All-in-one Multi-degradation Image Restoration Network via Hierarchical Degradation Representation

Cheng Zhang
zhangcheng233@mail.nwpu.edu.cn
Northwestern Polytechnical
University
Xi'an, China

Yu Zhu*
yuzhu@nwpu.edu.cn
Northwestern Polytechnical
University
Xi'an, China

Qingsen Yan
qingsenyan@nwpu.edu.cn
Northwestern Polytechnical
University
Xi'an, China

Jinqiu Sun
sunjinqiu@nwpu.edu.cn
Northwestern Polytechnical
University
Xi'an, China

Yanning Zhang
ynzhang@nwpu.edu.cn
Northwestern Polytechnical
University
Xi'an, China

ABSTRACT

The aim of image restoration is to recover high-quality images from distorted ones. However, current methods usually focus on a single task (*e.g.*, denoising, deblurring or super-resolution) which cannot address the needs of real-world multi-task processing, especially on mobile devices. Thus, developing an all-in-one method that can restore images from various unknown distortions is a significant challenge. Previous works have employed contrastive learning to learn the degradation representation from observed images, but this often leads to representation drift caused by deficient positive and negative pairs. To address this issue, we propose a novel All-in-one Multi-degradation Image Restoration Network (AMIRNet) that can effectively capture and utilize accurate degradation representation for image restoration. AMIRNet learns a degradation representation for unknown degraded images by progressively constructing a tree structure through clustering, without any prior knowledge of degradation information. This tree-structured representation explicitly reflects the consistency and discrepancy of various distortions, providing a specific clue for image restoration. To further enhance the performance of the image restoration network and overcome domain gaps caused by unknown distortions, we design a feature transform block (FTB) that aligns domains and refines features with the guidance of the degradation representation. We conduct extensive experiments on multiple distorted datasets, demonstrating the effectiveness of our method and its advantages over state-of-the-art restoration methods both qualitatively and quantitatively.

CCS CONCEPTS

• **Computing methodologies** → **Image and video acquisition**; **Computational photography**.

KEYWORDS

Neural Network, Image Restoration, Degradation Representation

1 INTRODUCTION

Image restoration is a critical topic in low-level vision, which generates a high-quality image from a damaged image caused by degradation, *e.g.*, blurriness, noise, and low illumination. While a particular type of degradation typically prevails in an image, it is common to encounter situations where multiple degradations [1] need to be processed in the real world. For instance, such a scenario may arise when capturing photography with different camera parameters [4, 38], collecting image data from the internet [24], and taking pictures under adverse weather conditions [29].

The majority of the existing restoration methods [2, 5, 6, 14, 20, 32, 35, 37, 39, 40] are designed for a single degradation (*e.g.* denoising, deblurring, or super-resolution), and even some methods [5, 20, 37] that claim to handle multiple degradations require separate training on categorized and specific degraded images, which is computationally expensive, time-consuming for optimization, and not friendly for storage on mobile devices. Moreover, specifying the degradation types adds complexity to the usage and increases the risk of performance degradation. Therefore, considering these issues, an all-in-one approach is the optimal solution for restoring images with multiple unknown degradations, as it allows for unified training and convenient testing with the same parameters and architecture, thus reducing complexity and increasing ease of use.

Within the all-in-one framework, the crucial problem to be tackled is how to represent and leverage the degradation information in the restoration network, since improved representation leads to enhanced restoration performance, particularly in situations involving multiple degradations. Some degradation estimation methods [7, 12, 14, 15, 21, 28, 36] typically assume a predefined degradation category and estimate degradation level parameters, making them less suitable for scenarios with multiple unknown degradations. While some methods [16, 30] rely on contrastive learning for degradation representation, the deficient selection of positive and negative samples in these approaches makes it difficult to comprehensively describe the relationships between different degradations, which can lead to representation drift and ultimately impacts the performance of image restoration.

To tackle the aforementioned challenges, we propose a novel All-in-one network, named AMIRNet, to handle multi-degraded images via learning hierarchical degradation representation. Specifically, we observe that the degradations exhibit a characteristic of

---

*Corresponding author

---

[1] Following [16], the multi-degradation that this paper focuses on refers to a dataset with multiple degradations, differing from mixed degradations in an image.

hierarchical subordination, as illustrated in Figure 1. As an example, in a dataset with multiple degradations, two images may be grouped together based on the presence of blur, but a more specific lower-level cluster can further categorize them into defocus blur and motion blur, respectively. The hierarchical structure enables the modeling of commonalities and distinctions among image degradations, which provides a beneficial clue for all-in-one image restoration under multiple degradations.

Therefore, we propose a tree-structured representation to capture the relationship between multi-degradations and progressively construct the representation through clustering from coarse to fine. Similar to training networks separately on individual degraded datasets [5, 20, 37], the use of the hierarchical degradation representation is aimed at making feature distributions of similar degradations more compact and easily distinguishable from dissimilar degradation features. Additionally, to sufficiently leverage the degradation representation in the all-in-one restoration network, we devise a feature transform block (FTB) to integrate the image feature and corresponding degradation representation. Considering the domain gap caused by distortions, we draw inspiration from [3] and incorporate degradation-related layer normalization in the FTB to align domains. We also introduce a degradation-related gating mechanism in the FTB to control the information flow in the restoration network. The FTB can refine image features, enabling the network to adapt to various degraded images. Furthermore, we conduct comprehensive experiments and ablation studies on multi-degradation datasets to verify the effectiveness of our method.

Our main findings and contributions can be summarized as follows:

- We propose a novel All-in-one Multi-degradation Image Restoration Network (AMIRNet) to handle adverse degraded images in the real world.
- Based on the observation that multi-degraded images follow a hierarchical structure, we propose to progressively construct a tree-structured representation by clustering to characterize the similarity and difference between degradations.
- We devise a feature transform block (FTB) to overcome domain gaps caused by various distortions and refine image features with the guidance of degradation representation.
- Extensive experiments are conducted to confirm the effectiveness of our method, which not only has the potential for promising results but also achieves state-of-the-art performance on multi-degradation datasets.

