

Supplementary File of Degradation-Adaptive Neural Network for Jointly Single Image Dehazing and Desnowing

Due to the page limit of the LETTER format (3 printed pages), we provide the supplementary file of “Degradation-Adaptive Neural Network for Jointly Single Image Dehazing and Desnowing”. This file includes the following materials:

- Section 1 shows our main motivation and contribution.
- Section 2 provides the details of the task-specific expert networks.
- Section 3 describes the proposed DAN-Net.
- Section 4 shows the implementation details of our algorithm.
- Section 5 reveals more comparisons with state-of-the-art dehazing and desnowing methods.
- Section 6 provides the ablation studies to prove the contribution of each proposed module.

1 Motivation and Contribution

In daily winter scenes, the uneven haze often occurs with diverse snowy degradations and the widely used haze imaging model [1] is:

$$I(x) = \mathcal{J}(x)t(x) + \mathcal{A}(1 - t(x)) \quad (1)$$

where $t(x)$ is the transmission map of hazy image $I(x)$, and $\mathcal{A}$ is the global atmospheric light. Single image dehazing methods aim to recover the clean image $\mathcal{J}(x)$ from $I(x)$.

On the other hand, based on previous works [2, 3], the imaging of snow scene can be modeled as:

$$I(x) = \mathcal{K}(x)\mathcal{T}(x) + \mathcal{A}(x)(1 - \mathcal{T}(x)) \quad (2)$$

where $\mathcal{K}(x) = \mathcal{J}(x)(1 - \mathcal{Z}(x)\mathcal{R}(x)) + C(x)\mathcal{Z}(x)\mathcal{R}(x)$, $I(x)$ represents the snowy image, $\mathcal{K}(x)$ denotes the veiling-free snowy image, $\mathcal{T}(x)$ is the media transmission map, $\mathcal{A}(x)$ is the atmospheric light and $\mathcal{J}(x)$ is the snow-free image. $C(x)$ and $\mathcal{Z}(x)$ are the chromatic aberration map for snow images and the snow mask, respectively. $\mathcal{R}(x)$ is the binary mask, presenting the snow location information.

Obviously, the incompatibility of physical models implies a considerable challenge to remove haze and snow for a single compact network coordinately. In addition, the varying

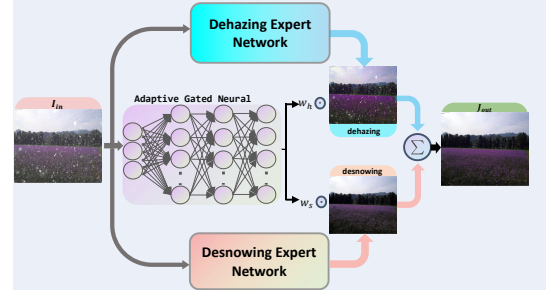


Fig. 1: The schematic overview of the proposed degradation-adaptive neural network (DAN-Net).

scales of snow degradation and uneven density of natural haze pose a significant challenge for efficient implementation.

Based on the aforementioned analysis, most existing dehazing and desnowing methods [2–20] have been designed to overcome only one type of degradation. For instance, Ju *et al.* proposed several traditional dehazing methods such as IDE [4] and IDBP [5] to estimate the key parameters of atmospheric model or enhance the quality of the hazy images, which contributes significantly to the task of image dehazing. Recently, a CNN-based universal method, *i.e.*, All-in-One [17], is proposed to remove all weather conditions at once. As a Neural Architecture Search (NAS) method, it consists of a task-specific encoder and a task-shared decoder to handle multiple degradations adverse weather removal problems. However, due to its high number of parameters, this method is not suitable for real-world applications. Moreover, it also ignores the implicit useful knowledge in different types of degradation. A two-stage knowledge learning method [19] has also been proposed to leverage a multi-teacher and contrastive learning strategy to achieve multiple adverse weather removal. However, it also suffers from high number of parameters.

