textures with the linear complexity, and reconstruct degraded features into a better representation that is close to the clean feature by utilizing high-quality codebook priors.

Differ with vanilla cross-attention operation, our DCA aims to dynamically sample two distinct features from different sources in a sparse and global manner. By employing sparse sampling, it is possible to reshape each feature to a pre-determined resolution size (c, h, w, N_x^w, N_y^w), thereby enabling efficient global cross-attention with linear complexity. As shown in Fig. 4, the light sub-network consists of an average pooling layer, a max pooling layer, a light convolution layer with 1×1 kernel and activation function. The two distinct feature F_i^d, F_i^G are concatenated in the channel dimension, and then the maxpool and avgpool operations in sub-network are used to aggregate global information respectively, and then the two are added together to get feature for the predicting of scales and offsets.

$$S_i^d, O_i^d, S_i^G, O_i^G = \text{Sub-network} \circ \text{Concat}(F_i^d, F_i^G) \quad (2)$$

Where $S_i^{d,G}$ and $O_i^{d,G} \in \mathcal{R}^{2N}$ represent the predicted scales and offsets for sparse sampling. We leverage these scales and offsets to sample $V_i^{p^d} \in \mathcal{R}^{\frac{N}{M^2} \times C}$, $K_i^{p^d} \in \mathcal{R}^{\frac{N}{M^2} \times C}$ and $Q_i^{p^G} \in \mathcal{R}^{\frac{N}{M^2} \times C}$ from $F_i^d \in \mathcal{R}^{N \times C}$ and $F_i^G \in \mathcal{R}^{N \times C}$, respectively. Here, p represents the window size. Therefore, for each window feature $F_i^{p^d} \in \mathcal{R}^{\frac{N}{M^2} \times C}$ and $F_i^{p^G} \in \mathcal{R}^{\frac{N}{M^2} \times C}$, cross-attention can be expressed as follows:

$$\text{Attn}_i\left(Q_i^{p^G}, K_i^{p^d}, V_i^{p^d}\right) = \textbf{Softmax}\left(\frac{Q_i^{p^G} K_i^{p^d \mathrm{T}}}{\sqrt{D}} + p\right) V_i^{p^d} \quad (3)$$

Please refer to our Supplemental Materials for details and deeper analysis of our deformable cross-attention layer.

Parallel Decoder Design. Through observation and analysis we mentioned above, we design a parallel decoder architecture to gradually perform interaction between degraded features and high-quality features by deformabel cross-attention layers. Our decoder design differs from that used in face restoration method VQFR [13], as our main branch restores features from degraded sources while another branch directly inherits from the pre-trained VQGAN decoder. This design allows us to maintain consistency between restored results and the clean background, while preserving the quality of the prior features.

Hierarchical Texture Warping Head. For further alleviating the texture defects in results, we futher provide a solution to refine high-resolution features by the guidance of hierarchical high-quality codebook features. Specifically, in ith stage of our Hierarchical Texture Warping Head, we employ the deformable convolution and channel/spatial attention layers to build texture warping module, align the features F_i^G from VQGAN Decoder G with the feature F_i^d from D_d:

$$F_i^{tw} = \text{TWM} \circ \text{Concat}(F_i^d, F_i^G) \tag{4}$$

where F_i^{tw} denotes the warped features by our Texture Warping Module (TWM). F_i^{tw} will be fused with features of corresponding stage in the head for the texture refinement. We notice that the utilization of channel/spatial attention could further boost the feature warping of F_i^{tw} in a cheap way. The effectiveness of the proposed Hierarchical Texture Warping Head module is detailed in our ablation study section.