2 RELATED WORKS

2.1 Single Degradation Restoration

Image restoration is a fundamental task in computer vision, which aims to recover the degraded images to their original high-quality versions, including deblurring, denoising, inpainting, low-light enhancement, and so on. Traditional approaches focus on the exploration of the image prior, such as sparse [23, 25, 33], low-rank [13, 32], self-similarity [9] *etc.*.

Recently, with the support of a large number of collected paired images, many deep neural networks (DNN) methods [2, 5, 6, 20, 31, 34, 35, 37, 39] have produced impressive results on each sub-task of restoration. Those works emphasize the design of network
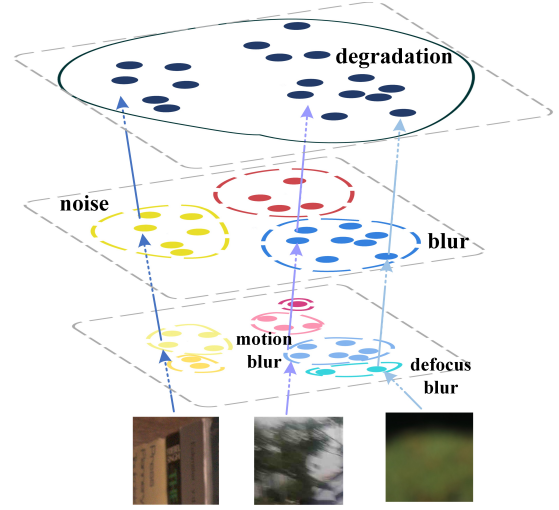

Figure 1: Degradation possesses a characteristic of hierarchical subordination *e.g.* degradation → blur → motion blur from coarse to fine. As presented in the figure, two blurred images may be clustered as blur, but at a lower level of the hierarchy, they are distinguished as motion blur and defocus blur.

architecture and loss functions. By utilizing images that are usually captured and categorized manually by humans according to degradation type, DNN-based methods focus on learning the implicit mapping between the distorted image and the high-quality image. Although some models [5, 6, 20, 37] can be adapted to handle multiple types of degraded images, they typically require separate training on specific degraded datasets, and they may not generalize well to other typed of degraded images without further adaptation. For instance, MPR [37] trained well for image denoising has limited performance on image deblurring, which is not expected in practice. Therefore, it is vital to consider the fact that images are often corrupted by multiple degradations and devise all-in-one solutions to meet the requirement of real-world multi-task processing.

2.2 Multi-degradation Restoration

Currently, there has been increased interest in developing all-in-one models that can handle various degraded images in a single network after being trained. For instance, [19] proposes a network with multiple encoders to process each degradation using a specific encoder. Similarly, Transweather [29] introduces a decoder with learnable embeddings in a transformer architecture to address multiple degradation types.

Instead of only focusing on designing network structures, there are also techniques that incorporate contrastive learning to enable the network to differentiate between different types of corruptions and handle multi-degraded images. One such example is presented in [8], which suggests using both soft and hard contrastive regularization to enhance the performance of both specific and multiple degradations. Other approaches, DASR [30] and AirNet [16] consider patches from the same image as positive samples and patches from different images as negative samples. However, this selection
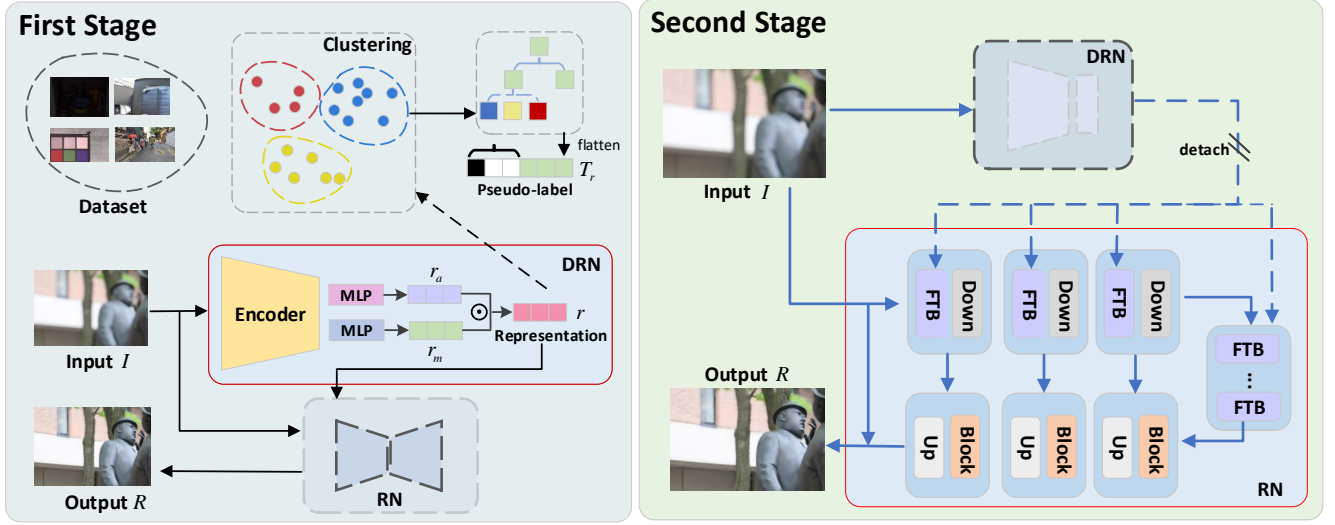
Figure 2: The overview of our proposed All-in-one Multi-degradation Image Restoration Network (AMIRNet), which consists of two main components, the degradation representation network (DRN) for learning hierarchical degradation representation and the restoration sub-network (RN) for recovering a high-quality image from a distorted image. In the first stage, we focus on constructing the degradation representation and training the DRN. In the second stage, we fix the parameters of the DRN and retrain the RN.

of positive and negative samples may not always be adequate, since it does not take into account situations where different degradations are related or where different images may belong to the same degradation, which could result in representation drift and performance drop of restoration network.

