Visual Recognition-Driven Image Restoration for Multiple Degradation with Intrinsic Semantics Recovery

Zizheng Yang* Jie Huang* Jiahao Chang Man Zhou
Hu Yu Jinghao Zhang Feng Zhao†
University of Science and Technology of China
{yzz6000, hj0117, changjh, manman, yuhu520, jhaozhang}@mail.ustc.edu.cn
fzhao956@ustc.edu.cn

Abstract

Deep image recognition models suffer a significant performance drop when applied to low-quality images since they are trained on high-quality images. Although many studies have investigated to solve the issue through image restoration or domain adaptation, the former focuses on visual quality rather than recognition quality, while the latter requires semantic annotations for task-specific training. In this paper, to address more practical scenarios, we propose a Visual Recognition-Driven Image Restoration network for multiple degradation, dubbed VRD-IR, to recover high-quality images from various unknown corruption types from the perspective of visual recognition within one model. Concretely, we harmonize the semantic representations of diverse degraded images into a unified space in a dynamic manner, and then optimize them towards intrinsic semantics recovery. Moreover, a prior-ascribing optimization strategy is introduced to encourage VRD-IR to couple with various downstream recognition tasks better. Our VRD-IR is corruption- and recognition-agnostic, and can be inserted into various recognition tasks directly as an image enhancement module. Extensive experiments on multiple image distortions demonstrate that our VRD-IR surpasses existing image restoration methods and show superior performance on diverse high-level tasks, including classification, detection, and person re-identification.

1. Introduction

We have witnessed the remarkable success made by deep learning in image recognition tasks in recent years, such as classification [25, 33, 66], detection [23, 46, 62, 68], and segmentation [9, 51]. However, most of these approaches leverage the public datasets with high-quality images (*e.g.*, ImageNet [64], CoCo [47]) for training, and they suffer a significant drop when applied to low-quality images (*e.g.*, hazy, rainy, and noisy), since the statistical properties of pixels are ruined by image degradation [75].

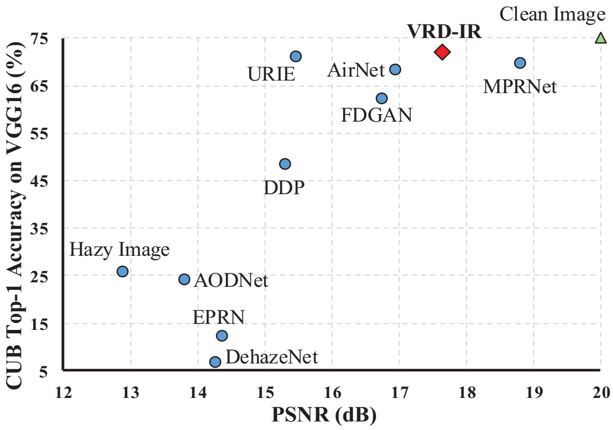

Figure 1. Illustration of visual quality and recognition quality using different dehazing methods on hazy CUB [73]. The top-1 accuracy is evaluated by VGG16 [66] pre-trained on clean CUB. Our method, VRD-IR, is shown in bold. As we can see, higher visual quality doesn't mean higher recognition quality.

An intuitive approach to tackle this issue is to restore the distorted images first, and then feed them into the succeeding recognition models. With this line, various image enhancement methods have been developed to improve the human visual quality of corrupted images [15, 90]. However, the visual quality and the recognition quality of an image differ fundamentally from one another. As shown in Fig. 1, the restored image with higher visual effect cannot guarantee satisfactory performance on downstream high-level vision tasks [57, 67, 81].

Another feasible solution is to encourage the recognition models to learn corruption-invariant feature representations, which can be applied to low-quality images directly without image recovery. For that purpose, numerous datasets have

[*] Equal contribution
[†] Corresponding Author

been created [26,50,80]. One common method is to narrow the distribution distance between low- and high- quality images in feature space [32,37,67,81]. While promising, most of these methods neglect the fact that the adverse impacts of different degradation are quite different on semantic level. On the other hand, they either assume that the task-specific annotations is available during training, or just could handle a single corruption/recognition task only, which hinders the timely adaptation to changing external environment and adjustment to flexible high-level tasks in real-world.

In this paper, we propose a visual recognition-driven image restoration (VRD-IR) for multiple degradation, to recover the recognition-friendly high-quality image from its given degraded version without knowing the specific degradation type and downstream high-level tasks. We first harmonize the semantic features suffered from different degradation into a unified representation spaces, and then optimize them towards semantic recovery. Specifically, we design a model paradigm: Intrinsic Semantics Enhancement (ISE), which can restore different degraded semantic representations in a dynamic manner. It consists of a Degradation Normalization and Compensation (DNC) module for mapping different degraded features to a degradation-invariant space, and a Fourier Guided Modulation (FGM) for guiding the feature enhancement from the statistical properties in amplitude spectrum. For better perception of different semantics, a prior-ascribing optimization strategy is proposed. A semantic aware decoder (SAD) is first pre-trained on both low- and high- quality images with the objective to reconstruct the high-quality image from the corresponding semantic features. To make full use of semantic information and provide good guidance for ISE, a similarity ranking loss is enforced during the pre-training of SAD. Then, we fix the pre-trained SAD and force the ISE to improve the quality of images reconstructed by SAD through enhancing the degraded semantic representations. In this way, we encourage the ISE to modulate the degraded input features from the perspective of machine vision.

Moreover, the proposed VRD-IR can be plugged into pre-trained recognition models directly as a data enhancement module. Compared with feature distillation-based methods that require task-specific annotations for training, our VRD-IR enjoys better flexibility and practicality.

We summarize our main contributions as follows:

- To the best of our knowledge, VRD-IR is the first attempt towards a pure universal image restoration framework for high-level vision. As the VRD-IR can be integrated with various recognition models directly, it is more practical in real world scenario.
- Considering the adverse impacts of different degradation in semantics, we design an Intrinsic Semantic Enhancement (ISE) module to modulate the degraded semantic representation in a dynamic manner.
- A prior-ascribing optimization strategy is proposed to endow VRD-IR with capability to perceive degradation effects on semantic level. Guided by this, our ISE can modulate degraded features from the perspective of machine vision.
- We verify the effectiveness of our framework on diverse high-level vision tasks, including classification, detection, and person re-identification. Experiments results show the superiority of our method in recognition tasks under multiple degradation.