3.3. Model Objective

We present the detailed model objective of VQGAN and AWRCP in this section. Let's denote the degraded image as x_d, the restored result by AWRCP as y_r, the corresponding ground truth image as y_g, and the high-quality input image for VQGAN training as x_h. Furthermore, the reconstructed result by VQGAN is denoted as x_r. In the first training stage, the VQGAN only learns from high-quality images without any degradation or noises. In the second training stage, the codebook of VQGAN and its Decoder G are frozen, and a new encoder E_d is developed for better code matching and improved restoration performance. Accordingly, the objectives of VQGAN and AWRCP for training are as follows.

3.3.1 VQGAN

Due to the non-differentiable vector-quantized operation, we train VQGAN and it's codebook by copying the gradients of $\mathbf{G}$ to $\mathbf{E}$. And we adopt four image-level reconstruction losses for VQGAN: L1 loss $\mathcal{L}_{rec}$ for basic pixel reconstruction, perceptual loss $\mathcal{L}_{per}$ [31] for perceptual quality, adversarial loss $\mathcal{L}_{adv}$ for texture generation, and semantic guided loss $\mathcal{L}_{semantic}$ to encourage the texture to be conditioned on semantics.

Pixel Reconstruction Loss. We utilize the L1 loss in the RGB color space as the basic reconstruction loss, which can be denoted as:

$$\mathcal{L}_{rec} = \left\| x^r - x^h \right\|_1 \tag{5}$$

Codebook Loss. For codebook optimization, we employ codebook loss to reduce the distance between codebook and input feature embeddings and update codebook:

$$\mathcal{L}_{codebook} = \left\| sg(z_d) - z_q^d \right\|_2^2 + \beta \left\| sg\left(z_q^d\right) - z_d \right\|_2^2 \tag{6}$$

where $sg(\cdot)$ is the stop-gradient operation, and $\beta = 0.25$.

Adversarial Loss. Following VQGAN [12], we utilize the adversarial loss item to improve texture quality of reconstructed x_r:

$$\mathcal{L}_{adv} = \left[\log D\left(x_h\right) + \log\left(1 - D\left(x_r\right)\right)\right] \tag{7}$$

Semantic Guidance Loss. The codebook is learned purely by gradient descent where similar patterns are clustered independent of their semantics. In order to maintain consistency between the semantic information and textures of codebook embedding, we regularize the learning of codebook by incorporating perceptual features that hold rich semantic information by adding semantic guidance loss follow FeMaSR [2]:

$$\mathcal{L}_{semantic} = \left\| Conv\left(z_d^q\right) - \phi(x_h) \right\|_2^2 \tag{8}$$

where $conv$ denotes a 1×1 convolution layer to adjust the dimension of z_d^q and $\phi(x_h)$. ϕ denotes the pre-trained VGG19 network.

Finally, the training objective of VQGAN is:

$$\begin{aligned} \mathcal{L}_{vq} = \lambda_r \mathcal{L}_{rec} + \lambda_c \mathcal{L}_{codebook} + \lambda_a \mathcal{L}_{adv} + \\ \lambda_s \mathcal{L}_{semantic} + \lambda_p \mathcal{L}_{per} \end{aligned} \tag{9}$$

3.3.2 AWRCP

For better restoration performance, we adopt PSNR loss [4] as the reconstruction loss of AWRCP. The loss function can be calculated as:

$$\mathcal{L}_{psnr} = -\text{PSNR}(\textbf{AWRCP}(x_d), y_g), \tag{10}$$

In addition, perceptual level of the restored image is also critical. We also applied the perceptual loss to improve the restoration performance of AWRCP. Overall loss function can be expressed as:

$$\mathcal{L}_{AWRCP} = \lambda_1 \mathcal{L}_{psnr} + \lambda_2 \mathcal{L}_{per}, \tag{11}$$

where the λ_1 and λ_2 are set to 1 and 0.2.

4. Experiments

4.1. Datasets

We utilize five standard benchmark image restoration datasets considering various adverse weather conditions, such as real and synthetic snow, heavy rain with haze, and real and synthetic raindrops and rain streaks.