2.3 Degradation Representation

The representation of degradation is a crucial step in image restoration and serves as a prerequisite for accurately restoring a degraded image. Generally, the type of degradation is known before the estimation and the main task is to estimate the parameters of the degradation model. For example, in denoising [14, 21], the core is to estimate the noise level when the noise type is known. And in deblurring, many methods [7, 12, 15, 28, 36] usually estimate the blur kernel before non-blind deblurring. Additionally, there exist approaches that implicitly represent the degradation by learning a feature vector that serves as a proxy for the degradation and is subsequently fed into the restoration method. DASR [30] and AirNet [16] realize the learning of representation by contrastive learning. [17] proposes to learn degradation representations with a blurry-sharp cycle framework. [18] proposes to learn a latent representation space for degradations in super-resolution. Our method falls into this category, in which we guide the network to restore clear images by constructing hierarchical degradation representations of degraded images and adapting to different degraded image features.

3 METHODOLOGY

We aim to develop an **All-in-One** model that can effectively handle multiple types of degraded images, eliminating the requirement for retraining or fine-tuning once the training process is complete. As

mentioned above, the primary obstacle in developing an all-in-one method for multi-degradation is how to represent and leverage the degradation information within the restoration network. In this section, we will address the issue and present a detailed explanation of our proposed solution.

3.1 Overview

Our solution is a two-stage All-in-One approach, the first stage is designed to construct the hierarchical representation of degraded images, and the second stage is to remove degradation artifacts to produce a high-quality image with the supervision of degradation representation. The overview of our method is depicted in Figure 2. Despite serving different purposes, the two stages share a common network architecture. The network we propose comprises of two main components: a Degradation representation sub-Network (DRN) and a restoration sub-network (RN). Initially, a corrupted image I is fed into the encoder of the degradation representation sub-network to extract its features z, which are then transformed by projectors to yield the hierarchical degradation representation r in latent space. Subsequently, the degradation representation r is integrated into the restoration via the newly proposed feature transform block. With the assistance of the degradation representation r, our network is able to recover a high-quality image from the input distorted image.

3.2 Hierarchical Degradation Representation

As mentioned above and shown in Figure 1, the degradation has a characteristic of hierarchical subordination. The characteristic indicates that the images with adverse distortions can be classified into different categories in each layer of the hierarchy. An instance could be a motion-blurred image and a defocused image both categorized under "blur", however, their classification in the subgroup

---

Algorithm 1 Degradation Representation Construction

Require: Image pairs $D = (I_1, Y_1), (I_2, Y_2), ..., (I_n, Y_n)$, number of hierarchical layers L, number of clusters in each layer $C = c_1, c_2, ..., c_L$

Ensure: Tree-structured representation $T = T_1, T_2, ..., T_n$

1: Initialize T with the root node
2: **for** $i = 1$ to L **do**
3:     **for** $j = 1$ to c_i **do**
4:         Select $CurrentNode\ N_{i,j}$
5:         Obtain $r_k, k \in N_{i,j}$
6:         Cluster $T_{k,i} \leftarrow KMeans(r_k)$
7:         Update $T_k \leftarrow concat(T_k, T_{k,i})$
8:     **end for**
9:     Optimize DRN and RN
10: **end for**
11: **return** Tree-structured representation $T = T_1, T_2, ..., T_n$

---

of blur types is different. To model the commonality and the distinction between multiple distortions, naturally, we propose a novel tree-structured representation, which corresponds to the categories of the hierarchical structure of degradations. The tree-structured representation T can be flattened into a vector when performing a level-order traversal of the hierarchy, which is formulated as follows:

$$T = [v_{1,1}, ..., v_{i,j}, ..., v_{L,c_L}], v_{i,j} \in \{0, 1\}, \tag{1}$$

where L and c_L denote the number of levels in the tree structure and the number of nodes in each level, respectively. The binary value $v_{i,j}$ in j_{th} node of i_{th} level represents whether the image belongs to that node or not. In our method, we predefine the structure of the tree as a 4-layer binary tree, so that it can be used consistently during both training and testing. After establishing the specific form of the hierarchical representation, we need to take into account how to construct the representation for each corrupted image without additional degradation information and learn it with a degradation representation network.

Representation Construction. We recommend a progressive strategy to construct the degradation representation from the top layer to the bottom layer in the hierarchy, as described in Algorithm 1. Given the corrupted images I, clear images Y, the number of hierarchy L, and the number of clusters in each layer $c_1, c_2, ..., c_L$, our goal is to build a tree-structured representation T. The degradation representation of each input image is initially assigned to the root node and updated with the increasing depth of the tree during the outer loop. In each iteration of a layer, a node is sequentially selected as the current node, and the samples belonging to the current node are located to obtain their degradation representation using DRN. After clustering, the clustering results are concatenated with the original representations to update the degradation representations.

Representation Learning. To learn the degradation representation and facilitate usage during testing, a sub-network DRN is employed to extract degradation features and transform them into a low-dimensional embedding space under the supervision of the hierarchical tree-structured representation. Our DRN includes an encoder and two parallel projectors. The encoder $\mathcal{F}$ is composed of multiple convolutional layers with residual connections, activation

layers, and pooling layers. The encoder's objective is to extract features about degradations from degraded image I, and the process is expressed as $z = \mathcal{F}(I)$. The projectors are based on MLPs, which are composed of fully-connected layers and activation functions. One of the projectors, named mask projector $\mathcal{P}_m$, is designed to predict the binary mask $r_m = \mathcal{P}_m(z)$. The length of vector r_m is always consistent with the current clustering layer in the construction process of tree-structured representation and will increase with the number of nodes in the degradation tree.