2. Related Works

2.1. Image Restoration

Image Restoration for Single Degradation. Image Restoration methods for single degradation (IRSD) focus on recovering clean images from those suffer from a specific degradation. SRCNN [15] is the first work to introduce convolution neural network (CNN) to image restoration. After that, numerous image restoration methods emerge and have achieved great success, such as super-resolution (SR) [36, 45, 91, 92], denoising [1, 3, 6, 22, 55, 90], dehazing [2, 10, 17, 18, 24, 38, 40, 48, 86], deraining [19, 43, 52, 58, 61, 79, 87], and deblurring [5, 11, 21, 34, 35, 63]. Some works also try to handle multiple kinds of degradation with one designed network, such as DnCNN [88], MPRNet [85], and HINet [8]. Recently, transformer [72] is also applied in image restoration tasks [44, 76, 84]. However, these methods cannot handle multiple degradation simultaneously, which limits their application in real-world.

Image Restoration for Multiple Degradation. Recently, image restoration for multiple degradation (IRMD) methods are proposed to handle different types of degradation simultaneously with a single network. All-in-One Network [42] is proposed to remove different weather degradation with one network. IPT [7] achieve all-in-one image restoration with multiple heads, multiple tails, and a shared transformer-based backbone. AirNet [39] distinguishes different image degradation in latent space with contrastive learning. TransWeather [70] uses transformer to handle various weather degradation. Although these methods enjoy better flexibility, they consider less on recognition quality.

2.2. High-level Vision in Degraded Scenarios.

With the development of the autonomous driving and surveillance analysis, robust visual recognition under different distorted scenes has garnered increasing attention in recent years. Some works [13, 14, 71] have revealed that the performance of high-level task based on CNN will decrease when facing corrupted images. DDP [75] recon-

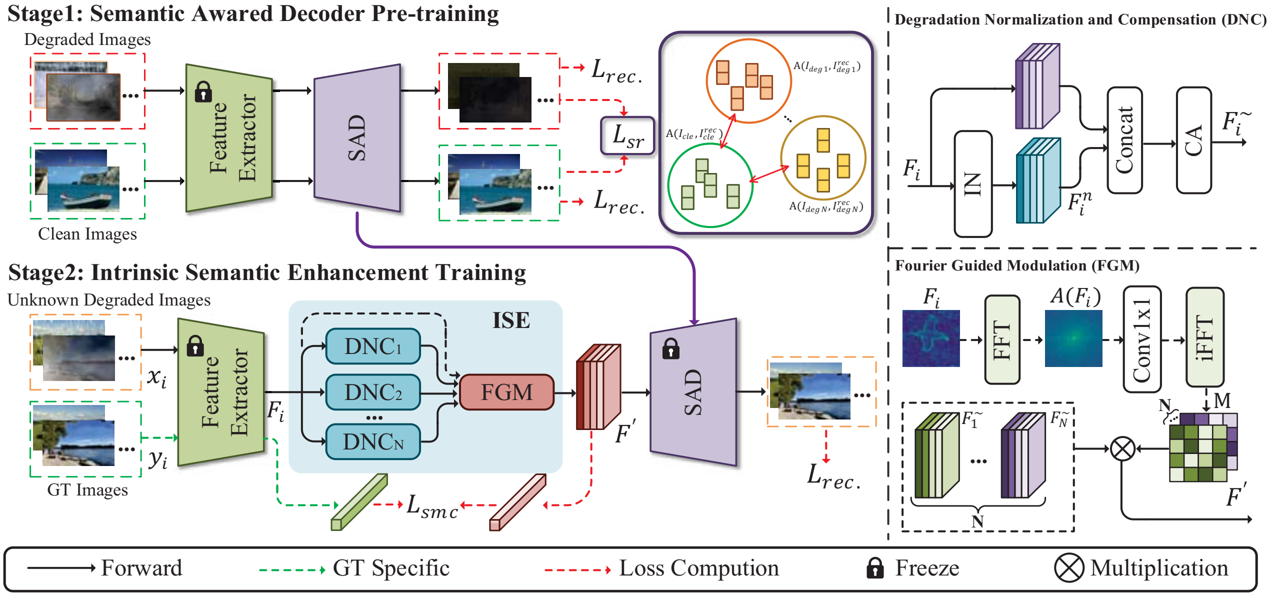

Figure 2. Architecture of the proposed VRD-IR, consisting of an intrinsic semantic enhancement (ISE) module and a semantic aware decoder (SAD). Given an unknown degradation image, we first get the semantic representation, and then send the obtained feature to ISE to enhance semantics in a dynamic manner with its statistical properties on amplitude spectrum. Finally, we feed the restored features to SAD to reconstruct the recognition-friendly image. The entire training is based on our designed prior-ascribing optimization strategy, which is instantiated as a two-stage training: the SAD is pre-trained at the first stage; then we fix SAD and train ISE at the second stage.

structs the shallow features of low-quality images for degraded images classification. QualNet [32] try to achieve quality-agnostic feature learning with invertible decoder. URIE [67] aims to handle various image degradation with a task-specific recognition model. SFDUNet [81] attempts to reconstruct high-quality features from low-quality images by self-feature distillation and uncertainty modeling. FIFO [37] encourage the segmentation model to learn fog-invariant features with a fog-pass filter module. However, these methods require task-specific annotations for training. SACC [74] directly recover recognition-friendly normal images from low-light images with a self-supervised pre-text task. Despite this, it only considers single degradation.

3. Methods

In this section, we will give a detailed introduction to our proposed VRD-IR. We first overview the whole pipeline of our VRD-IR in Sec. 3.1, and then introduce the Intrinsic Semantic Enhancement (ISE) and the prior-ascribing optimization strategy in Sec. 3.2 and Sec. 3.3, respectively.

