Multi-weather Image Restoration via Domain Translation

Prashant W. Patil[1], Sunil Gupta[1], Santu Rana[1], Svetha Venkatesh[1], and Subrahmanyam Murala[2]
[1]Deakin University, Geelong Warun Ponds Campus, VIC, Australia, [2] Trinity College Dublin, Ireland
prashant.patil@deakin.edu.au

Abstract

Weather degraded conditions such as rain, haze, snow, etc. may degrade the performance of most computer vision systems. Therefore, effective restoration of multi-weather degraded images is an essential prerequisite for successful functioning of such systems. The current multi-weather image restoration approaches utilize a model that is trained on a combined dataset consisting of individual images for rainy, snowy, and hazy weather degradations. These methods may face challenges when dealing with real-world situations where the images may have multiple, more intricate weather conditions. To address this issue, we propose a domain translation-based unified method for multi-weather image restoration. In this approach, the proposed network learns multiple weather degradations simultaneously, making it immune for real-world conditions. Specifically, we first propose an instance-level domain (weather) translation with multi-attentive feature learning approach to get different weather-degraded variants of the same scenario. Next, the original and translated images are used as input to the proposed novel multi-weather restoration network which utilizes a progressive multi-domain deformable alignment (PMDA) with cascaded multi-head attention (CMA). The proposed PMDA facilitates the restoration network to learn weather-invariant clues effectively. Further, PMDA and respective decoder features are merged via proposed CMA module for restoration. Extensive experimental results on synthetic and real-world hazy, rainy, and snowy image databases clearly demonstrate that our model outperforms the state-of-the-art multi-weather image restoration methods. Code is available at https://github.com/pwp1208/Domain_Translation_Multi-weather_Restoration.

1. Introduction

Fog, snow, rain, haze or their combintations often degrade the quality of images recorded for computer vision applications such as surveillance, traffic monitoring, and autonomous driving. These applications routinely involve sub-tasks such as optical flow estimation [36], object detection [8], depth estimation [27], *etc.*, which use algorithms or models expecting clean image as their input. Therefore, restoration of images degraded by one or many weather conditions is an important problem.

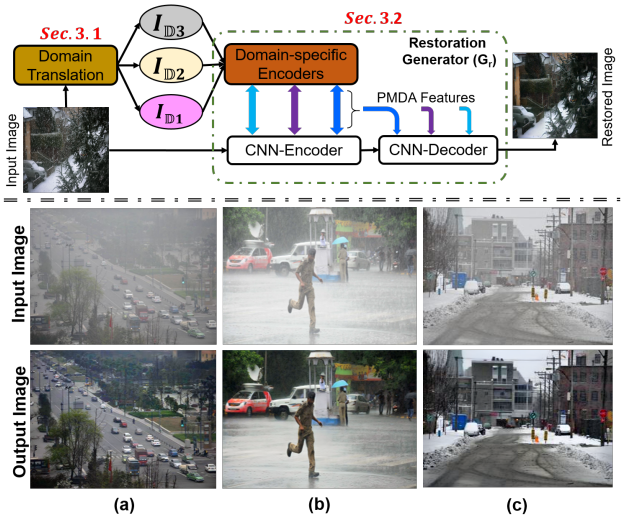

Figure 1. *Upper-part*: Thumbnail of the proposed domain translation based multi-weather image restoration approach. *Lower-part*: Sample restoration results: (a) de-hazing, (b) de-raining with veiling effect, and (c) de-snowing with veiling effect.

Although remarkable success has been achieved in restoration of particular domain such as image de-hazing, image de-raining, and image de-snowing, multi-weather degraded image restoration is still an open problem. Existing works handle multiple weather degradations *via* separate weather specific encoders [15], or separate training for each weather [1]. To handle multiple degradations with separate training-based models, typically these weather-specific models are put in a cascade, *e.g.,* haze removal followed by rain removal. But, such an approach may produce undesirable results as the erroneous output of any one of the earlier stages affects the overall result in an unpredictable manner. Also, in order to select the appropriate restoration model, prior information about the degradation must be available. To overcome these limitations, researchers intro-

duced various unified (*single trainable parameters*) models based on knowledge distillation [4], transformer [33], contrastive degradation guidance [12] and degradation removel with feature corrector [11] for multi-weather image restoration. These models are trained using a combined dataset that contains individual degraded images representing rainy, snowy, and hazy weather, *etc*. As real-world scenario may have more intricate weather conditions, which limits the performance of [4], [33], [11], [12] methods. Therefore, the restoration architecture should be able to perform significantly for any weather degradation.

To address the above challenges, we propose multi-weather image restoration framework where we first create different weather degradations of a degraded image and use them to learn a set of features that jointly represent all the degraded images. Since the common part across diverse weather degraded images is the clean image (*i.e.,* everything except weather degradations), our feature extractor learns to suppress the weather specific information and thus learns weather-invariant features. A practical advantage of this learning scheme is that our feature extractor can be also used for real data even though it is trained using synthetic data. This is because the synthetic and real-world images mainly differ in the weather-degradations, and since our feature extractor learns to suppress it, it can close the domain gap between synthetic and real-world data.

Following this idea, we propose a domain translation based unified approach for multi-weather image restoration. In order to achieve this, we first used a domain translation approach, where any of the degraded image can be translated to specific domain (*e.g., hazy, rain with veil, snow with veil*). We then propose a restoration architecture which takes domain translated and original degraded images as input to generate degradation free output. The proposed restoration network comprises of two components: namely a progressive multi-domain deformable alignment (PMDA) and cascaded multi-head attention (CMA) blocks. PMDA module perform inter domain offset feature alignment which facilitates the restoration network to learn weather-invariant representation effectively. Further, the CMA block is proposed to aggregate the respective PMDA and decoder features for effective restoration. Our main contributions are:

- A novel unified architecture for multi-weather image restoration. *It utilizes a single trainable model to restore all weather degradations*.
- A domain-translation with multi-attentive feature learning to achieve unified model's generalizability.
- A progressive multi-domain deformable alignment and cascaded multi-head attention modules for multi-weather image restoration.