Given a degraded image I and its degradation r, we calculate Cross Entropy loss to optimize the parameters of the encoder $\mathcal{F}$ and the mask projector $\mathcal{P}_m$:

$$\mathcal{L}_{cls} = CrossEntropy(r_m, T). \tag{2}$$

And the other parallel projector, named attribute projector P_a is only used in the first training stage, its output has the same size as the output of the first projector. The value on each vector dimension represents the attribute value at the corresponding node. In the first training stage, the degradation representation can be described as a product of mask r_m and attribute value $r_a = \mathcal{P}_a(z)$:

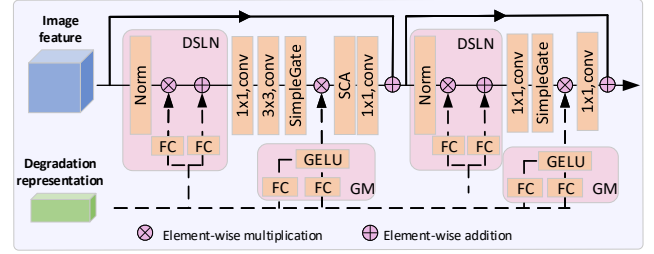$$r = r_m \cdot r_a. \tag{3}$$

Figure 3: The architecture of feature transform block (FTB) used in the restoration network. FTB is employed to align the domain gaps caused by adverse distortions and refine degraded image features.

3.3 Restoration with Degradation Guidance

Feature Transform Block. To leverage the hierarchical degradation representation in image restoration, we introduce a feature transform block to modulate the image feature transformation in latent feature space according to the distortion information. The structure of this block is depicted in Figure 3. Inspired by Domain-Specific Batch Normalization (DSBN) [3] in domain adaptation, which transforms domain-specific information into domain-invariant representation using the parameters of BN, we employ degradation-specific parameters in layer normalization (LN) to refine the degraded image features. LN is a widely used technique in restoration networks, which is expressed as

$$LN(x) = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} * \gamma + \beta, \tag{4}$$

where μ and σ denote the mean and standard deviation of the image feature $x \in \mathbb{R}^{C \times H \times W}$, and ϵ is a small constant to avoid dividing by zero, and γ and β are learnable affine parameters. In order to make the layer normalization adapt to various degraded image features,

Table 1: Quantitative comparison of our method and other state-of-the-art methods on multi-degraded image datasets. The datasets include various degradations such as blur, noise, low illumination, and defocus. All results are measured in terms of PSNR, SSIM, and LPIPS [43]. The red color indicates the best results, and the blue color indicates the second-best results.

Model	Param	FLOPs	Average			RED4[27]			SIDD[1]			LOL[31]			DPDD[2]		
			PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
SwinIR	11.46	752.1	27.97	0.7797	0.3725	25.58	0.7297	0.4084	36.24	0.9019	0.3149	17.74	0.7509	0.3989	24.59	0.7267	0.3945
MPR	20.13	1707.4	28.89	0.8098	0.3540	26.53	0.7599	0.4070	37.54	0.9125	0.3267	22.72	0.8406	0.2537	25.10	0.7641	0.3628
NAFNet	17.06	15.97	28.87	0.8089	0.3302	26.80	0.7611	0.3969	37.80	0.9211	0.2614	22.29	0.8280	0.2658	24.91	0.7584	0.3565
TransWeather	37.68	6.13	26.90	0.7508	0.4062	25.03	0.7210	0.4358	34.59	0.8571	0.4231	21.58	0.8071	0.2748	23.49	0.6979	0.3979
AirNet	5.77	301.3	28.29	0.7840	0.3796	25.82	0.7360	0.4125	37.25	0.9030	0.3041	13.79	0.7166	0.3639	24.78	0.7342	0.4136
Ours	71.76	73.23	29.39	0.8204	0.3204	26.77	0.7671	0.3875	38.46	0.9280	0.2608	22.83	0.8259	0.2662	25.45	0.7739	0.3412

we propose a Degradation-Specific Layer Normalization (DSLN). Formally, DSLN allocates degradation-specific affine parameters $\gamma(r)$ and $\beta(r)$ for different degradations.

Hence, DSLN can be written as:

$$\text{DSLN}(x) = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} * \gamma(r) + \beta(r), \qquad (5)$$

where $\gamma(r)$ and $\beta(r)$ are connected with the degradation representation r by using linear transformation matrixes, which is realized by a fully-connection layer in the implementation, i.e., $\gamma(r) = Wr + b$.

Moreover, we incorporate a Gating Mechanism (GM) to activate the channels of image features according to degradation. The gating mechanism is formulated as the element-wise product of image feature x and $\phi(r)$, and $\phi(r) = GELU(W^1 r + b^1)(W^2 r + b^2)$. The gating mechanism controls the information flow with the guidance of degradation representation r, thereby allowing the network to focus on the degradation-specific channels. Overall, with DSLN and gating mechanism, the FTB has the capability to align degradation domains and refine the image features under the guidance of degradation representation r, allowing the restoration network to produce high-quality images.

Training and Loss. Our restoration network is constructed based on NAFNet [5], which is a variant Unet with symmetric encoder-decoder architecture. In the encoder of our restoration network, the FTB serves as the fundamental module and modifies the degradation-related image features. In the decoder, efficient NAF-Blocks [5] are employed to transform modulated image features. The encoder features are concatenated with the decoder features via skip connections to avoid the gradient vanishing. Finally, a convolution layer is applied to generate a residual image. The specific structure of our restoration network can be referred to in Figure 2.