3.1. Overview

We denote the training set with N kinds of degradation as $D = \{(x_{i,j}, y_{i,j})_{j=1}^{M_i}\}_{i=1}^{N}$, where $x_{i,j}$ is the j-th image from i-th degradation, $y_{i,j}$ is the ground-truth image of $x_{i,j}$, and M_i represents the number of samples in i-th degradation. Given a degradation-agnostic image x_{deg}, we aim to recover the recognition-friendly high-quality image I_{hq}.

As illustrated in Fig. 2, the proposed VRD-IR comprises an intrinsic semantic enhancement module f_{ISE} and a semantic-awared decoder f_{SAD}. The entire training process is based on our prior-ascribing scheme, which is instantiated as a two-stage training scheme.

For the degraded image x_{deg}, we can get the degraded feature F_{deg} through a fix-weight feature extractor. Then, the F_{deg} is mapped to a degradation-invariant space through the multi-branch Degradation Normalization and Compensation (DNC) modules. A Fourier Guided Modulation (FGM) is introduced to guide the restoration of F_{deg} based on its statistical characteristics, which is detailed in Sec. 3.2.

Moreover, a prior-ascribing training strategy is introduced. Specifically, the semantic aware decoder f_{SAD} is first pre-trained on both clean and degraded images with a similarity ranking loss to perceive different semantics. And then, we fixed the pre-trained f_{SAD}, and train our feature enhancement module f_{ISE} with a semantic maximum loss, which is described in detail in Sec. 3.3.

As for the feature extractor f_{ext}, we utilize the shallow layers of classical network (*e.g.*, VGG16 [66], ResNet50 [25]) pre-trained on large-scale datasets (*e.g.*, ImageNet [64], CoCo [47]) to provide a representation space rich in semantic information. So, we can get the set $D' = \{(F_{i,j}, y_{i,j})_{j=1}^{M_i}\}_{i=1}^{N}$, where $F_{i,j}$ is the semantic feature of $x_{i,j}$ extracted by f_{ext}. Note that $F_{i,j}$ and $x_{i,j}$ are one-to-one correspondence since f_{ext} is fixed-weight during training.

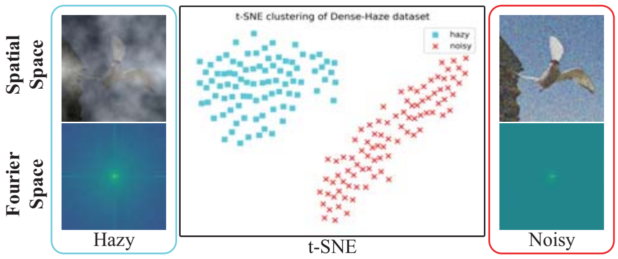

Figure 3. The t-SNE map of the amplitude spectrum of semantic features suffered from different degradation.

3.2. Intrinsic Semantic Enhancement

As shown in Fig. 2, we implement our ISE module with two main parts: the Degradation Normalization and Compensation (DNC) that is designed for mapping various degradation features to degradation invariant feature space, and the Fourier Guided Modulation (FGM) which is proposed to integrate the outputs of multi-branch DNC based on the statistical properties of the input feature.

Degradation Normalization and Compensation. Since the effects of various degradation differ significantly on semantic level, we introduce Instance Normalization (IN) to align different distorted features. Given a degraded semantic feature F_i that belongs to i-th degradation ($i \in \{1, ..., N\}$), we employ IN to map it to the degradation invariant feature space:

$$F_i^n = IN(x_i) = \gamma \frac{F_i - \mu(F_i)}{\sigma(F_i)} + \beta, \tag{1}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation, and $\gamma, \beta \in \mathbb{R}^c$ are the parameters learned from data. IN could filter out instance-specific style information [31]. Since each degradation can be viewed as a kind of style, different degradation features can be aligned through IN by decreasing the discrepancy between them in style.

However, IN inevitably discards discriminative information [29, 31, 56] for semantic restoration and image reconstruction. To alleviate this problem, we integrate the original features before normalization and those go through IN to guarantee the integrity of information [49]. For the feature F_i^n, we concatenate it with its counterpart before normalization (*i.e.*, F_i) in the channel dimension. We utilize a SE-like [28] channel attention integrate them adaptively and get the output $\widetilde{F_i}$, as shown in Fig. 2.

Fourier Guided Modulation. Although the DNC align different representation coarsely, we need to modulate them further. Based on the observation that, the distribution of different degradation in amplitude spectrum via deep Fourier transform differs significantly (see Fig. 3), we propose a Fourier guided modulation (FGM) to utilize the statistical properties of input to guide the feature adjustment.

Given the degraded input F_i, we first transform it from the spatial domain to its frequency domain Fourier transformation $\mathcal{F}(F_i)$:

$$\mathcal{F}(F_i)(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} F_i(h,w)e^{-j2\pi(\frac{h}{H}u+\frac{w}{W}v)}, \tag{2}$$

The $\mathcal{F}(F_i)$ is denoted as $\mathcal{F}(F_i) = \mathcal{R}(F_i) + j\mathcal{I}(F_i)$, where $\mathcal{R}(F_i)$ and $\mathcal{I}(F_i)$ are the real and imaginary part of $\mathcal{F}(F_i)$.

As the distorted effects caused by degradation mainly manifest in the amplitude spectrum [83], we can get the amplitude spectrum $\mathcal{A}(F_i)$ by:

$$\mathcal{A}(F_i)(u,v) = [\mathcal{R}^2(F_i)(u,v) + \mathcal{I}^2(F_i)(u,v)]^{1/2}. \tag{3}$$

We extract the degraded information from $\mathcal{A}(F_i)$ using a 1×1 convolution, which is formulated by:

$$\mathcal{A}^{'}(F_i)(u,v) = \mathcal{A}(F_i)(u,v) * kernel_1, \tag{4}$$

where $*$ denotes the convolution operator, and $kernel_1$ means the 1×1 convolution kernel. We can also get the $\mathcal{F}^{'}(F_i)$ corresponding to $\mathcal{A}^{'}(F_i)$.