Snow100K [39] is a widely-used desnowing benchmark for evaluation of snow removal methods. Its test datasets consists of three synthetic sub-test sets: Small, Middle, Large (Snow100k-S/M/L).

	Snow100K-S [39]		Snow100K-L [39]	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
SPANet [51]	29.92	0.8260	23.70	0.7930
JSTASR [8]	31.40	0.9012	25.32	0.8076
RESCAN [34]	31.51	0.9032	26.08	0.8108
DesnowNet [39]	32.33	0.9500	27.17	0.8983
DDMSNet [66]	34.34	0.9445	28.85	0.8772
MPRNet [62]	34.97	0.9457	29.76	0.8949
NAFNet [3]	34.79	0.9497	30.06	0.9017
Restormer [64]	35.03	0.9487	<u>30.83</u>	<u>0.9121</u>
All-in-One [33]	-	-	28.33	0.8820
TransWeather [49]	32.51	0.9341	29.31	0.8879
TKL&MR [10]	34.80	0.9483	30.24	0.9020
WeatherDiff$_{64}$[41]	<u>35.83</u>	<u>0.9566</u>	30.09	0.9041
WeatherDiff$_{128}$[41]	35.02	0.9516	29.58	0.8941
AWRCP(Ours)	**36.92**	**0.9652**	**31.92**	**0.9341**

(a) Image Desnowing

	Outdoor-Rain [32]	
	PSNR ↑	SSIM ↑
CycleGAN [69]	17.62	0.6560
pix2pix [21]	19.09	0.7100
HRGAN [32]	21.56	0.8550
PCNet [25]	26.19	0.9015
MPRNet [63]	28.03	0.9192
NAFNet [3]	29.59	0.9027
Restormer [64]	<u>29.97</u>	0.9215
All-in-One [33]	24.71	0.8980
TransWeather [49]	28.83	0.9000
TKL&MR [10]	29.92	0.9167
WeatherDiff$_{64}$[41]	29.64	<u>0.9312</u>
WeatherDiff$_{128}$[41]	29.72	0.9216
AWRCP(Ours)	**31.39**	**0.9329**

(b) Image Deraining & Dehazing

	RainDrop [43]	
	PSNR ↑	SSIM ↑
pix2pix [21]	28.02	0.8547
DuRN [36]	31.24	0.9259
RaindropAttn [45]	31.44	0.9263
AttentiveGAN [43]	31.59	0.9170
IDT [56]	<u>31.87</u>	<u>0.9313</u>
All-in-One [33]	31.12	0.9268
TransWeather [49]	30.17	0.9157
TKL&MR [10]	30.99	0.9274
WeatherDiff$_{64}$[41]	30.71	0.9312
WeatherDiff$_{128}$[41]	29.66	0.9225
AWRCP(Ours)	**31.93**	**0.9314**

(c) Removing Raindrops

Table 1. Quantitative comparisons in terms of PSNR and SSIM (higher is better) with state-of-the-art image desnowing, deraining, adverse weather removal methods. Best and second best values are indicated with **bold** text and <u>underlined text</u> respectively. Above half of the tables show comparisons of methods individually evaluated for each task. Bottom half of the tables show performance results of our unified multi-weather method AWRCP on all four test sets with the state-of-the-art adverse weather removal methods All-in-One [33], TransWeather [49], TKL&MR [10] and WeatherDiff [41].

Outdoor-Rain [32] is a classical dataset for simultaneous image deraining and dehazing. This dataset contains dense synthetic rain streaks and heavy hazy degradation. Its testing dataset has 750 high-resolution images for evaluation.

RainDrop [43] contains images of raindrops that introduce real artifacts on the camera sensor and obstruct the view. For quantitative evaluations, the dataset includes a testing subset, referred to as Raindrop-A in prior research [41, 49], which comprises 58 images.

4.2. Training Details

VQGAN. For the training of VQGAN, we employ Adam optimizer with a fixed learning rate of 1×10^{-4}. We utilize the high-quality images from widely-used DIV2k [1] and Flickr2k [35] to train our VQGAN in the first training stage.