The restoration network takes the degraded image I and degradation representation r as input and performs adaptive processing to generate a high-quality image R. In the training phase, to optimize the parameters in the network, we adopt two commonly used losses, smoothL1 [11] and SSIM loss, which can measure the discrepancy of the restored result R and ground truth image Y at the pixel-wise level and patch-wise level, respectively. The optimization objective of our restoration network is the combination of two losses, which is formulated as follows:

$$\mathcal{L}_{res} = \mathcal{L}_{smoothL_1}(R, Y) + \alpha \mathcal{L}_{ssim}(R, Y), \qquad (6)$$

where α is a hyper-parameter to balance the two losses. In the first stage, the DRN is required to be optimized by cross-entropy

loss, therefore, the total loss in our method can be summarized as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{res}. \qquad (7)$$

In the second stage, the degradation representation r is produced by detached DRN, and we focus on the optimization of the restoration network. Hence, the total loss in the second stage is $\mathcal{L}_{res}$.

4 EXPERIMENTS

In this section, we conduct extensive experiments to verify the effectiveness of the proposed method. We first introduce the dataset and implementation details. Then we present the comparison with state-of-the-art methods across distortion types and across distortion levels. Finally, we conduct an ablation study to evaluate the effect of each component in our network and give a visualization of the hierarchical degradation representation.

4.1 Datasets and Implementation Details

Datasets. As our work primarily focuses on multi-degradation restoration, the dataset used should contain adverse types of degraded images. For this purpose, we train our network on a combination of different degradation datasets, including RED4[27] for deblurring and jpeg-compression removal, SIDD[1] for denoising, LOL [31] for low-light enhancement, and DPDD [2] for defocus deblurring. The training data consists of 1920 images uniformly sampled from the four datasets. And we randomly sample 539 images for testing to validate the effectiveness of our method. Moreover, to measure the network's performance on images with different degradation levels, we follow [16] to train our network on WED [24] and test it on synthetic noisy images with different noise levels from CBSD68 [26].

Implementation Details. We adopt NAFNet [5] as our restoration backbone, as it has demonstrated remarkable performance and computational advantages across multiple restoration tasks. We implement our approach using Pytorch framework and train the network on two NVIDIA A100 GPUs in a distributed manner. An AdamW [22] optimizer is adopted to optimize network parameters. The learning rate is initialized to 5e-4 and decreased with the CosineAnnealingLR decay strategy. The network is trained for 600 epochs. The batch size is set to 28 and the patch size is 256. In the first stage, the clustering algorithm is conducted every 150 epochs.

Motion Blur

(a) Input **(b) SwinIR** **(c) MPR** **(d) NAFNet**

(e) Transweather **(f) AirNet** **(g) Ours** **(h) Ground truth**

Defocus Blur

(a) Input **(b) SwinIR** **(c) MPR** **(d) NAFNet**

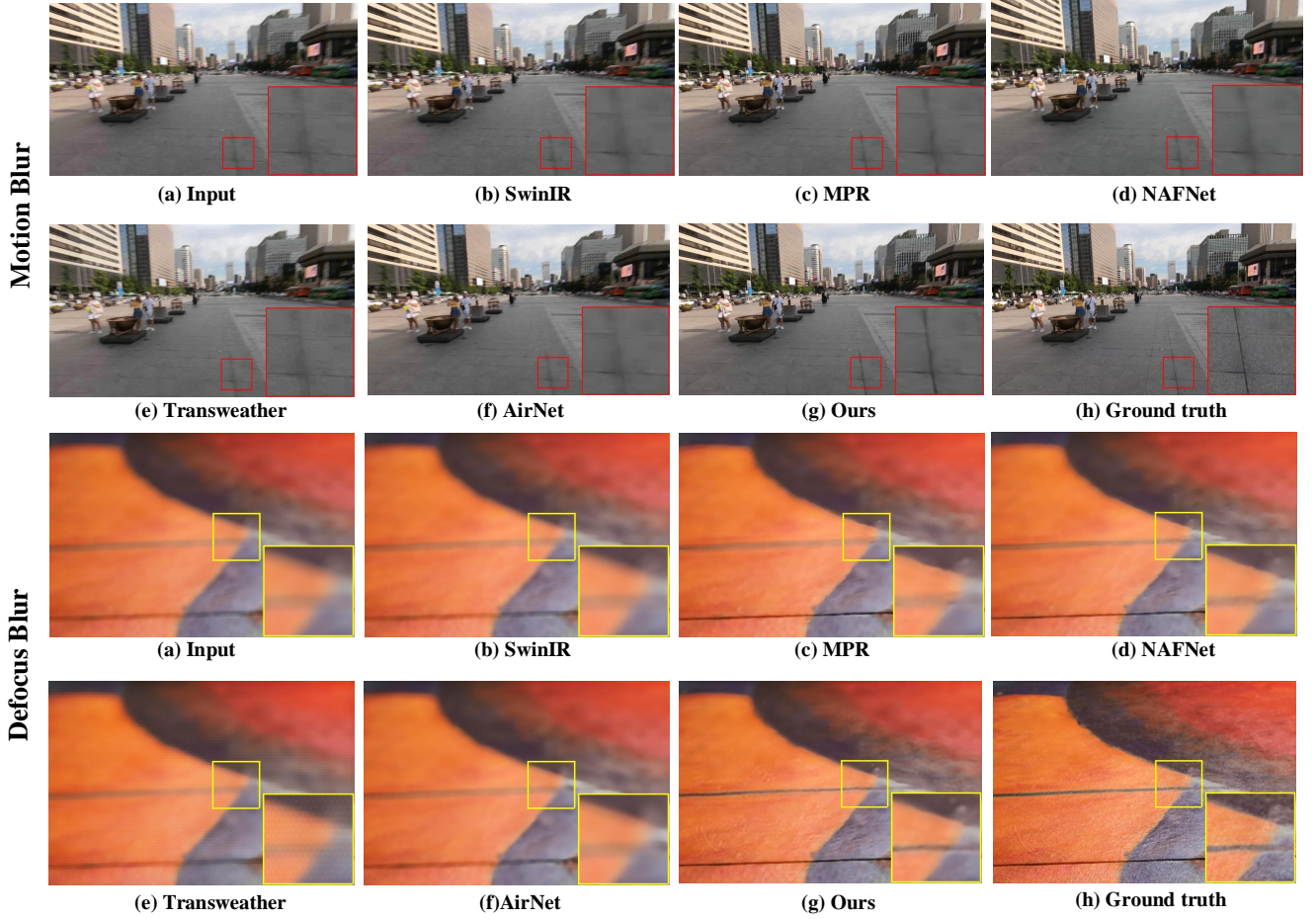**(e) Transweather** **(f)AirNet** **(g) Ours** **(h) Ground truth**

Figure 4: Visual comparison of our results and other state-of-the-art methods on blurry images from RED4[27] and DPDD[2] datasets. Our method achieves sharper details. Zoom in for better visualization.