To achieve practical real-world winter image restoration, we propose a lightweight Degradation-Adaptive Neural Network (DAN-Net) based on the Mixture of Experts (MoE) strategy [21] to jointly achieve dehazing and desnowing from a single degraded image, as shown in Fig. 1. The proposed DAN-Net consists of three parts: two task-specific expert networks and an adaptive gated neural module. More specif-

ically, three novel high-performance components are carefully devised to construct the expert network: Multi-branch Spectral Transform Block (MSTB) based on Fast Fourier Convolution and multi-scale filter to address the diverse scale degradation, Dual-Pool Attention Module (DPAM) to refine the uneven features of real-world degradation, and Cross-Layer Activation Gated Module (CLAGM) to optimize information-flow of features with different scales. To our surprise, the designed lightweight dehazing and desnowing expert networks can achieve best performance-parameter trade-off compared to previous state-of-the-art single-task algorithms. In addition, to skillfully combine the expert network with different task knowledge, a novel adaptive gated neural module is developed to learn the inner knowledge of expert networks for simultaneously removing haze and snow degradations. Experiments show that the proposed DAN-Net, as an efficient and effective image restoration solution, addresses diverse degradations in winter scenes, including complex, uneven hazy effects, as well as snow particles, streaks, and patches. The main contributions are summarized as follows:

- Considering multiple degrading effects (haze and snow) under real-world winter scenes, a lightweight DAN-Net consisting of two compact task-specific expert networks and an adaptive gated neural module is proposed to achieve jointly single image dehazing and desnowing.
- Three effective modules (MSTB, DPAB and CLAGM) as the fundamental components are designed to construct a task-specific expert network, which achieves best trade-off between restoration performance and parameters for image dehazing and desnowing.
- To fully make use of the potential of two task-specific expert networks, the presented adaptive gated neural module serves as an effective degradation-adaptive guider to control the contributions of these two pre-trained expert networks, which facilitates the removal of multiple degradations.
- Extensive experiments demonstrate that the proposed DAN-Net can provide satisfactory results for both synthetic and real-world degraded images under winter scenes, which performs favorably against other state-of-the-art dehazing or desnowing algorithms in subjective and objective evaluations.

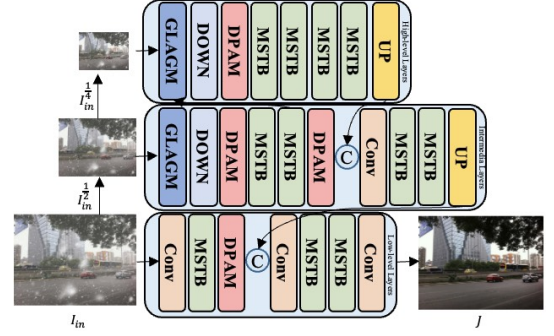


Fig. 2: The detailed architecture of the proposed Expert Network.

2 Task-Specific Expert Network

In this section, we present the overall architecture of the proposed task-specific expert network, shown in Fig. 2. Concretely, three high-performance convolution-based components (i.e. MSTB, DPAM and CLAGM) and three-level layers are carefully designed to construct the lightweight expert network, as depicted in Fig. 3(a)-(c). Different from previous learning-based methods [2, 9, 10, 14] that simply stack the components repeatedly to achieve impressive performance, our presented expert network deliberately avoids this strategy and achieves excellent parameter-performance trade-off through careful design in each layer and as few components as possible. Next, we describe the details of three components and three-level layers.

Multi-branch Spectral Transform Block (MSTB). Considering the uneven distribution of features in channel dimension in neural networks, we assume that the information from a part of the channel in incoming features can help network to establish global dependencies by spectral transformation. Consequently, the designed MSTB splits the incoming features F_{in} in $\mathbb{R}^{H \times W \times C}$ as:

$$\mathbb{R}_{in}^{H \times W \times C} \rightarrow \mathbb{R}_g^{H \times W \times \frac{C}{2}}, \mathbb{R}_l^{H \times W \times \frac{C}{2}} \quad (3)$$

Then, the global branch of MSTB performs Fast Fourier Convolution (FFC) on the $\mathbb{R}_g^{H \times W \times \frac{C}{2}}$. As shown in Fig. 3(a), the global branch adopts real FFT to account for global context. This parallel multi-branch structure combines different scale operation in single block, effectively capturing local information and global dependencies simultaneously and avoiding heavy computational burden. Specifically, the global branch first applies 2D Real FFT to the input features:

$$\text{Real FFT 2d} : \mathbb{R}_g^{H \times W \times \frac{C}{2}} \rightarrow \mathbb{C}^{H \times W \times \frac{C}{2}} \quad (4)$$