AWRCP. For the training setting of AWRCP, our manner only need train at once on a mixed adverse weather dataset *Allweather* from TransWeather [49], which has 18,069 adverse weather samples from the training dataset of Snow100K, Outdoor-Rain and Raindrop. We augment the training dataset with randomly rotated by $90, 180, 270$ degrees and horizontal flip. The training image patches with the size 256×256 are extracted as input data of our AWRCP. We utilize Adam optimizer with the initial learning rate of 5×10^{-4}, and adopt the CyclicLR to adjust the learning rate progressively, where on the triangular mode, and the gamma weight is 1.0, the base momentum is 0.9, the max learning rate is 8×10^{-4} and the base learning rate is the same as initial one. We utilize Pytorch [42] framework to implement our method with 8 NVIDIA A100 GPU with total batch size of 160. For training stage, we train on the mixed adverse weather dataset for 2000 epochs totally. We empirically set $\lambda_r, \lambda_c, \lambda_a, \lambda_s$ and λ_p as $\{1, 1, 1, 0.25, 0.2\}$. Please refer to our Supplemental Materials for details of AWRCP's architecture.

4.3. Quantitative Comparison.

We present a comparative analysis of metrics between synthetic and real datasets, as summarized in the Table 1a, 1b, 1c. We re-trained recent multiple degradation removal methods MPRNet [62], NAFNet [3] and Restormer [64] as weather-specific methods on each benchmark for a fair and convinced comparison. And We also re-trained unified adverse weather removal method TKL&MR [10] on *allweather* [49, 33] training dataset for an exhaustive comparison. The results indicate that our proposed method outperforms existing approaches by a significant margin across three different degradation types.

4.4. Visual Comparison.

We also perform the visual comparison and the results are shown in Fig.5 and 6. The results show that our method can comprehensively eliminate snow degradation, including fine snow marks and large snow spots. In contrast, the latest WeatherDiffusion [41] method still exhibits some slight snow degradation, and its ability to restore details is not ideal. As for restoring harsh weather conditions, our AWRCP method is very effective in removing complex haze and rain marks and produces more attractive visuals compared to previous methods.

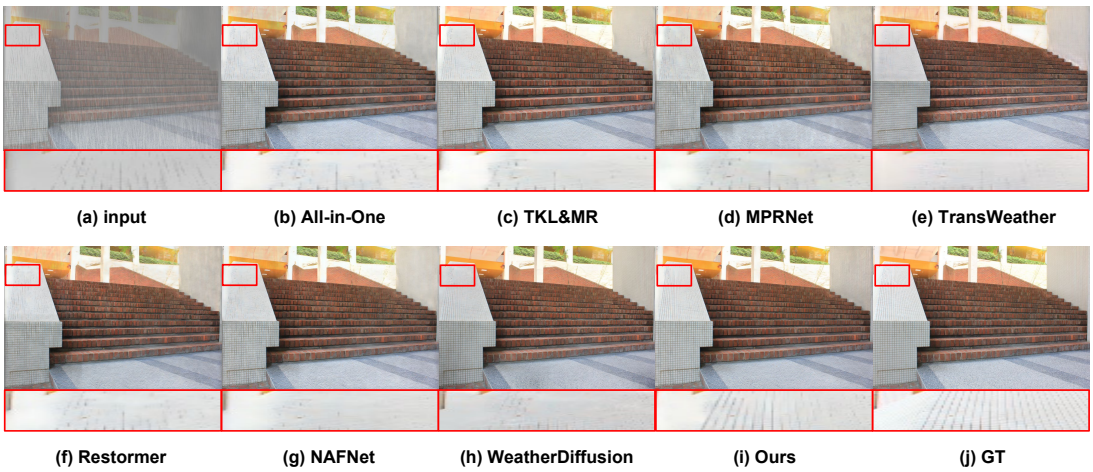

Figure 5. Visual comparisons of synthetic rain and fog image from the Outdoor-Rain [33] testing dataset. The image can be zoomed in for improved visualization.