Finally, we transform the frequency domain feature $\mathcal{F}^{'}(F_i)$ back to the spatial domain feature by inverse Fourier transform, and get the guidance map M through a convolution followed by softmax.

Joint Training for ISE. In order to handle diverse degradation more flexibly, we propose a multi-branch DNC architecture [94], as shown in Fig. 2(a). Specifically, we prepare a DNC for each degradation, and pre-train each DNC with degradation-specific data independently.

After that, we fine-tune the whole ISE with both DNCs and the FGM. Given an unknown degraded feature F, we send it to the multi-branch DNCs and get a set of features $[\widetilde{F_1}, ..., \widetilde{F_N}]$. Then, we fuse them with the guidance map M generated by FGM: $F^{'} = M \otimes [\widetilde{F_1}, ..., \widetilde{F_N}]$.

To further maintain the semantic property of the integrated feature $F^{'}$, we propose a Semantic Maximum Constraint for feature distillation between $F^{'}$ and the clean features $F_{cle}^{'}$. We first extract the most semantic parts through max pooling, then we utilize cosine similarity to measure their semantic distance:

$$\mathcal{L}_{smc} = 1 - \frac{f^{'}}{\|f^{'}\|_2} \cdot \frac{f_{cle}^{'}}{\|f_{cle}^{'}\|_2}, \tag{5}$$

where $\|\cdot\|_2$ is the L_2 normalization, $f^{'}$ and $f_{cle}^{'}$ represent the max pooling results of $F^{'}$ and $F_{cle}^{'}$, respectively. Unlike the common-used L_1 or L_2 regularization that treat each pixel equally, L_{smc} maintains the semantic consistency between the restored feature $F^{'}$ and the corresponding clean feature $F_{cle}^{'}$, which encourages ISE to modulate degraded features from the perspective of semantics rather than visualization.

The reconstruction loss in image-level is also computed to regularize the restoration of ISE:

$$\mathcal{L}_{rec.} = \|f_{SAD}(f_{ISE}(F)) - y\|_1, \tag{6}$$

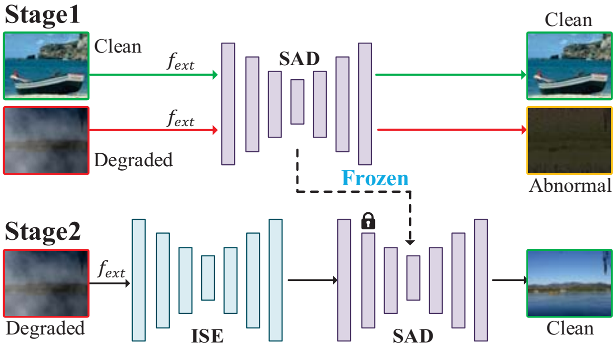

Figure 4. Illustration of the prior-ascribing training strategy. The SAD is first pre-trained on both degraded and clean images. Then, we fix the SAD and train ISE with degraded image.

where $\|\cdot\|_1$ denotes the L_1 regularization, y means the clean image of F, and f_{SAD} is a pre-trained decoder in advance, which will be described in Sec. 3.3. So the whole training loss for ISE is defined as:

$$\mathcal{L}_{ISE} = \mathcal{L}_{smc} + \mathcal{L}_{rec.}, \tag{7}$$

Note that we assign a DNC for each degradation and pretrain them independently. This is because the distribution information of various degradation is quite different, using different DNCs in different degradation can better preserve degradation intrinsic properties for semantic recovery. And then, we fine-tune the ISE with a statistical properties-based guidance generated by FGM to aggregate multiple parallel DNCs dynamically, which are input dependent. It not only endows the VRD-IR with more representation power than methods [39, 42] based on static network, but also encourages the ISE to explore correlations between the procedures of correcting different degradation implicitly.

3.3. Prior-Ascribing Optimization Strategy

Our goal is to recover recognition-friendly images with acceptable visual quality. The challenge lies in narrowing the semantic gap between low- and high- quality images. To tackle this issue, we propose a prior-ascribing optimization strategy, which is instantiated as a two-stage training sheme. A semantic-awared decoder (SAD) is first pre-trained as a pretext task to perceive the semantics at the first stage, then we fix the SAD and encourage the ISE to bridge the semantic gap between degraded and clean images at the second stage, as shown in Fig. 4.

Specifically, given a clean image I_{cle} and a degraded one I_{deg}, we can get their semantic feature F_{cle}, F_{deg} through the fixed feature extractor. Then we feed them into the semantic aware decoder f_{SAD} to self-reconstruct the input images I^{rec}_{cle} and I^{rec}_{deg}, respectively: $I^{rec}_{cle} = f_{SAD}(F_{cle})$, and $I^{rec}_{deg} = f_{SAD}(F_{deg})$. A reconstruction loss is computed to ensure the reconstruction ability of the SAD:

$$\mathcal{L}^{s}_{rec.} = \|I^{rec}_{cle} - I_{cle}\|_1 + \lambda \cdot \left\|I^{rec}_{deg} - I_{deg}\right\|_1 . \tag{8}$$

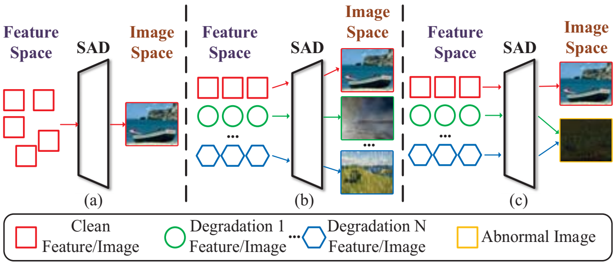

Figure 5. Illustration of different training strategy for SAD. (a) training on clean images only, (b) training on both clean and degraded images with objective to reconstruct both of them accurately, (c) training with our proposed similarity ranking loss.

where λ is the weighting parameter.