Extensive experimental analysis on various weather-degraded synthetic and real-world datasets demonstrate that the proposed multi-weather restoration framework outperforms the existing state-of-the-art (SOTA) methods. An overview and sample results of the proposed method for real-world images are given in Figure 1.

2. Related Work

2.1. Single Weather Degradation Removal

De-hazing: Earlier research on single image de-hazing was directed towards hand-crafted based approaches [6], [32]. The emergence of deep learning networks brought advancements in degradation removal methods, and the researchers proposed various approaches towards the task of haze, rain or snow removal from images. Zhou *et al.* [40] proposed feedback spatial attention-based architecture with iterative extraction of the high-level information. With iterative process, the improper feature learning at the initial steps may cause an up-stretched error at the last stage and lead to ineffective results. Recently, Liu *et al.* [17] proposed three-stage visibility enhancement approach with multi-scale attention with multi-feature fusion module. The hazy-image and depth-map-based feature learning with cross-connection and residual channel attention module is proposed in [7] for the de-hazing task. However, getting an accurate depth map for dense hazy images is a challenging task which may produce unrealistic de-hazing results. Density and depth information is integrated with unpaired learning for image de-hazing [37].

De-raining: Deng *et al.* [5] considered rain removal and detail recovery as separate tasks and proposed two parallel sub-networks which make use of dilated convolutions and squeeze-and-excitation networks. Zamir *et al.* [38], proposed a multi-stage approach which progressively restores a degraded image by injecting supervision at each stage, while processing different scales of the input images.

De-snowing: Snow removal from single image is even more challenging task due to the non-transparency property of snow, which results in occluded background region and loss of details. In [39], the authors analyzed the difference between snow streaks and clear background edges, and applied a multi-guided filter to remove snowflakes. A three-stage hierarchical approach is proposed in [34], which combine image decomposition and dictionary learning. Even though these algorithms work effectively for specific weather degradation removal, performance may degrade in the presence of real-world weather degradation. Because along with rain or snow streaks, there is a strong veiling effect present, particularly in heavy rain or snow.

2.2. Multi-weather Degradation Removal

In 2019, the first attempt was made in [1], to handle single image de-hazing and the de-raining problem with specific weather training. A two-stage approach is proposed [14], in which the first stage estimates the rain streaks, trans-

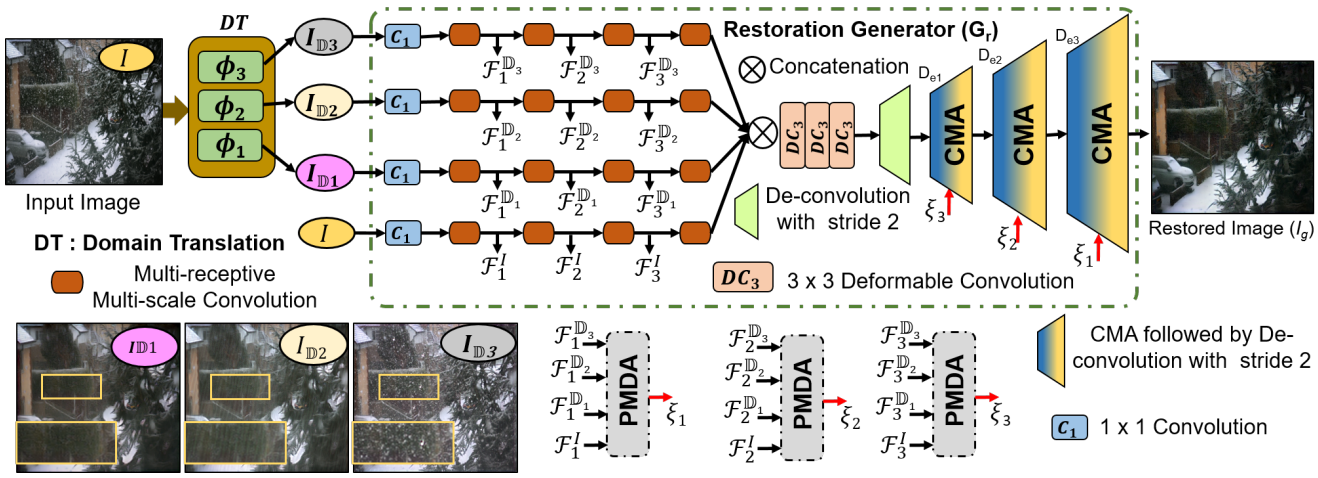

Figure 2. Overview of the proposed multi-weather image restoration algorithm. Initially, the weather degraded image is translated to different domains ($I_{\mathbb{D}1}$*: hazy,* $I_{\mathbb{D}2}$*: rain with veil and* $I_{\mathbb{D}3}$*: snow with veil*). Next, these translated and input images are processed through separate encoders and aligned using proposed progressive multi-domain feature alignment with cascaded multi-head attention module.

mission map, and atmospheric light from the input image. Li *et al.* [15] proposed a multi-weather image restoration approach where authors leverage neural architecture search for processing the features extracted from the task-specific encoders of rain with veil, rain-drop, and snow. Also, the selection of particular encoder while dealing with real-world weather degraded images is one of the limitation of [15]. Similar to heavy rain, heavy snow will also show a strong veiling effect in the real scene. The first attempt was made by [2] to handle snow and veiling effects simultaneously. They used the modified partial convolution and the differentiable dark channel prior layer. Chen *et al.* [3] proposed a dual-tree wavelet transform based approach with complex wavelet loss. In [38], authors proposed a multi-stage architecture that progressively learns restoration functions from degraded images. They handle image de-raining, de-blurring, and de-noising degradations *via* specific training. Recently, a contrastive-based degraded encoder and degradation-guided restoration network is proposed in [12] to handle image de-nosing, de-raining, and de-hazing tasks. Valanarasu *et al.* [33] proposed a transformer-based approach for rain with veil, snow and rain-drop removal. They also introduced learnable weather-type embeddings to handle different weather degradation effectively. Chen [4] propsoed two-stage *i.e.* knowledge collation and knowledge examination with multi-teacher and student architecture for adverse weather removal. Kulkarni *et al.* [11] proposed dual stream (appearing degradation and veiling effect removal) architecture for multi-weather image restoration. These [4] (28M) [15] (41M), [33] (31M) methods achieved superior performance but at the cost of increased computational complexity.