4.2 Comparisons cross Degradation Types

Compared Methods. In this section, we evaluate our method with the state-of-the-art methods of image restoration. Initially, we selected some restoration approaches that are tailored for individual restoration tasks, including deblurring, denoising, low-light image enhancement, and defocus. Those methods like MPR [37], NAFNet [5], which are CNN-based methods, and SwinIR [20] is a transformer-based method. Besides, we also compare the performance of our method with other all-in-one approaches, like TransWeather [29] and AirNet [16]. For a fair competition with the compared methods, we follow them to conduct experiments with default optimal hyperparameters and settings. During the experiments, all the comparative methods were trained and tested on multi-degraded datasets in order to explore and compare their performance in handling multi-degraded images.

Quantitative Comparison. In this comparison, we adopt three commonly used in image restoration reference-based image quality assessment metrics, PSNR, SSIM, and LPIPS [43] to evaluate the quality of restored images. Better results are indicated by higher values of PSNR and SSIM, while LPIPS is the opposite. During the test, we calculate the metrics not only on all testing degraded images

but also on each type of degradation according to the dataset type. As illustrated in Table 1, our method performs better than other methods in terms of the average metrics on all types of degraded data. Additionally, when classified by types of degradation, our method achieves the best or second-best results on different types of degraded images, suggesting that the performance improvement of our method can effectively distinguish and represent each type of degradation, rather than relying on the improvement on a single type of degradation. We also measured the computational costs of different models using a 3x256x256 image as input. It is evident that our method exhibits significantly reduced computational cost compared to AirNe t[16], MPR [37], and SwinIR [20]. While there is an increase in computational cost compared to NAFNet [5] and TransWeather [29], our method holds an advantage in handling multiple degradations.

Visual Comparison. We conduct a visual comparison of our method and several state-of-the-art techniques on the aforementioned types of degraded images. Figure4 presents a visual comparison against other image restoration methods on RED4 [27] and DPDD [2] datasets, and it demonstrates that our method is capable

(a) Input   **(b) SwinIR**   **(c) MPR**   **(d) NAFNet**

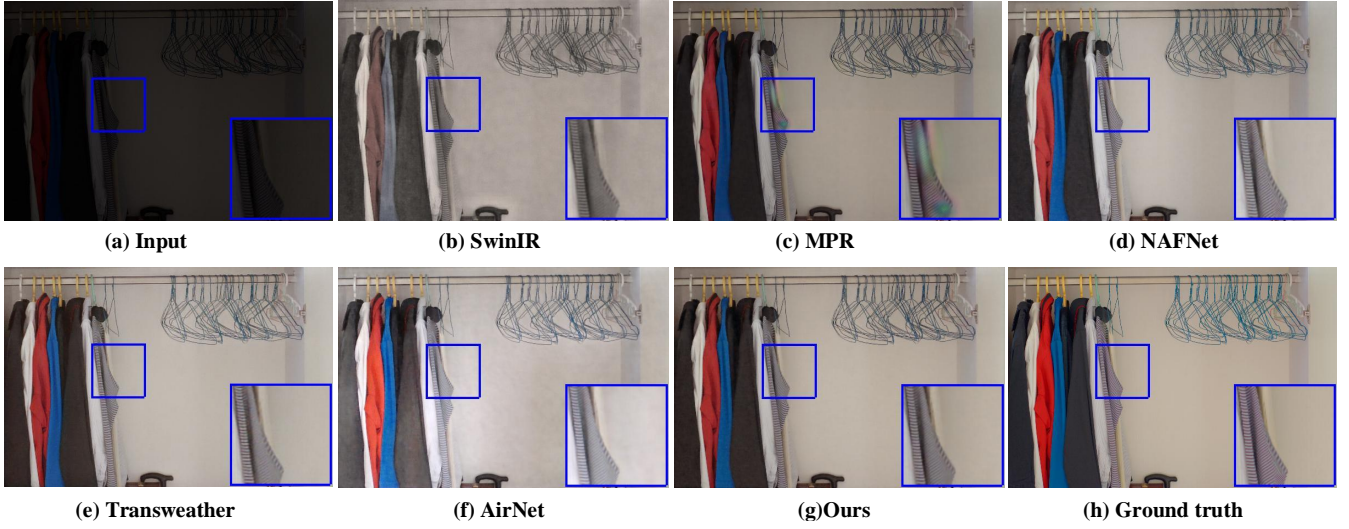**(e) Transweather**   **(f) AirNet**   **(g)Ours**   **(h) Ground truth**

Figure 5: Visual comparison of our results and other state-of-the-art methods on low-light images from the LOL dataset [31]. Our network effectively enhances the low-light image and generates harmonious tones. Zoom in for better visualization.