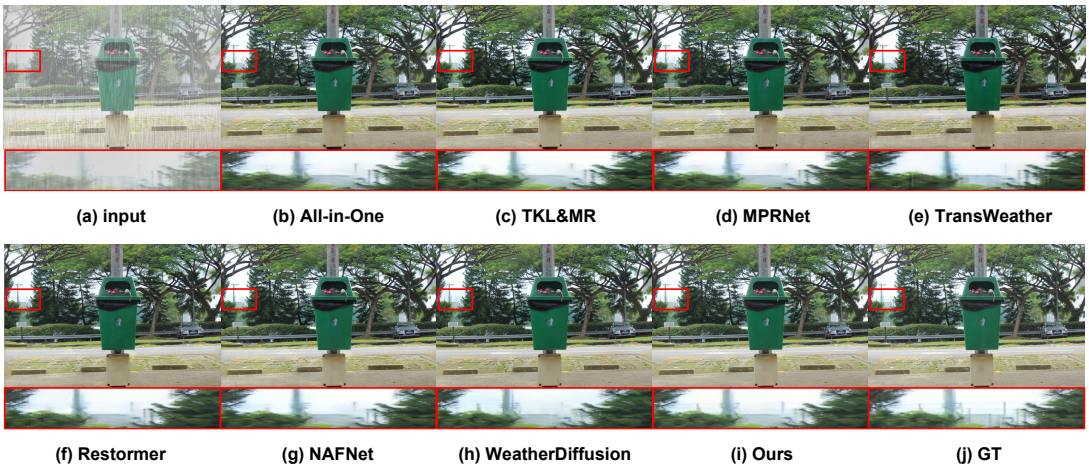

Figure 6. Visual comparisons of the synthetic rain and fog image from the Outdoor-Rain [33] testing dataset. The image can be zoomed in for improved visualization.

4.5. Further analyzing VQ-GAN's distortion in object structures.

There are two main factors that cause structure distortion when using VQ-GAN: (i) According to rate-distortion theory[48], the best achievable reconstruction result depends on the number of bits utilized. To achieve the best possible reconstruction result, VQ-GAN requires $HWlog_2K$ bits to represent a $H \times W$ image as codes, where K is the codebook size. Therefore, a larger codebook size is necessary in order to maintain reconstruction quality while reducing resolution. However, VQ-GAN with a large codebook can become inefficient and unstable because of the codebook collapse problem during the learning phase. Therefore, to trade efficiency and reconstruction performance, the version of VQ-GAN incorporated in AWRCP has a codebook with a limited size, leading to information loss and distortion. (ii) Furthermore, **the texture quality** of VQ-GAN is optimized by an adversarial loss in our work, but **the structure quality** is not optimized by any specific loss. So the reconstruction results of VQ-GAN can have high-quality texture but suffer from structure distortion problem.

Table 2. Ablation studies on Parallel Decoder Design.

Setting	Model	PSNR	SSIM
i	D_d	28.12	0.8924
ii	Fixed G	23.01	0.8642
iii	Finetune G	27.23	0.8719
iv	Ours	**31.39**	**0.9329**

Table 3. Ablation studies on Deformable Cross-Attention.