2.3. Domain Translation

Since our proposed restoration model is built upon domain translation, we briefly review it in this section. With a recent deep learning advancement, many algorithms are proposed for domain transfer task like structure-guided arbitrary [18], semantic context-aware [16], *etc.* In [10], authors proposed a memory-guided unsupervised approach to perform diverse translation between two visual domains *e.g., Sunny to night, rainy to sunny*. They have used separate content and style encoders to learn content and style features from input and target domains. Inspired by this, we have proposed an architecture for instance-level image-to-image translation between original degraded input to required domains. To the best of author's knowledge, this is the first weather-to-weather translation approach for restoration tasks.

3. Proposed Framework

In this work, a solution to handle multiple weather degradations is proposed with a unified network without any pre-trained weights or domain specific knowledge.

Overall Pipeline: Overview of the proposed framework is shown in Figure 2. Given a weather degraded image (I), we first aim to learn a weather invariant representation of the image making the restoration task for multi-weather scenario easier. To learn a weather-invariant representation for I, we first propose the domain translation (DT) of given image I into its different weather variants. For example, we take any weather degraded image I and translate it into a corresponding hazy, rain with veil and snow with veil images. We refer to each weather condition as a *domain*, denoting them as hazy ($\mathbb{D}_1$), rain with veil ($\mathbb{D}_2$), and snow with veil ($\mathbb{D}_3$), *etc.* In using domain translation, our main

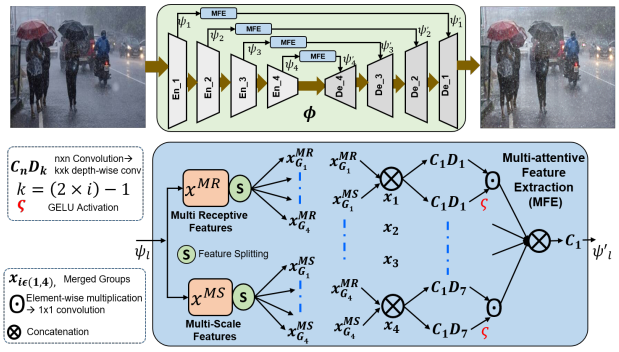

Figure 3. Overview of the proposed domain translation architecture with multi-attentive feature extraction (MFE) module.

idea is to create different weather degradations of the same image and use them to train a feature extractor to learn a common set of features that simultaneously represent all the degraded images. Since the common part across diverse weather degraded images is the clean image (everything except weather degradations), our feature extractor learns to suppress the weather specific information and thus learning weather-invariant features. Another practical advantage is that this feature extractor can be used for real data even though it is learnt using synthetic data. This is because only difference between the synthetic generated and real-world images lies in the weather-degradations and since our feature extractor learns to suppress it, the domain gap between synthetic and real-world data is closed.

We propose a restoration generator (G_r) where these weather degraded images (I, $I_{\mathbb{D}1}$, $I_{\mathbb{D}2}$ and $I_{\mathbb{D}3}$) are fed as input. The proposed G_r encodes each of the inputs via multi-receptive and multi-scale convolution block based encoders (see in Figure 2) in order learn to domain-relevant features (see $F_l^{\mathbb{D}_1}$, $F_l^{\mathbb{D}_2}$, $F_l^{\mathbb{D}_3}$, $F_{l\epsilon(1,3)}^{I}$ in Figure 2). Except first encoder, every encoder down-samples the input features. At last encoder layer, these separately learned features are aggregated and forwarded to the decoder via successive deformable convolutions (DC_3) for final reconstruction. These extracted features are then processed through successive decoder layers to have final restored image. While reconstruction, the independently learned encoder features processed via proposed progressive multi-domain feature alignment (PMDA) are forwarded to the respective decoder through skip connection. Further, the PMDA features and respective decoder features are merged effectively through proposed cascaded multi-head attention (CMA) to provide the restored image (I_g). Summarizing the overall pipeline of the proposed framework as: 1) domain translation (*refer Section 3.1*), and 2) Multi-weather image restoration (*refer Section 3.2*). Detailed exposition of each step of the proposed pipeline is provided in following subsections.

Figure 4. Visual results of the proposed DT network.

3.1. Domain-translation Architecture

The proposed DT approach takes weather degraded image (I) as input and translates it to other weather domains $I_{\mathbb{D}_i \in (1,2,3)}$. We have used U-Net [28] architecture as generator to perform translation of any weather degraded image over the required domain. This architecture mainly comprises of four encoders and four decoders (*see DT in Figure 2*). The encoder block is defined as: $\{En_{l\_l\times f};\ [l \in (1,4),\ f = 16]\}$. Similar to this, the decoder block is defined as: $\{De_{l\_l\times f};\ [l \in (4,1),\ f = 16]\}$, where, l and $(l \times f)$ represent encoder or decoder level and number of filters in encoder or decoder respectively.

Multi-attentive Feature Extraction: The input image (I) is given to the encoder which will give features $\left(\psi_{l\epsilon(1,4)}\right)$. *While giving a skip connection from the respective encoder to the decoder, we have processed these features via the proposed multi-attentive feature extraction (MFE) module (see MFE block in DT of 2 for more details).* In MFE, the respective l^{th} encoder-level features (ψ_l) are processed as:

$$x_G^{MR} = \left\langle C_{d\epsilon(1,2,3)}^{3}\left(\psi_l\right)\right\rangle ; x_G^{MS} = \left\langle C_1^{k\epsilon(1,3,5)}\left(\psi_l\right)\right\rangle \quad (1)$$