(a) Noisy   **(b) SwinIR**   **(c) MPR**   **(d) NAFNet**

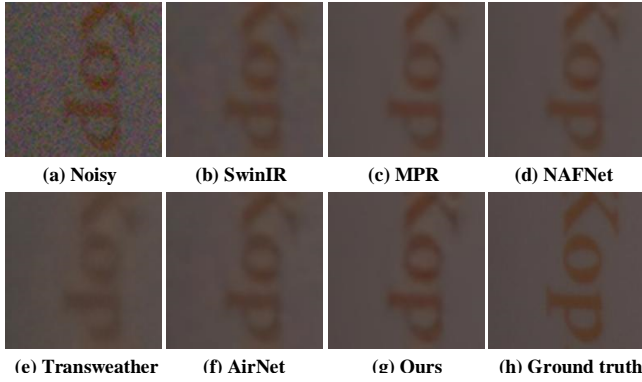**(e) Transweather**   **(f) AirNet**   **(g) Ours**   **(h) Ground truth**

Figure 6: Visual comparison of our results and other state-of-the-art methods on a noisy image from SIDD [1]. Our method produces results with less noise and sharper details. Zoom in for better visualization.

of restoring the intricate details of the image while the image content is corrupted seriously by blur. Meanwhile, the results on LOL [31] dataset are shown in Figure 5, our method effectively removes artifacts from low-light images when enhancing them, ensuring that the processed images have good visual quality. Figure 6 is the comparison of denoising results on SIDD [1] dataset, our approach can remove the noise in distorted images and recover a high-quality image. There are more results about the multi-degradation experiment, which can be found in the supplement material.

4.3 Comparison cross Degradation Levels

To evaluate our network's performance on distorted images with different degradation levels, we compare the denoising results of our method and other state-of-the-art denoising methods, including BM3D [9], DnCNN [40], IRCNN [41], DL [10], FFDNet [42], and

Table 2: Quantitative comparison of our method and other state-of-the-art methods on denoising dataset CBSD68 [26]. The best results are shown in boldface

Model	Average		σ=15		σ=25		σ=50	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
BM3D[9]	30.54	0.8505	33.52	0.9215	30.71	0.8672	27.38	0.7627
DnCNN[40]	31.03	0.8672	33.90	0.9290	31.24	0.8830	27.95	0.7896
IRCNN [41]	30.98	0.8669	33.87	0.9285	31.18	0.8824	27.88	0.7898
DL [10]	30.10	0.8440	33.25	0.9225	30.38	0.8679	26.68	0.7415
FFDNet[42]	31.01	0.8666	33.87	0.9290	31.21	0.8821	27.96	0.7887
MPR[37]	31.15	0.8747	34.01	0.9334	31.34	0.8892	28.10	0.8014
Ours	**31.21**	**0.8783**	**34.05**	**0.9357**	**31.40**	**0.8929**	**28.17**	**0.8064**

MPR [37]. The noisy images are synthetic by adding White Gaussian noise with different levels (*i.e.* $\sigma = 15, 25, 50$) to clear images. The comparison result is reported in Table 2. Our approach exhibits superior performance compared to other denoising methods and delivers optimal results in quantitative evaluations. This highlights the versatility of our method in restoring images degraded at varying levels.

4.4 Ablation Study

Impact of Supervision on Degradation Types. To investigate the impact of degradation category supervision on the restoration network, we retrain some separate networks. As listed in Figure 7, the backbone model is a restoration network that does not utilize any degradation information. And backbone+CN is a restoration network that uses a classification network trained with degradation-type supervision and utilizes the classification results. The comparative results in Figure 7 show that the performance of networks that use degradation representation for processing multiple degradations has been significantly improved. However, even with the supervision of degradation types in the classification network, it

cannot surpass our proposed method of using hierarchical degradation representation for restoration. This implies that degradation types are not mutually exclusive in multi-degraded image restoration, and exploring the relationships between degradations and leveraging them for restoration is highly necessary.
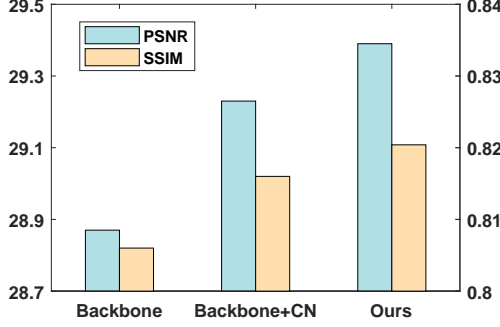

Figure 7: Performance comparison of models with/without the supervision of degradation information on multi-degraded images.

Impact of Hierarchical Representation Layers. We evaluate the performance of networks with different layers in hierarchical degradation representation. The visual results are shown in Figure 8. From the results shown in Table 3, it can be observed that as the number of layers in the tree structure increases, the restoration performance also improves with the enhancement of representation capacity. When using the two-stage strategy, the degradation representation provided by the fixed DRN allows the training to focus on RN. As a result, the performance of the restoration network is further improved.

Table 3: Ablation experiments for representation layers and FTB on multi-degraded datasets.

Number of layers	1 layer	2 layers	3 layers	4 layers
PSNR	29.15	29.23	29.26	29.28
SSIM	0.8163	0.8172	0.8179	0.8185
Models	w/o FTB	w/o DSLN	w/o GM	full
PSNR	28.87	29.17	29.22	29.39
SSIM	0.8060	0.8149	0.8166	0.8204

Impact of FTB. We also conduct an ablation study to validate the effectiveness of FTB in our network. We remove FTB, DSLN, and GM separately in our model, to observe the effect of each component. The results are shown in Table 3. Removing either DLSN or GM leads to a performance drop in our network. However, as both have the ability to refine degraded image features by utilizing degradation information, the absence of one does not hinder the other from being effective. FTB's ability to modulate image features is maximized when both are used.