In fact, we hope the reconstructed images from clean semantic features are much better than those from degraded semantic features, which makes the SAD sensitive to the input semantic features. To this end, we design a similarity ranking loss. For the clean images, we calculate the relative similarity between I_{cle} and I^{rec}_{cle}:

$$A(I_{cle}, I^{rec}_{cle}) = \frac{I_{cle} \cdot I^{rec}_{cle}}{\|I_{cle}\|_2 \cdot \|I^{rec}_{cle}\|_2} \tag{9}$$

$A(\cdot)$ denotes the relative similarity calculation function between two images. Similarly, we perform the same technique to the degraded images I_{deg} and I^{rec}_{deg} and get the similarity $A(I_{deg}, I^{rec}_{deg})$.

After that, we employ a ranking metric [27] to measure the distance between two similarity and construct a similarity ranking loss based on it:

$$\mathcal{L}_{sr} = max\left(0, -(A(I_{cle}, I^{rec}_{cle}) - A(I_{deg}, I^{rec}_{deg})) + m\right), \tag{10}$$

where m represents a threshold value.

The total optimization objective for pre-training SAD is defined as:

$$\mathcal{L}_{SAD} = \mathcal{L}^{s}_{rec.} + \mathcal{L}_{sr}, \tag{11}$$

We endow the SAD with ability to perceive both clean and degraded semantic representation through pre-training, which can serve as a semantic prior to supervise the ISE training. After pre-training, we fix the pre-trained SAD, and force the ISE to improve the quality of images reconstructed by SAD through modulating the degraded features with the training objective in Eq. 7.

The similarity ranking loss L_{sr} guarantees the f_{SAD} to be monotonic, which is critical to the SAD pre-training, as shown in Fig. 5(c). We also show two possible alternative pre-training schemes in Fig. 5(a) that trains f_{SAD} on clean images only, and Fig. 5(b) which trains f_{SAD} with both clean and degraded images with the objective to reconstruct all of them accurately. However, they either cannot ensure the sensitivity of f_{SAD} to degraded semantics, or requires larger number of parameter and a complex structure for f_{SAD}. We will analyze this in details in the **Appendix**.

Table 1. Performance comparisons with state-of-the-art IRSD (*i.e.*, image restoration for single degradation) and IRMD (*i.e.*, image restoration for multiple degradation) approaches for image classification on CUB dataset among three different degradation. "Top-1 V" and "Top-1 R" refer to the Top-1 Accuracy (%) on pre-trained VGG16 and ResNet50, respectively. The best results are marked as **bold** and the second ones are masked by underline. More comparison results can be found in **Appendix**.

Method	Dehazing Method					IRMD Method		
	DehazeNet [4]	AODNet [40]	EPRN [60]	FDGAN [16]	DDP [75]	MPRNet [85]	AirNet [39]	VRD-IR
Top-1 V (%)	6.67	24.01	12.05	62.04	48.26	69.59	68.19	**72.11**
Top-1 R (%)	17.24	42.06	25.93	74.23	63.02	78.15	76.81	**80.55**
PSNR/SSIM	14.29/0.5225	13.28/0.6415	14.39/0.6864	16.76/0.7545	15.32/0.7002	**18.83/0.8000**	16.97/0.7692	17.64/0.7790

(a) dehazing

Method	Denoising Method					IRMD Method		
	CBM3D [12]	DnCNN [88]	IRCNN [89]	FFDNet [90]	BRDNet [69]	MPRNet [85]	AirNet [39]	VRD-IR
Top-1 V (%)	20.16	24.48	26.98	22.69	25.05	25.18	27.77	**32.00**
Top-1 R (%)	25.41	38.92	43.37	37.51	41.96	42.51	45.72	**50.26**
PSNR/SSIM	22.67/0.5237	23.01/0.5474	25.52/0.7121	25.16/0.6982	26.28/0.7552	**26.74/0.7764**	26.42/0.7653	26.41/0.7669

(b) denosing

Method	Deraining Method					IRMD Method		
	DIDMDN [87]	UMRL [82]	SIRR [77]	MSPFN [30]	LPNet [20]	MPRNet [85]	AirNet [39]	VRD-IR
Top-1 V (%)	50.29	55.89	73.34	73.46	73.56	74.25	73.08	**74.68**
Top-1 R (%)	62.35	68.59	81.19	81.26	81.37	82.03	81.65	**82.21**
PSNR/SSIM	18.63/0.6845	20.80/0.7197	27.56/0.8655	27.68/0.8664	27.73/0.8692	**28.54/0.8844**	27.45/0.8693	27.41/0.8805

(c) deraining

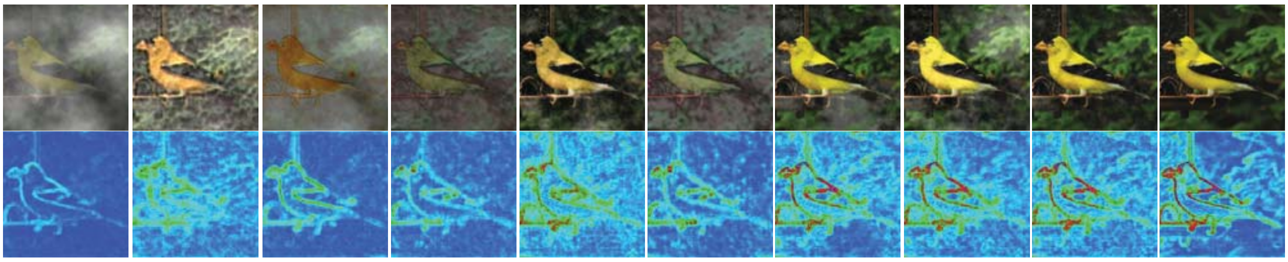
(a) Hazy (b) DehazeNet (c) AODNet (d) EPRN (e) FDGAN (f) DDP (g) MPRNet (h) AirNet **(i) VRD-IR** (j) GT

Figure 6. Qualitative results of different methods for dehazing on the CUB dataset. Our method is shown in bold.