Setting	Model	PSNR	SSIM
i	WCA	30.21	0.9108
ii	WCA+DeConv	30.71	0.9218
iii	SFT	30.12	0.9185
iv	Ours	**31.39**	**0.9329**

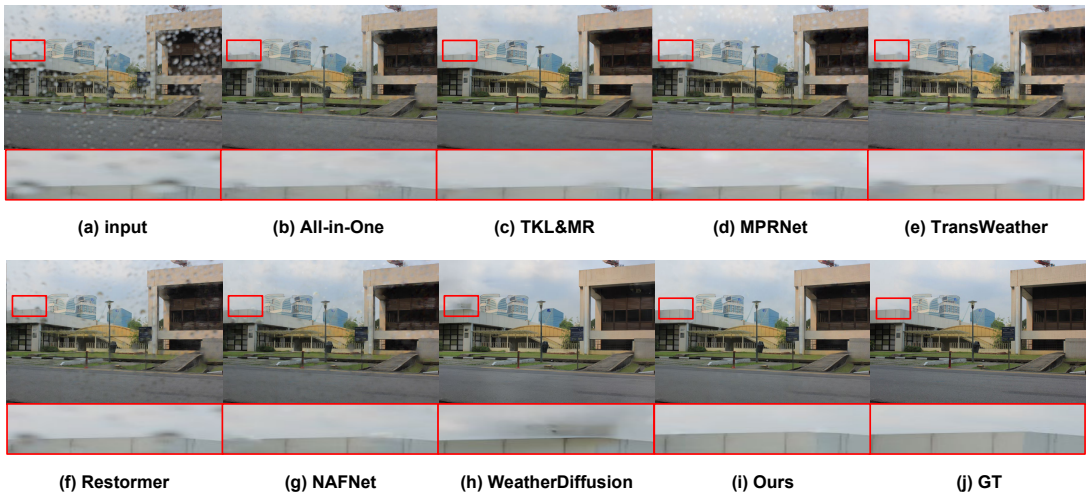

Figure 7. Visual comparisons of the real-world raindrop image. The image can be zoomed in for improved visualization.

Table 4. Ablation studies on Hierarchical Texture Warping Head.

Setting	Model	PSNR	SSIM
i	w/o HTWH	30.91	0.9261
ii	w/o DeConv	31.21	0.9301
iii	w/o CA&SA	31.44	0.9312
iv	Ours	**31.39**	**0.9329**

4.6. Ablation Study

In order to verify the efficacy of each key component of AWRCP, we conducted a series of ablation experiments. Specifically, we discussed the effectiveness of Parallel Decoder Design, Deformable Cross-Attention and Hierarchical Texture Warping Head on the Outdoor-Rain test set. All of these variants were trained using the same settings as our AWRCP.

Effectiveness of Parallel Decoder Design. In this subsection, we discuss the performance impact of decoder design. For exploring it, we propose 3 variants to replace the parallel decoder design in AWRCP, which are: *(i) using only D_d to generate high-quality features from the codebook; (ii) employing a fixed, pretrained G of VQGAN as the decoder; and (iii) fine-tuning the fixed, pretrained G of VQGAN as the decoder.* Table. 2 shows that utilizing only D_d is not powerful enough to achieve excellent performance. Using a fixed, pretrained G or fine-tuning it yields unsatisfactory results due to the huge divergence between image reconstruction and image restoration. Our full decoder solution delivers the best results in terms of PSNR and SSIM.

Effectiveness of Deformable Cross-Attention. To demonstrate the superior performance of our Deformable Cross-Attention, we compared it with several classical feature fusion manners. As shown in Table. 3, we compared with three different solutions: *(i) Vanilla window-based cross-attention [5] (WCA) to perform feature fusing; (ii) utilizing Deformable convolutional layers to warp high-quality features and employing WCA to fuse them (WCA+DeConv); (iii) using classical Spatial Feature Transform [52] to fuse different features (SFT).* We can observe that our manner achieves better restoration performance compared with other solutions.

Effectiveness of Hierarchical Texture Warping Head. To study the effectiveness of the proposed Hierarchical Texture Warping Head, three variant settings are presented in Table. 4: *(i) Without Hierarchical Texture Warping Head (w/o HTWH); (ii) removing deformable convolution layer in Texture Warping Module (w/o DeConv); (iii) removing both channel attention layer and spatial attention layers in Texture Warping Module (w/o CA&SA).* Results indicate that each component of the TWM is essential for the head, as it can improve the ability of the deformable convolutions to cope with misalignment between different features due to the presence of channel and spatial attention layers.