where, x_G^{MR} and x_G^{MS} are multi-receptive and multi-scale features, C_d^k is convolution with $k \times k$ kernel and d dilation rate, $\langle\cdot\rangle$ is concatenation followed by convolution with 1×1 kernel (*refer Figure 3 for more details*). Required contextual and spatial information is extracted using dilated and multi-scale convolutions through large receptive fields. Here, we form four different groups by splitting the features into groups of $x_{G_{i\epsilon(1,4)}}^{MR}$ and $x_{G_{i\epsilon(1,4)}}^{MS}$ containing equal number of channels in each group and merge respective $x_{G_i}^{MR}$ and $x_{G_i}^{MS}$ to form $x_{i\epsilon(1,4)}$. Further, these merged features $x_{i\epsilon(1,4)}$ are processed via multi-scale attention as : $C_1D_k(x_i) \odot \varsigma(C_1D_k(x_i))$, where ς is GeLu activation, $\odot$ is dot product, $k = (2 \times i) - 1$ (see Figure 3). This splitting and merging of multi-receptive and multi-scale features through multi-scale attention help the network for robust feature representation learning. Sample visual results for domain translation network are shown in Figure 4. From Figure 4, it is clear that the proposed DT network is able to generate three different degradation ($I_{\mathbb{D}1}$, $I_{\mathbb{D}2}$, $I_{\mathbb{D}3}$) irrespective of input degradation. With this intuition, the three outputs of the DT architecture ($I_{\mathbb{D}1}$, $I_{\mathbb{D}2}$, $I_{\mathbb{D}3}$) are used along with original weather degraded image (I) as input to the proposed restoration architecture.

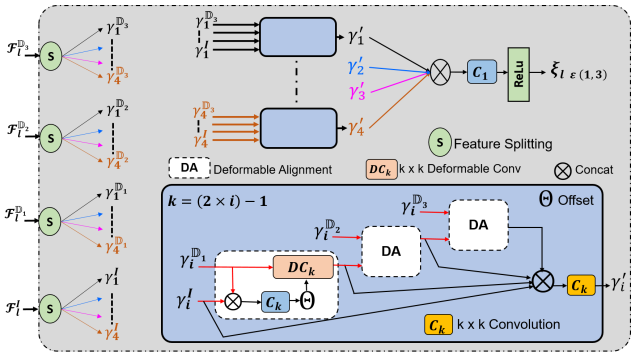

Figure 5. Overview of the proposed progressive multi-domain deformable alignment module for effective feature aggregation.

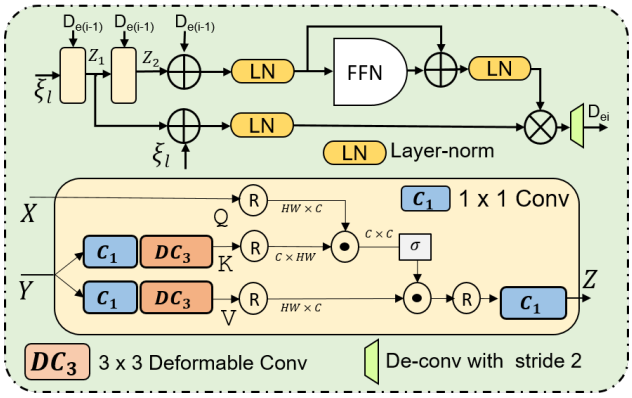

Figure 6. Overview of the proposed cascaded multi-head attention module for feature merging.

3.2. Multi-weather Image Restoration Architecture

The proposed restoration framework is shown in Figure 2 which comprises two major modules: 1) Progressive Multi-domain Deformable Alignment, and 2) Cascaded Multi-head Attention module. Details are given below. **Progressive Multi-domain Deformable Alignment:** As various weather degradations have different representation (*refer DT images of Figure 4*), the unified architecture urges alignment of domain (weather) translated features for effective multi-weather restoration. To achieve effective feature alignment, we propose a progressive multi-domain deformable alignment module. The proposed PMDA facilitates the network to learn weather-invariant representation effectively. The independently learned features from all the translated domains *i.e.* $F_l^{\mathbb{D}_1}, F_l^{\mathbb{D}_2}, F_l^{\mathbb{D}_3}$ and input image F_l^I of l^{th} encoder layer are forwarded to the proposed PMDA module (*refer the Figure 5 for more details*). Initially, every domain features are split into four groups *i.e.* γ_i^{domain}, $i \in (1,4)$, $domain \in (\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3, I)$. Further, we process two input domains progressively in proposed inter-domain offset feature alignment (*see* ▭ *in Figure 5*). Here, in the inter domain offset feature alignment, inputs from any two domains are used to generate the offsets for actual deformable convolution layer (*see DA in Figure 5*). Initially, the $\mathbb{D}_1$ domain features are aligned with input image features. Because, the rainy and snowy degradations are more complex as compared to hazy degradation. All the aligned features of all the respective i_{th} group γ_i^{domain} are finally merged to generate γ_i'. All the groups are processed in similar way through inter-domain offset feature alignment (see ▭ in Figure 5) and concatenation followed by a `convolution` $\rightarrow$ `Relu` layer as:

$$\gamma_l =< \gamma_1',\ \gamma_2',\ \gamma_3',\ \gamma_4' > \tag{2}$$

where, l is l^{th} encoder layer, $< \cdot >$ is concatenation followed by 1×1 convolution and ReLu operation. In the progressive inter-domain offset feature alignment for every group γ_i, the offset generation and deformable alignment is done with the offset convolution and deformable convolution respectively with different kernel size $k = (2 \times i) - 1$ *i.e.* for $i = 4$ the kernel size is $k = 7$ for both offset convolution (C_k) and deformable convolution (DC_k) (*see DA in Figure 5*). Overall response of progressive inter-domain feature alignment (*see* ▭ *in Figure 5*) for i^{th} group as:

$$\gamma_i' =< \gamma_i^I,\ \chi,\ \varphi(\chi, \gamma_i^{\mathbb{D}_2}),\ \varphi(\varphi(\chi, \gamma_i^{\mathbb{D}_2}), \gamma_i^{\mathbb{D}_3}) > \tag{3}$$