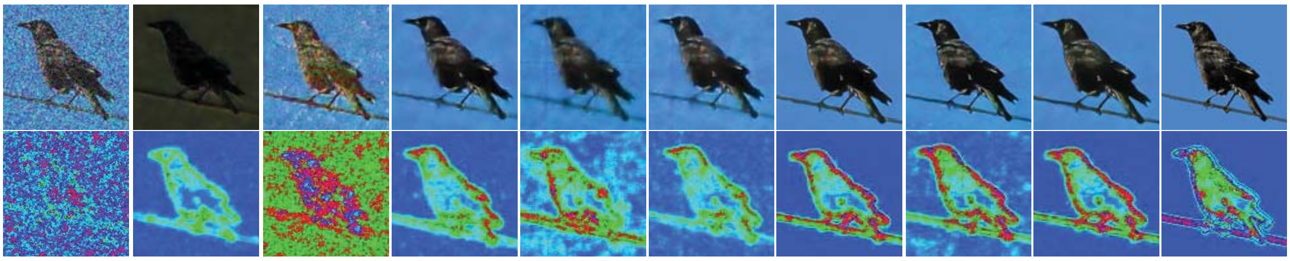
(a) Noisy (b) CBM3D (c) DnCNN (d) IRCNN (e) FFDNet (f) BRDNet (g) MPRNet (h) AirNet **(i) VRD-IR** (j) GT

Figure 7. Qualitative results of different methods for denoising on the CUB dataset. Our method is shown in bold.

4. Experiments

4.1. Implementation Details

To evaluate the effectiveness of our method, we test our method on three different recognition tasks, involving classification, detection, and person re-identification, across three types of degradation, hazy, noisy, and rainy.

Datasets. For training datasets, we follow the setting in [39] for fair comparison. To be specific, we use the combination of BSD400 [54] and WED [53] as training set for image denoising. For image deraining, we conduct experiments on Rain100L [78]. And for image dehazing, we utilize the RESIDE [41] as the training datasets. More details about datasets are provided in **Appendix**.

Training Details. The proposed VRD-IR is trained by Adam optimizer, where β_1 , β_2, and γ are set to 0.9, 0.999, and 0.5, respectively. The initial learning rate is set to 2×10^{-4} and reduced as the training epoch increases. The

Table 2. Performance comparisons with state-of-the-art IRSD and IRMD approaches for object detection on CrowdHuman dataset among three different degradation. ↑ means higher the better. The best results are marked as **bold** and the second ones are masked by underline. More comparison results can be found in **Appendix**.

Method	Dehazing Method					IRMD Method		
	DehazeNet [4]	AODNet [40]	EPRN [60]	FDGAN [16]	FFANet [59]	MPRNet [85]	AirNet [39]	VRD-IR
AP ↑	46.77	61.08	55.45	74.75	74.71	78.64	78.24	**79.33**
JI ↑	41.39	53.18	47.69	63.43	63.11	66.52	65.87	**67.58**
MR ↓	85.53	75.79	81.04	65.32	65.56	63.60	64.03	**63.21**
PSNR/SSIM	13.13/0.4935	12.13/0.5638	13.26/0.5784	17.27/0.7956	17.35/0.7986	**19.27/0.8569**	19.00/0.8427	18.25/0.8256

(a) dehazing

Method	Denoising Method					IRMD Method		
	CBM3D [12]	DnCNN [88]	IRCNN [89]	FFDNet [90]	BRDNet [69]	MPRNet [85]	AirNet [39]	VRD-IR
AP ↑	48.51	56.48	59.08	57.88	58.05	58.98	59.36	**59.80**
JI ↑	41.93	48.89	50.76	50.63	50.98	51.61	51.95	**52.57**
MR ↓	83.69	78.62	78.12	78.27	78.20	76.89	77.08	**75.89**
PSNR/SSIM	20.15/0.5426	22.59/0.5529	24.10/0.6739	23.94/0.6773	24.52/0.6874	**24.98/0.7221**	24.66/0.7126	24.69/0.7160

(b) denosing

Method	Deraining Method					IRMD Method		
	DIDMDN [87]	UMRL [82]	SIRR [77]	MSPFN [30]	LPNet [20]	MPRNet [85]	AirNet [39]	VRD-IR
AP ↑	74.44	76.36	78.05	78.21	78.28	78.49	77.76	**78.68**
JI ↑	63.10	64.49	66.26	66.54	66.68	66.81	66.29	**67.20**
MR ↓	67.89	66.51	64.47	63.99	63.79	63.73	64.53	**63.51**
PSNR/SSIM	22.72/0.7865	23.16/0.8089	26.21/0.8651	26.44/0.8720	26.48/0.8731	**27.41/0.8934**	26.61/0.8792	26.53/0.8886

(c) deraining

mini-batch is set to 40, and the images are resized, cropped to 128 × 128 with being flipped horizontally randomly. The λ in Eq. 8 and m in Eq. 11 is set to 0.1 and 1.0, respectively. The whole model is trained with one 3090Ti GPU.

4.2. Improvement on Image Classification.

In this section, we show the superiority of our VRD-IR by comparing with other image restoration methods on the most fundamental machine analysis: image classification. Following the DDP [75], we choose CUB [73] as our test dataset and synthesize the degraded images as [26]. We use VGG16 [66] and ResNet50 [25] pre-trained on clean CUB as the recognition models to evaluate images restored by different methods. To be fair, we compare the VRD-IR with both IRSD and IRMD methods , where IRMD methods are trained following the setting in [39], and can handle various degradation simultaneously. Unlike [32, 67, 81], we do not introduce any task-specific annotation during training.

Tab. 1 shows that our VRD-IR has better performance than compared restoration methods in two recognition models among three different degradation on the CUB test set. Ordinary restoration methods enhance degraded input images from the perspective of visual quality rather than high-level vision, thus leading to higher performance in PSNR/SSIM but unsatisfying results in accuracy. The performance improvement is more noticeable on denoising (see Tab. 1b), which is the most challenging tasks among three degradation restoration for machine vision. On image deraining where the advantage is not obvious, the VRD-IR still holds the lead of MPRNet by 0.43%/0.18% accuracy improvement (see Tab. 1c). Fig. 6 and Fig. 7 describe the qualitative results of different methods along with their feature maps on image dehazing and denoising, respectively.