5. Conclusion

In this paper, we propose a novel paradigm for improving adverse weather removal using codebook priors. Through careful observation and analysis, we introduce a high-quality codebook obtained from a well-trained VQGAN model for the task of adverse weather removal. By effectively leveraging codebook priors, AWRCP is able to recover realistic texture details and achieve superior restoration performance. Extensive experiments demonstrate that the proposed framework achieves state-of-the-art performance.

Acknowledgment. This work was supported by Natural Science Foundation of Fujian Province (2021J01867), and the National Natural Science Foundation of China (Grant No. 61902275).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

[2] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1329–1338, 2022.

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022.

[4] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–192, 2021.

[5] Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2022.

[6] Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Yi Ye, and Erkang Chen. Msp-former: Multi-scale projection transformer for single image desnowing. *arXiv preprint arXiv:2207.05621*, 2022.

[7] Sixiang Chen, Tian Ye, Jun Shi, Yun Liu, JingXia Jiang, Erkang Chen, and Peng Chen. Dehrformer: Real-time transformer for depth estimation and haze removal from varicolored haze scenes. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[8] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *European Conference on Computer Vision*, pages 754–770. Springer, 2020.

[9] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021.

[10] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen 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*, pages 17653–17662, 2022.

[11] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020.

[12] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021.

[13] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, pages 126–143. Springer, 2022.

[14] Geoffrey E Hinton and Richard Zemel. Autoencoders, minimum description length and helmholtz free energy. *Advances in neural information processing systems*, 6, 1993.

[15] Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6043–6052, June 2022.

[16] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *European Conference on Computer Vision*, pages 163–180. Springer, 2022.

[17] Jie Huang, Zhiwei Xiong, Xueyang Fu, Dong Liu, and Zheng-Jun Zha. Hybrid image enhancement with progressive laplacian enhancing unit. In *Proceedings of the 27th ACM International Conference on Multimedia*, page 1614–1622, 2019.

[18] Jie Huang, Feng Zhao, Man Zhou, Jie Xiao, Naishan Zheng, Kaiwen Zheng, and Zhiwei Xiong. Learning sample relationship for exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9904–9913, June 2023.

[19] Jie Huang, Man Zhou, Yajing Liu, Mingde Yao, Feng Zhao, and Zhiwei Xiong. Exposure-consistency representation learning for exposure correction. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6309–6317, 2022.

[20] Jie Huang, Pengfei Zhu, Mingrui Geng, Jiewen Ran, Xingguang Zhou, Chen Xing, Pengfei Wan, and Xiangyang Ji. Range scaling global u-net for perceptual image enhancement on mobile devices. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[21] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[22] Jingxia Jiang, Jinbin Bai, Yun Liu, Junjie Yin, Sixiang Chen, Tian Ye, and Erkang Chen. Rsfdm-net: Real-time spatial and frequency domains modulation network for underwater image enhancement. *arXiv preprint arXiv:2302.12186*, 2023.

[23] Jingxia Jiang, Tian Ye, Jinbin Bai, Sixiang Chen, Wenhao Chai, Shi Jun, Yun Liu, and Erkang Chen. Five a^{+} network: You only need 9k parameters for underwater image enhancement. *arXiv preprint arXiv:2305.08824*, 2023.

[24] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale

progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8346–8355, 2020.

[25] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zheng Wang, Xiao Wang, Junjun Jiang, and Chia-Wen Lin. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Transactions on Image Processing*, 30:7404–7418, 2021.

[26] Yeying Jin, Ruoteng Li, Wenhan Yang, and Robby T Tan. Estimating reflectance layer from a single image: Integrating reflectance guidance and shadow/specular aware learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 1069–1077, 2023.

[27] Yeying Jin, Aashish Sharma, and Robby T Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network (supplementary paper).