where, $\chi = \varphi(\gamma_i^I, \gamma_i^{\mathbb{D}_1})$, φ is the progressive inter-domain feature alignment. This PMDA module helps for disentanglement of task-relevant data of restoration and task-irrelevant data of various degradation. Therefore, these PMDA features are then forwarded to the respective decoder level via skip connection. **Cascaded Multi-head Attention:** To effectively merge the decoder features with the respective processed encoder features via PMDA module, we propose a cascaded multi-head attention (CMA). The multi-head attention proposed here consists of deformable convolutions to capture maximum receptive fields. Like transformer layers, the proposed CMA has the structure as `attention`$\rightarrow$`layer-norm`$\rightarrow$`feed forward network`. So the proposed CMA is able to capture the long term dependencies along-with the high receptive fields due to deformable attention. We use the term cascaded as we are using two successive multi-head attentions (MHA) (*see* ▭ *in Figure 6*) with two inputs from previous decoder $\left(D_{e(i-1)}\right)$ and respective PMDA level feature (ξ_l). This cascading allows to capture global dependencies effectively. The output of MHA is given as:

$$Z = \sigma(Q \cdot K^T) \cdot V \tag{4}$$

where, $Q = X$, $K = DC_3(C_1(Y))$, $V = DC_3(C_1(Y))$, and X, Y are the inputs to MHA. So the input to first MHA are $D_{e(i-1)}$ and ξ_l and second MHA are Z_1 and $D_{e(i-1)}$ (*see Figure 6*). To further process these features, we pass the response of first and second MHA through `layer-norm` (LN) with respect to ξ_l and $D_{e(i-1)}$ respectively. So the responses of both the LN are given as:

$$LN_1 = LN(Z_2 + D_{e(i-1)}); \quad LN_2 = LN(Z_1 + \xi_l) \quad (5)$$

Further, the output of first LN is processed through `feed forward network` (FFN) followed by a LN. Finally, the output of the CMA (D_{ei}) is given as:

$$D_{ei} = \langle LN_2, LN(FFN(LN_1) + LN_1) \rangle \quad (6)$$

where, $< \cdot >$ is concatenation operation.

4. Training Settings

The proposed restoration network is trained using synthetic data with original and domain translated images as input to get restored image under adversarial settings similar to [9], [23], [22]. To get domain transalted images, domain translation network is trained separately for each required domain like $\mathbb{D}_1$: hazy, $\mathbb{D}_2$: rain with veil and $\mathbb{D}_3$: snow with veil under adversarial settings [9]. The initial learning rate is set to 2×10^{-4} which is reduced by factor of 0.1 as the losses plateau. The networks are trained with PyTorch library on NVIDIA DGX Tesla V100 32 GB GPU. While training the networks, the discriminator is same as in [9].

4.1. Domain Translation Losses

We train the proposed DT network separately (see ϕ_1, ϕ_2, ϕ_3 in Figure 2) for generating each weather degraded image ($I_{\mathbb{D}1}, I_{\mathbb{D}2}, I_{\mathbb{D}3}$). While training of the network to generate particular weather degraded image (*e.g.,* $I_{\mathbb{D}2}$), any of the synthetically degraded images (I) is provided as input. The loss is calculated between the domain translated image (*e.g.,* $I_{\mathbb{D}2}$) and the respective target (*e.g.,* I_{Rv}). We denote by I_H, I_{Rv}, I_{Sv} are the synthetically degraded target images for $I_{\mathbb{D}1}, I_{\mathbb{D}2}, I_{\mathbb{D}3}$ domains respectively. Using these notations, the loss is calculated between the image pairs ($I_{\mathbb{D}1}, I_H$), ($I_{\mathbb{D}2}, I_{Rv}$), and ($I_{\mathbb{D}3}, I_{Sv}$) for generation of $I_{\mathbb{D}1}$, $I_{\mathbb{D}2}$ and $I_{\mathbb{D}3}$ images respectively. The adversarial learning is adopted for DT network training. Along with $\mathbb{L}_1$ loss, adversarial loss ($\mathbb{L}_A$) for DT network is calculated. To guide the network for textural and structural information, the perceptual loss ($\mathbb{L}_V$) is calculated between the translated and original target image by passing them through the pre-trained VGG19 model [30]. Further, the contrastive loss ($\mathbb{L}_c$) [35] in a common latent feature space is used to maximize and minimize the difference between $(I_{\mathbb{D}_{i\epsilon(1,2,3)}}, I)$ and $(I_{\mathbb{D}_{i\epsilon(1,2,3)}}, I_{\mathbb{T}_{i\epsilon(H,Rv,Sv)}})$ respectively. Also, the style loss ($\mathbb{L}_{S_d}$) with mean and standard deviation are considered. The total loss ($\mathbb{L}_d$) is:

$$\mathbb{L}_d(\mathbb{D}_i, \mathbb{T}_i) = \lambda_1 \mathbb{L}_1 + \lambda_2 \mathbb{L}_A + \lambda_3 \mathbb{L}_V + \lambda_4 \mathbb{L}_c + \lambda_5 \mathbb{L}_s \quad (7)$$

where, the weights $\lambda_{k\epsilon(1,5)}$ for each loss are set empirically.

4.2. Image Restoration Losses

Along with domain-translated images ($I_{\mathbb{D}_{i\epsilon(1,2,3)}}$), the original degraded image (I) is used as input to get restored image (I_g). The proposed restoration network is required to generate the restored image similar to the clean image (I_c) in adversarial manner. The $\mathbb{L}_1$ loss is used to optimize the network for better reconstruction. The adversarial loss ($\mathbb{L}_A$) is the min-max problem between generator and discriminator, respectively. Along with this loss, the perceptual loss ($\mathbb{L}_V$) [30] between (I_g, I_c), contrastive loss ($\mathbb{L}_c$) [35] between (I, I_g, I_c) are considered. So, the total loss ($\mathbb{L}_r$) is:

$$\mathbb{L}_r(I_g, I_c) = \lambda_1 \mathbb{L}_1 + \lambda_2 \mathbb{L}_A + \lambda_3 \mathbb{L}_V + \lambda_4 \mathbb{L}_c \quad (8)$$

where, the weights $\lambda_{k\epsilon(1,4)}$ for each loss are set empirically. (*loss functions are explained in supplementary material*).