4.5 Visualization

We randomly sample degraded images from the training set and extract their degradation features through DRN. By utilizing t-SNE on the degradation representations, we can evaluate the performance

of degradation representation in low-dimension embedding space. The visualized result is presented in Figure 8. From the illustration, we can see that the hierarchical degradation representations generated by our method exhibit greater inter-class separation and intra-class compactness compared to the results of AirNet [16]. This result is consistent with our previous conjecture that tighter constraints on samples with similar degradation types can help improve the performance of the restoration network.
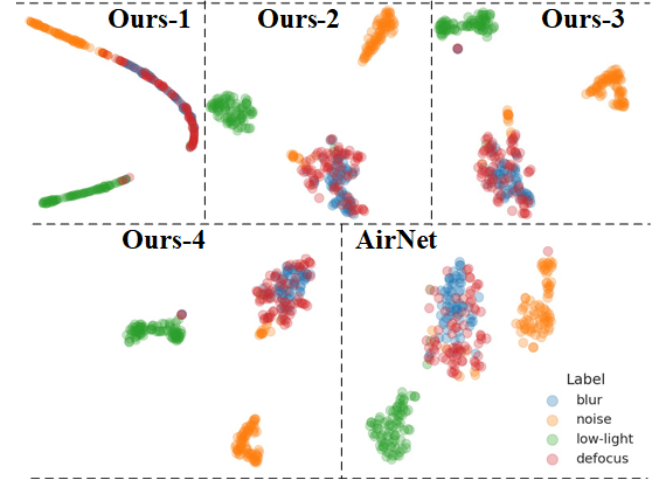

Figure 8: Visual comparison of degradation representation by t-SNE. The number in "Ours-number" represents the representation layers of AMIRNet.

5 CONCLUSION

We present an all-in-one multi-degradation image restoration network, AMIRNet, an efficient and practical method to handle multi-degraded images. By progressively constructing a hierarchical degradation representation, AMIRNet can effectively model the similarity and differences among degradations. To overcome the domain gaps and sufficiently utilize the degradation information, AMIRNet uses FTB to refine features with the guidance of the degradation representation. Extensive experimental results show that AMIRNet outperforms other state-of-the-art methods in multi-degradation restoration. The effectiveness of our approach is also verified by ablation studies and visualization.

REFERENCES

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. 2018. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1692–1700.

[2] Abdullah Abuolaim and Michael S Brown. 2020. Defocus deblurring using dual-pixel data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*. Springer, 111–126.

[3] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. 2019. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*. 7354–7362.

[4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3291–3300.

[5] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. 2022. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*. Springer, 17–33.

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

[7] Liang Chen, Jiawei Zhang, Songnan Lin, Faming Fang, and Jimmy S Ren. 2021. Blind deblurring for saturated images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6308–6316.

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

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing* 16, 8 (2007), 2080–2095.

[10] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. 2019. A general decoupled learning framework for parameterized image operators. *IEEE transactions on pattern analysis and machine intelligence* 43, 1 (2019), 33–47.

[11] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

[12] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. 2017. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2319–2328.

[13] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. 2014. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2862–2869.

[14] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. 2019. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1712–1722.

[15] Zhe Hu, Sunghyun Cho, Jue Wang, and Ming-Hsuan Yang. 2014. Deblurring low-light images with light streaks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3382–3389.

[16] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. 2022. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17452–17462.

[17] Dasong Li, Yi Zhang, Ka Chun Cheung, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. 2022. Learning Degradation Representations for Image Deblurring. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*. Springer, 736–753.

[18] Fengjun Li, Xin Feng, Fanglin Chen, Guangming Lu, and Wenjie Pei. 2022. Learning Generalizable Latent Representations for Novel Degradations in Super-Resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1797–1807.

[19] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. 2020. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3175–3185.

[20] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1833–1844.

[21] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi. 2013. Single-image noise level estimation for blind denoising. *IEEE transactions on image processing* 22, 12 (2013), 5226–5237.

[22] Ilya Loshchilov and Frank Hutter. 2017. Fixing weight decay regularization in adam. (2017).

[23] Yu Luo, Yong Xu, and Hui Ji. 2015. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE international conference on computer vision*. 3397–3405.

[24] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. 2016. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing* 26, 2 (2016), 1004–1016.

[25] Julien Mairal, Michael Elad, and Guillermo Sapiro. 2007. Sparse representation for color image restoration. *IEEE Transactions on image processing* 17, 1 (2007), 53–69.

[26] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vol. 2. IEEE, 416–423.

[27] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. 2019. NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study. In *CVPR Workshops*.

[28] Jinshan Pan, Zhouchen Lin, Zhixun Su, and Ming-Hsuan Yang. 2016. Robust kernel estimation with outliers handling for image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2800–2808.

[29] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2353–2363.

[30] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. 2021. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10581–10590.

[31] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference*.

[32] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. 2017. Multi-channel weighted nuclear norm minimization for real color image denoising. In *Proceedings of the IEEE international conference on computer vision*. 1096–1104.

[33] Li Xu, Shicheng Zheng, and Jiaya Jia. 2013. Unnatural l0 sparse representation for natural image deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1107–1114.

[34] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. 2019. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1751–1760.

[35] Qingsen Yan, Dong Gong, Pei Wang, Zhen Zhang, Yanning Zhang, and Javen Qinfeng Shi. 2023. SharpFormer: Learning Local Feature Preserving Global Representations for Image Deblurring. *IEEE Transactions on Image Processing* (2023).

[36] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. 2017. Image deblurring via extreme channels prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4003–4011.

[37] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14821–14831.

[38] Cheng Zhang, Shaolin Su, Yu Zhu, Qingsen Yan, Jinqiu Sun, and Yanning Zhang. 2022. Exploring and evaluating image restoration potential in dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2067–2076.

[39] Cheng Zhang, Qingsen Yan, Yu Zhu, Xianjun Li, Jinqiu Sun, and Yanning Zhang. 2020. Attention-based network for low-light image enhancement. In *2020 IEEE international conference on multimedia and expo (ICME)*. IEEE, 1–6.

[40] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing* 26, 7 (2017), 3142–3155.

[41] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. 2017. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3929–3938.

[42] Kai Zhang, Wangmeng Zuo, and Lei Zhang. 2018. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing* 27, 9 (2018), 4608–4622.

[43] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.