Table 3. Performance (%) comparisons of different methods for person ReID on Market1501 dataset in dehazing and denoising. We show more comparison results in **Appendix**.

Category	Method	mAP (%)
Dehazing	FDGAN [16]	72.98
	FFANet [59]	73.74
	MPRNet [85]	75.12
	AirNet [39]	74.21
	VRD-IR	**75.83**
Denoising	FFDNet [90]	53.33
	BRDNet [69]	52.89
	MPRNet [85]	54.56
	AirNet [39]	54.45
	VRD-IR	**55.64**

4.3. Improvement on Object Detection

The VRD-IR can also benefit the object detection. To demonstrate this, we test different methods on CrowdHuman [65], a benchmark for human detection. RetinaNet [46] is employed as the downstream recognition model. The synthesis of degraded images for detection is the same as that for image classification.

As illustrated in Tab. 2, the VRD-IR outperforms all compared baseline networks in detection among all three types of degradation. Note that the VRD-IR for classification (in Tab. 1) and detection (in Tab. 2) are shared parameters. It further verifies that VRD-IR can achieve better practicality and generalization ability for high-level vision. Interestingly, Tab. 1 and Tab. 2 exhibit a similar phenomenon that, the restoration of diverse corruptions for visual quality and machine analysis differs significantly. For example, denoising can benefit PNSR better but fail in recognition tasks, while dehazing does the opposite. We conjecture that this is due to the various effects of diverse degradation on different frequency bands of the image.

4.4. Improvement on Person Re-Identification

We further conduct experiments on a fine-grained image retrieval task, person re-identification (ReID). We evaluate different methods on Market1501 [93]. The recognition baseline is the ResNet50 pre-trained on clean data. Tab. 3 shows that the VRD-IR attains the best performance.

4.5. Ablation Study

In this section, we perform comprehensive ablation studies to demonstrate the effectiveness of our designs in the proposed VRD-IR. Here, we conduct experiments on CUB classification to validate each component. More results of ablation studies are provided in **Appendix**.

Effectiveness of the ISE. To demonstrate the benefits of the proposed instrinsic semantic enhancement which consists of degradation normalization and compensation and Fourier guided modulation module, we design several variants as shown in Tab. 4. Among them, "VRD-IR w/o DNC" represents we replace DNC with a couple a convolution block which has the same number of parameter as DNC. We do a similar operation for "VRD-IR w/o FGM". "Baseline" means that both DNC and FGM are replaced.

As we can see, "VRD-IR w/o DNC" and "VRD-IR w/o FGM" outperform "Baseline" by **3.39%/1.76%** and **5.24%/3.48%** in terms of Top-1 Accuracy on pre-trained VGG16 in dehazing/denoising, respectively. With both two modules, "VRD-IR" achieves **72.11%/32.00%** in dehazing/denoising classification, which demonstrates that DNC and FGM are complementary and both vital to VRD-IR, joinly resulting in a superior performance.

Effectiveness of the Prior-Ascribing Optimization Strategy. As described in Tab. 5, the VRD-IR without PA has a significant performance drop on classification, where "VRD-IR w/o PA" means we train ISE and SAD in an end-to-end manner. Obviously, our prior-ascribing optimization strategy can help image recovery from the perspective of machine vision. We have shown three different training strategies for SAD in Fig. 5. We will further investigate the influence of them and show more results in **Appendix**.

Table 4. The ablation results of several variants of VRD-IR on CUB classification in dehazing and denoising. "Dehaze" and "Denoise" means Top-1 Accuracy (%) on pre-trained VGG16 in dehazing and denoising classification.

Model	DNC	FGM	Dehaze	Denoise
Baseline	×	×	61.18	25.17
VRD-IR w/o DNC	×	√	64.57	26.93
VRD-IR w/o FGM	√	×	66.42	28.65
VRD-IR	√	√	**72.11**	**32.00**

Table 5. Effectiveness of the prior-ascribing strategy in our VRD-IR on CUB. "Dehaze", "Denoise" and "Derain" means Top-1 Accuracy (%) on pre-trained VGG16 in dehazing, denoising and deraining classification.

Model	Dehaze	Denoise	Derain
VRD-IR w/o PA	67.55	25.14	73.23
VRD-IR w PA	**72.11**	**32.00**	**74.68**

Effectiveness of the Semantic Maximum Constraint. In order to better maintain semantic properties of modulated features, we introduce a semantic maximum constraint (SMC), which is implemented by calculating the cosine similarity of two features after max pooling. Previous methods adopt L1 or mean square error (MSE) [75] as their semantic loss function. In this section, we compare SMC with other possible solutions. As shown in Tab. 6, the SMC outperforms L1 by **3.15%/2.47%** in Dehaze/Denoise on classification, which also surpasses MSE by **3.28%/2.24%** in Dehaze/Denoise. Compared with L1 or MSE that treat each pixel in features equally, our SMC can better maintain the semantic consistency since it explores the most valuable semantic correspondence between two features.

Table 6. Effectiveness of the SMC in our VRD-IR.

Constraint	Dehaze	Denoise	Derain
No Constraint	63.51	23.79	71.06
L1	68.96	29.53	73.59
MSE	68.83	29.76	73.88
SMC	**72.11**	**32.00**	**74.68**

5. Conclusion

In this paper, we develop a Visual Recognition-Driven Image Restoration (VRD-IR) for multiple degradation, to recover degraded image from the perspective of high-level vision. It consists of a Intrinsic Semantic Enhancement (ISE) module and a Prior-Ascribing Optimization Strategy. Our VRD-IR can be plugged into existing recognition tasks as a image enhancement module. Extensive experiment on multiple degradation and diverse high-level tasks demonstrate the effectiveness of our method.