[28] Yeying Jin, Wending Yan, Wenhan Yang, and Robby T Tan. Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. In *Asian Conference on Computer Vision*, pages 155–172. Springer, 2022.

[29] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, pages 404–421. Springer, 2022.

[30] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Shadowdiffusion: Diffusion-based shadow removal using classifier-driven attention and structure preservation. *arXiv preprint arXiv:2211.08089*, 2022.

[31] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.

[32] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1633–1642, 2019.

[33] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3175–3185, 2020.

[34] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*, pages 254–269, 2018.

[35] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[36] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019.

[37] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, and Erkang Chen. Nighthazeformer: Single nighttime haze removal using prior query transformer. *arXiv preprint arXiv:2305.09533*, 2023.

[38] Yun Liu, Zhongsheng Yan, Aimin Wu, Tian Ye, and Yuche Li. Nighttime image dehazing based on variational decomposition model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 640–649, 2022.

[39] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073, 2018.

[40] Keyu Yan Hu Yu Xueyang Fu Aiping Liu Xian Wei Feng Zhao Man Zhou, Jie Huang. Spatial-frequency domain information integration for pan-sharpening. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

[41] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[42] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[43] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2482–2491, 2018.

[44] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020.

[45] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2463–2471, 2019.

[46] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3937–3946, 2019.

[47] Jingjing Ren, Qingqing Zheng, Yuanyuan Zhao, Xuemiao Xu, and Chen Li. Dlformer: Discrete latent transformer for video inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3511–3520, 2022.

[48] Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec*, 1959.

[49] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal 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*, pages 2353–2363, 2022.

[50] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

[51] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12270–12279, 2019.

[52] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 606–615, 2018.

[53] Zhouxia Wang, Jiawei Zhang, Runjian Chen, Wenping Wang, and Ping Luo. Restoreformer: High-quality blind face restoration from undegraded key-value pairs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17512–17521, 2022.

[54] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021.

[55] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[56] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[57] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Jinbin Bai, and Erkang Chen. Towards real-time high-definition image snow removal: Efficient pyramid network with asymmetrical encoder-decoder architecture. In *Proceedings of the Asian Conference on Computer Vision*, pages 366–381, 2022.

[58] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Erkang Chen, and Yuche Li. Underwater light field retention: Neural rendering for underwater imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 488–497, 2022.

[59] Tian Ye, Mingchao Jiang, Yunchen Zhang, Liang Chen, Erkang Chen, Pen Chen, and Zhiyong Lu. Perceiving and modeling density is all you need for image dehazing. *arXiv preprint arXiv:2111.09733*, 2021.

[60] Hu Yu, Jie Huang, Yajing Liu, Qi Zhu, Man Zhou, and Feng Zhao. Source-free domain adaptation for real-world image dehazing. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6645–6654, 2022.

[61] Hu Yu, Jie Huang, Yajing Liu, Qi Zhu, Man Zhou, and Feng Zhao. Source-free domain adaptation for real-world image dehazing. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6645–6654, 2022.

[62] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[63] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.

[64] Syed Waqas Zamir, Aditya Arora, Salman Khan, M unawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *arXiv preprint arXiv:2111.09881*, 2021.

[65] Jinghao Zhang, Jie Huang, Mingde Yao, Man Zhou, and Feng Zhao. Structure- and texture-aware learning for low-light image enhancement. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6483–6492, 2022.

[66] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30:7419–7431, 2021.

[67] Naishan Zheng, Jie Huang, Feng Zhao, Xueyang Fu, and Feng Wu. Unsupervised underexposed image enhancement via self-illuminated and perceptual guidance. *IEEE Transactions on Multimedia*, pages 1–16, 2022.

[68] Man Zhou, Keyu Yan, Jie Huang, Zihe Yang, Xueyang Fu, and Feng Zhao. Mutual information-driven pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1798–1808, June 2022.

[69] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.