5. Results and Discussion

5.1. Datasets and Evaluation Metrics

Synthetic Objective Testing Set (SOTS) [13]: 1000 indoor and outdoor images are provided for testing purpose. From training set, 5000 images are considered for training purpose. We have used outdoor images for result analysis.
Outdoor-Rain Database (ORD) [14]: This database images are degraded by rain and fog. 9000 and 750 images are provided for training and testing purpose respectively.
Comprehensive Snow Database (CSD) [3] : 8000 and 2000 images are provided for training and testing purpose respectively with combination of snow and fog degradation.
Real-world Databases: The image restoration algorithm is said to be domain generalised if it restores the synthetic as well as real-world degraded images efficiently. To scrutinize this, we have used three different real-world degraded image datasets named as Real-world Task-driven Testing Set (RTTS) [13], Snow realistic [19] and Rain In Driving (RID) [31] for experimental analysis .

We have considered all synthetic databases training sets in combination for training of the proposed restoration network. For testing purpose, the respective testing sets and real-world datasets are used. The reference based evaluation metrics includes average Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). Average Naturalness Image Quality Evaluator (NIQE) [21], Entropy and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [20] evaluation metrics are used in case of real-world weather degraded image restoration. We have used the released source code and provided trained models by respective authors to get the results on synthetic and real-world datasets.

5.2. Real-world Result Evaluation

The RTTS [13], Snow realistic [19] and RID [31] datasets are considered for real-world experimental analysis. The NIQE (↓), entropy (↑) and BRISQUE (↓) parameters are used for real-world quantitative analysis as provided in [29] (↓: *lower is better*, ↑: *higher is better*). The qualitative and quantitative analysis on real-world degraded images are provided in Figure 7 and Table 1 respectively. Proposed method de-hazing results follow the depth criteria. Whereas, UMVR, KD and TW methods are unable to

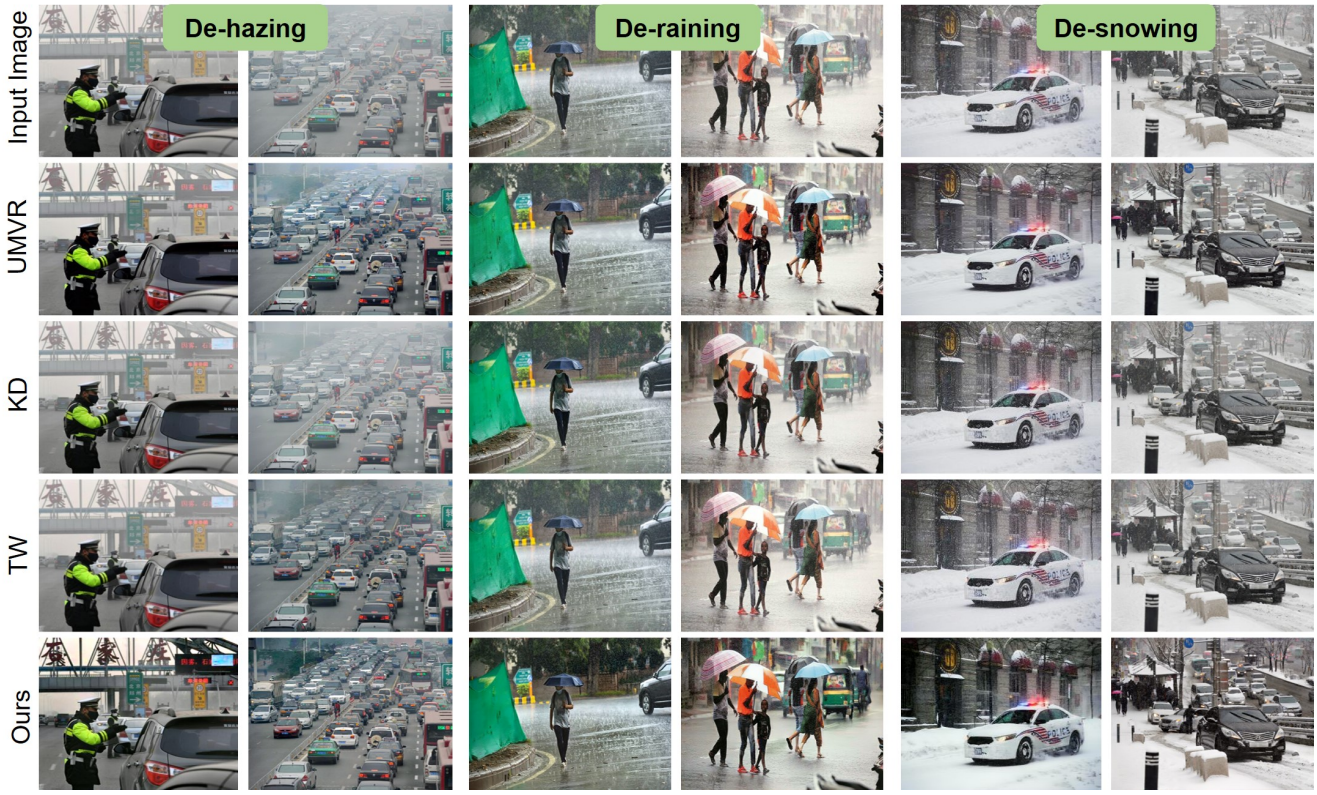

Figure 7. Qualitative analysis of the proposed and existing methods: UMVR [11], KD [4], TW [33] for multi-weather image restoration.

Table 1. Subjective results on real-world weather degraded images for de-hazing, de-raining with veil and snow with veil removal.