and then concatenates real and imaginary part for the subse-

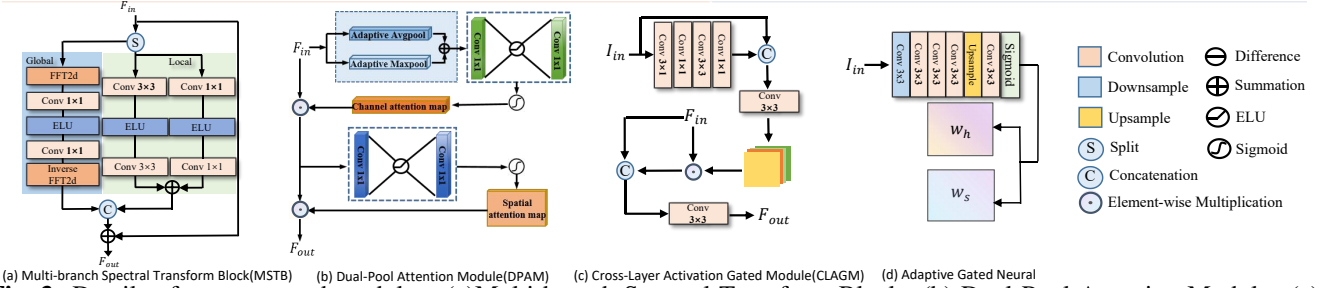


Fig. 3: Details of our proposed modules. (a) Multi-branch Spectral Transform Block. (b) Dual-Pool Attention Module. (c) Cross-Layer Activation Gated Module. (d) Adaptive Gated Neural.

quent feature extraction:

$$\text{ComplexToReal} : \mathbb{C}^{H \times W \times \frac{C}{2}} \rightarrow \mathbb{R}^{H \times W \times C} \quad (5)$$

Afterwards, we apply the convolution block with the kernel size of 1×1 in the frequency domain for capturing the long-distance dependence:

$$\text{Conv} \circ \text{ELU} \circ \text{Conv} : \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}^{H \times W \times C} \quad (6)$$

Finally, the inverse transformation is employed to recover the spatial structure of features:

$$\begin{aligned} \text{RealToComplex} &: \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{C}^{H \times W \times \frac{C}{2}} \\ \text{Inverse Real FFT2d} &: \mathbb{C}^{H \times W \times \frac{C}{2}} \rightarrow \mathbb{R}^{H \times W \times \frac{C}{2}}, \end{aligned} \quad (7)$$

In addition, we exploit multi-scale convolution blocks as the local branch to capture the local information, which is as important as global information in the procedure of details restoration:

$$\begin{aligned} \text{Conv} \circ \text{ELU} \circ \text{Conv with } 1 \times 1 &: \mathbb{R}_l^{H \times W \times \frac{C}{2}} \rightarrow \mathbb{R}_3^{H \times W \times \frac{C}{2}} \\ \text{Conv} \circ \text{ELU} \circ \text{Conv with } 3 \times 3 &: \mathbb{R}_l^{H \times W \times \frac{C}{2}} \rightarrow \mathbb{R}_1^{H \times W \times \frac{C}{2}}, \end{aligned} \quad (8)$$

Finally, we fuse all features from each branch and perform local residual learning by concatenate and add operator:

$$\text{Fusion} : \mathbb{R}_{ffc}^{H \times W \times \frac{C}{2}} \oplus (\mathbb{R}_3^{H \times W \times \frac{C}{2}} + \mathbb{R}_1^{H \times W \times \frac{C}{2}}) + \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}_{out}^{H \times W \times C} \quad (9)$$

The proposed MSTB effectively captures both global context information and local features, achieving the excellent trade-off between complexity and modeling ability due to the channel splitting solution and fast Fourier convolution.

Although some existing methods [22, 23] make use of Fourier transform for image dehazing, they only decompose the input image into high- and low frequency at shallow network layers. Compared with the above methods, our proposed MSTB incorporates Fourier transform into basic

blocks for efficient global information perception and utilization. More specifically, the proposed MSTB consists of multiple branches that perform spectral transformation with different filter sizes. This design allows the block to capture both local and global information, thereby enhancing feature extraction. By introducing Fourier transform into the basic block, we are able to efficiently capture global image information without compromising the computational efficiency of the network.