Acknowledgments. This work was supported by the JKW Research Funds under Grant 20-163-14-LZ-001-004-01, and the Anhui Provincial Natural Science Foundation under Grant 2108085UD12. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

[1] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3165–3173, 2019. 2

[2] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016. 2

[3] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11036–11045, 2019. 2

[4] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 6, 7

[5] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *European conference on computer vision*, pages 221–235. Springer, 2016. 2

[6] Ke-Chi Chang, Ren Wang, Hung-Jin Lin, Yu-Lun Liu, Chia-Ping Chen, Yu-Lin Chang, and Hwann-Tzong Chen. Learning camera-aware noise models. In *European Conference on Computer Vision*, pages 343–358. Springer, 2020. 2

[7] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021. 2

[8] 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. 2

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 1

[10] Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen Kuo. Pms-net: Robust haze removal based on patch map for single images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11681–11689, 2019. 2

[11] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 2

[12] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *2007 IEEE International Conference on Image Processing*, volume 1, pages I–313. IEEE, 2007. 6, 7

[13] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2016. 2

[14] Samuel Dodge and Lina Karam. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th international conference on computer communication and networks (ICCCN)*, pages 1–7. IEEE, 2017. 2

[15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 1, 2

[16] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10729–10736, 2020. 6, 7

[17] Raanan Fattal. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):1–9, 2008. 2

[18] Raanan Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1):1–14, 2014. 2

[19] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017. 2

[20] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight pyramid networks for image deraining. *IEEE transactions on neural networks and learning systems*, 31(6):1794–1807, 2019. 6, 7

[21] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3848–3856, 2019. 2

[22] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 2

[23] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1

[24] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 2

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 3, 7

[26] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 2, 7

[27] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 5

[28] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 4

[29] 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*, pages 6043–6052, 2022. 4

[30] 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. 6, 7

[31] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen, and Li Zhang. Style normalization and restitution for generalizable person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3143–3152, 2020. 4

[32] Insoo Kim, Seungju Han, Ji-won Baek, Seong-Jin Park, Jae-Joon Han, and Jinwoo Shin. Quality-agnostic image recognition via invertible decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12257–12266, 2021. 2, 3, 7

[33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 1

[34] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 2

[35] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019. 2

[36] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2

[37] Sohyun Lee, Taeyoung Son, and Suha Kwak. Fifo: Learning fog-invariant features for foggy scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18911–18921, 2022. 2, 3

[38] Boyun Li, Yuanbiao Gou, Jerry Zitao Liu, Hongyuan Zhu, Joey Tianyi Zhou, and Xi Peng. Zero-shot image dehazing. *IEEE Transactions on Image Processing*, 29:8457–8466, 2020. 2

[39] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022. 2, 5, 6, 7

[40] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 2, 6, 7

[41] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 6

[42] 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. 2, 5

[43] 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. 2

[44] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 2

[45] 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. 2

[46] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 1, 7

[47] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1, 3

[48] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7314–7323, 2019. 2

[49] Yajing Liu, Xinmei Tian, Ya Li, Zhiwei Xiong, and Feng Wu. Compact feature learning for multi-domain image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7193–7201, 2019. 4

[50] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 2

[51] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1

[52] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE international conference on computer vision*, pages 3397–3405, 2015. 2

[53] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality as-

sessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016. 6

[54] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. 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*, volume 2, pages 416–423. IEEE, 2001. 6

[55] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2502–2510, 2018. 2

[56] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 464–479, 2018. 4

[57] Yanting Pei, Yaping Huang, Qi Zou, Yuhang Lu, and Song Wang. Does haze removal help cnn-based image classification? In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 682–697, 2018. 1

[58] 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. 2

[59] 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. 7

[60] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019. 6, 7

[61] 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. 2

[62] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 1

[63] Wenqi Ren, Jiawei Zhang, Jinshan Pan, Sifei Liu, Jimmy S Ren, Junping Du, Xiaochun Cao, and Ming-Hsuan Yang. Deblurring dynamic scenes via spatially varying recurrent neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 44(8):3974–3987, 2021. 2

[64] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015. 1, 3

[65] Shuai Shao, Zijian Zhao, Boxun Li, Tete Xiao, Gang Yu, Xiangyu Zhang, and Jian Sun. Crowdhuman: A benchmark for detecting human in a crowd. *arXiv preprint arXiv:1805.00123*, 2018. 7

[66] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1, 3, 7

[67] Taeyoung Son, Juwon Kang, Namyup Kim, Sunghyun Cho, and Suha Kwak. Urie: Universal image enhancement for visual recognition in the wild. In *European Conference on Computer Vision*, pages 749–765. Springer, 2020. 1, 2, 3, 7

[68] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020. 1

[69] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 121:461–473, 2020. 6, 7

[70] 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. 2

[71] Igor Vasiljevic, Ayan Chakrabarti, and Gregory Shakhnarovich. Examining the impact of blur on recognition by convolutional networks. *arXiv preprint arXiv:1611.05760*, 2016. 2

[72] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2

[73] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 1, 7

[74] Wenjing Wang, Zhengbo Xu, Haofeng Huang, and Jiaying Liu. Self-aligned concave curve: Illumination enhancement for unsupervised adaptation. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2617–2626, 2022. 3

[75] Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, and Zhiwei Xiong. Deep degradation prior for low-quality image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11049–11058, 2020. 1, 2, 6, 7, 8

[76] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022. 2

[77] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3877–3886, 2019. 6, 7

[78] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(6):1377–1393, 2019. 6

[79] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection

and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017. 2

[80] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020. 2

[81] Zhou Yang, Weisheng Dong, Xin Li, Jinjian Wu, Leida Li, and Guangming Shi. Self-feature distillation with uncertainty modeling for degraded image recognition. 1, 2, 3, 7

[82] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8405–8414, 2019. 6, 7

[83] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. 4

[84] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 2

[85] 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. 2, 6, 7

[86] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018. 2

[87] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018. 2, 6, 7

[88] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 2, 6, 7

[89] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017. 6, 7

[90] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 1, 2, 6, 7

[91] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 2

[92] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 2

[93] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, 2015. 8

[94] Ziqi Zhou, Lei Qi, and Yinghuan Shi. Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In *European Conference on Computer Vision*, pages 420–436. Springer, 2022. 4