Database	Methods	NIQE (↓)	Entropy (↑)	BRISQUE (↓)
RTTS	UMVR [11]	5.009	7.221	28.625
	KD [4]	4.996	7.297	26.837
	TW [33]	5.703	7.263	29.874
	Ours	**4.859**	**7.505**	**24.761**
RID	UMVR [11]	**6.604**	7.486	24.296
	KD [4]	6.943	7.459	24.841
	TW [33]	7.496	7.393	24.165
	Ours	7.625	**7.492**	**23.931**
Snow realistic	UMVR [11]	4.296	7.354	23.761
	KD [4]	6.643	7.459	24.841
	TW [33]	5.628	7.331	25.377
	Ours	**4.196**	**7.572**	**21.884**

remove haze in depth (*see column 1 and 2 from Figure 7*). In presence of heavy rain, the baseline methods are not able to remove the rain effect completely due to strong veiling effect. Whereas, the proposed method can remove the rain with veiling effect significantly (*see column 3 and 4 from Figure 7*). In case of heavy snow, the effect of veiling effect is still present in the restored results of existing methods. However, the veiling effect is reduced substantially with the proposed restoration method (*see column 5 and 6 from Figure 7*). From the visual results, it is observed that the proposed method restores plausible results as compared to existing SOTA methods.

Table 2. Quantitative results analysis for de-hazing (SOTS), de-raining with veil (ORD) and snow with veil (CSD) removal in terms of average PSNR/SSIM.

Methods	SOTS	CSD	ORD
UMVR [11]	33.41/0.980	28.65/0.900	22.99/0.830
KD [4]	34.64/0.985	31.35/**0.950**	29.05/0.916
TW [33]	32.45/0.955	29.76/0.940	27.96/0.950
Ours	**36.26/0.987**	**32.95**/0.942	**31.24/0.951**

5.3. Synthetic Result Evaluation

We compared the proposed approach with SOTA unified restoration methods: UMVR [11], KD [4], TW [33] on SOTS-outdoor, CSD and outdoor-rain databases in terms of average PSNR, and SSIM. The quantitative analysis is provided in the Table 2 (*see supplementary material for qualitative results*). The provided qualitative and quantitative results validate the effectiveness of the proposed method compared to existing unified (UMVR [11], KD [4], TW

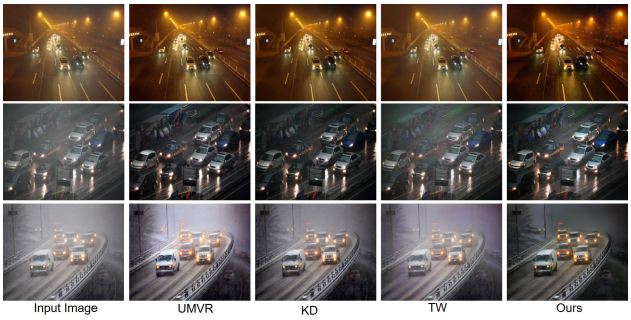

Figure 8. Visual analysis of the proposed and existing methods: UMVR [11], KD [4], TW [33] for night-time image restoration.

Table 3. Ablation study on the proposed modules with outdoor rain database (IL: Independent Learning, PMDA: Progressive Multi-domain Alignment and CMA: Cascaded Multi-head Attention)

Module	PSNR	SSIM
Baseline	26.49	0.887
Baseline + IL (**Net-1**)	27.89	0.905
Baseline + IL + PMDA (**Net-2**)	29.53	0.942
Baseline + IL + PMDA + CMA	**31.24**	**0.951**

[33]) methods on synthetic datasets. Specifically, our unified architecture has ∼11M parameters (including domain-translation architecture) whereas the existing unified architectures KD [4], and TW [33] have ∼28M and ∼31M parameters respectively.

As seen from the Table 1 and 2, our method's performance on synthetic data is usually comparable to SOTA that are trained using the respective synthetic datasets. We note that while our method learns from the synthetic data, it does not over-fit on synthetic data. Therefore, it performs well on both synthetic and real-world datasets. This proves that the domain generalization ability of the proposed method is better compared to SOTA unified restoration methods.

5.4. Night-time Multi-weather Restoration

Weather-degraded image/video restoration is equally important in night conditions as in day conditions [24]. Therefore, we examined the effectiveness of the proposed and existing networks for night-time multi-weather degradation removal. We labeled this analysis as cross-domain. We directly tested the proposed and existing models on real-world night-time multi-weather images downloaded from Internet. Sample results for night-time de-hazing, de-raining and de-snowing are provided in the Figure 8. From the visual results, it is observed that the proposed method performance is significantly better compared to existing methods for night-time multi-weather restoration.

5.5. Ablation Study

We analyse the contributions of individual modules (*independent feature learning, PMDA, and CMA*) of the proposed architecture in the ablation study conducted on ORD database. We start the ablation with single encoder-decoder architecture and named as "Baseline".

Figure 9. Object detection and depth estimation analysis on degraded and restored images by existing UMVR [11], KD [4], TW [33], and proposed multi-weather image restoration method.

Independent Feature Learning: In the proposed restoration approach, four inputs ($I_{\mathbb{D}1}$, $I_{\mathbb{D}2}$, $I_{\mathbb{D}3}$ and I) are processed parallelly to learn the features independently. **Is this independent feature learning effective for multi-weather image restoration?** To examine this, these inputs are processed through single **(baseline)** and parallel **Net-1** encoders. From the Table 3 results, it is clear that the proposed independent feature learning is effective for extracting the features from all domains.

Progressive Multi-domain Deformable Alignment: Along with effective feature extraction, the multi-domain feature alignment has significant importance in restoration. To do this task, we have proposed PMDA module and used at each encoder stage. **How does this proposed PMDA module helps the network to integrate the multi-domain features effectively?** To scrutinize this, the efficiency of proposed restoration network is examined without (**Net-1**) and with (**Net-2**) PMDA. The quantitative analysis is provided in Table 3. From these results, it is evident that the proposed PMDA works effectively for multi-domain feature alignment.

Cascaded Multi-head Attention: Applying attention on the encoder or decoder features affects differently while restoring the clean image. Therefore, we have applied cascaded multi-head attention on respective PMDA and decoder features. We analysed the network accuracy without (**Net-2**) and with (**baseline+IL+PMDA+CMA**) CMA module *i.e*. final proposed network. The results in Table 3 proves the effectiveness of the proposed CMA module for restoration. *The more ablation analysis for SOTS and CSD database is provided in the supplementary material*.

5.6. Real-world Applications

Restoration module can be used as a preprocessing step for down-stream applications like object detection, depth estimation, *etc*. To this end, we have evaluated the benefits of the proposed and existing restoration methods as a preprocessing step for the object detection [26] and depth estimation task [25]. The visual analysis for object detection

and depth predictions is provided in Figure 9. As seen from the Figure 9, depth predictions or confidence in detecting objects of the restored images with proposed approach are more accurate than existing methods and degraded image. Clearly, the proposed domain-translation based restoration approach performs favourably against the SOTA methods for object detection and depth predictions task.

6. Conclusion

In this work, we proposed the first weather-invariant representation learning based multi-domain progressive deform-able alignment architecture for multi-weather image restoration. The weather-invariant representation learning helps to learn a representation that is common across diverse weather conditions and also similar across real or synthetic weather degraded images. Further, the proposed progressive multi-domain deform-able alignment module effectively merges the independently learned multi-domain features. This progressive alignment of features helps the network to reduce the degradation in step-by-step manner. Finally, effective restoration is done via CMA to capture long-term dependencies with respective PMDA and decoder features. Results analysis is conducted on real-world and synthetic benchmark hazy, rainy and snowy datasets. Also, the night-time multi-weather restoration results are analysed as cross-domain analysis. Extensive comparisons demonstrate the generalizability of the proposed method over SOTA methods. Further, we have analysed the effectiveness of the proposed and existing methods for object detection and depth estimation tasks for verifying its usefulness in real-world applications.

References

[1] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 1375–1383. IEEE, 2019. 1, 2

[2] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. JSTASR: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *European Conference on Computer Vision*, pages 754–770. Springer, 2020. 3

[3] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021. 3, 6

[4] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17653–17662, 2022. 2, 3, 7, 8

[5] Sen Deng, Mingqiang Wei, Jun Wang, Yidan Feng, Luming Liang, Haoran Xie, Fu Lee Wang, and Meng Wang. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14560–14569, 2020. 2

[6] Sobhan Kanti Dhara, Mayukh Roy, Debashis Sen, and Prabir Kumar Biswas. Color cast dependent image dehazing via adaptive airlight refinement and non-linear color balancing. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5):2076–2081, 2020. 2

[7] Guodong Fan, Min Gan, Bi Fan, and CL Philip Chen. Multiscale cross-connected dehazing network with scene depth fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 2

[8] Zeyi Huang, Yang Zou, BVK Kumar, and Dong Huang. Comprehensive attention self-distillation for weakly-supervised object detection. *Advances in Neural Information Processing Systems*, 33, 2020. 1

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 6

[10] Somi Jeong, Youngjung Kim, Eungbean Lee, and Kwanghoon Sohn. Memory-guided unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6558–6567, 2021. 3

[11] Ashutosh Kulkarni, Prashant W Patil, Subrahmanyam Murala, and Sunil Gupta. Unified multi-weather visibility restoration. *IEEE Transactions on Multimedia*, 2022. 2, 3, 7, 8

[12] 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, 3

[13] 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

[14] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1633–1642, 2019. 2, 6

[15] 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. 1, 3

[16] Yi-Sheng Liao and Chun-Rong Huang. Semantic context-aware image style transfer. *IEEE Transactions on Image Processing*, 31:1911–1923, 2022. 3

[17] Ryan Wen Liu, Yu Guo, Yuxu Lu, Kwok Tai Chui, and Brij B Gupta. Deep network-enabled haze visibility enhancement for visual iot-driven intelligent transportation systems. *IEEE Transactions on Industrial Informatics*, 2022. 2

[18] Shiguang Liu and Ting Zhu. Structure-guided arbitrary style transfer for artistic image and video. *IEEE Transactions on Multimedia*, 24:1299–1312, 2021. 3

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

[20] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012. 6

[21] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6

[22] Prashant W Patil, Kuldeep M Biradar, Akshay Dudhane, and Subrahmanyam Murala. An end-to-end edge aggregation network for moving object segmentation. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8149–8158, 2020. 6

[23] Prashant W Patil, Sunil Gupta, Santu Rana, and Svetha Venkatesh. Dual-frame spatio-temporal feature modulation for video enhancement. *Pattern Recognition*, 130:108822, 2022. 6

[24] Prashant W Patil, Sunil Gupta, Santu Rana, and Svetha Venkatesh. Video restoration framework and its meta-adaptations to data-poor conditions. In *European Conference on Computer Vision*, pages 143–160. Springer, 2022. 8

[25] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12179–12188, 2021. 8

[26] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 8

[27] Wenqi Ren, Jingang Zhang, Xiangyu Xu, Lin Ma, Xiaochun Cao, Gaofeng Meng, and Wei Liu. Deep video dehazing with semantic segmentation. *IEEE Transactions on Image Processing*, 28(4):1895–1908, 2018. 1

[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[29] Joongchol Shin, Hasil Park, and Joonki Paik. Region-based dehazing via dual-supervised triple-convolutional network. *IEEE Transactions on Multimedia*, 24:245–260, 2021. 6

[30] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 6

[31] Wenqi Ren Zhangyang Wang Eric K. Tokuda Roberto Hirata Junior Roberto Cesar-Junior Jiawan Zhang Xiaojie Guo Siyuan Li, Iago Breno Araujo and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. *IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 6

[32] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2995–3000, 2014. 2

[33] 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, 3, 7, 8

[34] Yinglong Wang, Shuaicheng Liu, Chen Chen, and Bing Zeng. A hierarchical approach for rain or snow removing in a single color image. *IEEE Transactions on Image Processing*, 26(8):3936–3950, 2017. 2

[35] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 6

[36] Wending Yan, Aashish Sharma, and Robby T Tan. Optical flow in dense foggy scenes using semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13259–13268, 2020. 1

[37] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2037–2046, 2022. 2

[38] 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, 3

[39] Xianhui Zheng, Yinghao Liao, Wei Guo, Xueyang Fu, and Xinghao Ding. Single-image-based rain and snow removal using multi-guided filter. In *International conference on neural information processing*, pages 258–265. Springer, 2013. 2

[40] Yu Zhou, Zhihua Chen, Ping Li, Haitao Song, CL Philip Chen, and Bin Sheng. FSAD-Net: Feedback spatial attention dehazing network. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 2