Dual-pool Attention Module (DPAM). The effectiveness of the attention module for the feature representation ability of the network has been demonstrated in the image restoration task. To improve the performance of the expert network, we introduce the dual-pool attention module into our network. As observed in Fig. 3(b), we make use of the AvgPool and MaxPool to sample the incoming features for the generation of channel dimension weights:

$$\begin{aligned} \text{AvgPool} &: \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}_{avg}^{1 \times 1 \times C} \\ \text{MaxPool} &: \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}_{max}^{1 \times 1 \times C} \end{aligned} \quad (10)$$

Then, we add the sampled features and adopt the convolution block to generate the attention weights of channel dimension:

$$\text{Conv} \circ \text{ELU} \circ \text{Conv with } 1 \times 1 : \mathbb{R}_{avg}^{1 \times 1 \times C} + \mathbb{R}_{max}^{1 \times 1 \times C} \rightarrow \mathbb{R}^{1 \times 1 \times C} \quad (11)$$

The *Sigmoid* function is used to keep the boundary of final attention weights:

$$\text{Sigmoid} : \mathbb{R}^{1 \times 1 \times C} \rightarrow \mathbb{R}_c^{1 \times 1 \times C}, \quad (12)$$

The attention mechanism in spatial dimension is also necessary for refining the features. The plain convolution block can generate the attention map in 2D dimension:

$$\text{Conv} \circ \text{ELU} \circ \text{Conv with } 1 \times 1 : \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{R}^{H \times W \times \frac{C}{r}} \rightarrow \mathbb{R}^{H \times W \times 1}, \quad (13)$$

where r is the scale-factor and we empirically set it as 4 for all experiments. The *Sigmoid* function constrains the final

spatial attention map:

$$\text{Sigmoid} : \mathbb{R}^{H \times W \times 1} \rightarrow \mathbb{R}_s^{H \times W \times 1}. \quad (14)$$

Cross-Layer Activation Gated Module (CLAGM). The information loss caused by repetitive down-sampling seems unavoidable in a multi-scale feature flow network. The recent research [24] reveals that the retinal ganglion cells and cerebral visual cortex cells have a whole new way of cell communication, which is mediated by exosomes different from synapses. Inspired by this idea, we find that the additional external information can bring the valuable activation like an instructor for the current layer. Thus we aim to leverage external information from the down-sampled original image to effectively regulate the input features from another layer. As observed in Fig. 3(c), a series of operations are preformed on $I_{in}^{1/2}$ or $I_{in}^{1/4}$ to generate the activation gated map $\mathcal{M}^{H \times W \times C}$:

$$\mathcal{M}^{H \times W \times C} = \text{CLAGM}(I_{in}^{H \times W \times 3}), \quad (15)$$

where $\mathcal{M}^{H \times W \times C}$ is a unique attention map from additional information, whose size is the same as the F_{in} . Then, we multiply $\mathcal{M}$ with input feature F_{in} :

$$F_{gated}^{H \times W \times C} = \mathcal{M}^{H \times W \times C} \cdot F_{in}^{H \times W \times C}, \quad (16)$$

Afterwards, the original F_{in} and the weighted F_{gated} are fused by channel merging and the following compressing:

$$\text{Conv with } 3 \times 3 : \mathbb{R}^{H \times W \times 2C} \rightarrow \mathbb{R}_{out}^{H \times W \times C} \quad (17)$$

The proposed cross-scale mechanism can help the network better exploit the inner knowledge of the original degraded images and optimize the information flow over the network.

Lower-level Layers: First of all, the lower-level layers are designed to play the role of level one in extracting features from the degraded original images at their original resolution. As observed in Fig. 2, a 3×3 convolution is used to extract the features with 32 channels (16 channels in the tiny version). Second, these features will be sent into the intermediate layers and the subsequent attention module. Then, the concatenate operation is exploited to fuse the features that feed back from the intermediate layers and the refined features by the attention module, resulting in the features with 64 channels (32 channels in the tiny version):

$$\text{Concatenate: } \mathbb{R}_s^{H \times W \times 32} + \mathbb{R}_I^{H \times W \times 32} \rightarrow \mathbb{R}_s^{H \times W \times 64}, \quad (18)$$

where the $\mathbb{R}_I^{H \times W \times 32}$ is the features from the intermediate layers. In addition, we perform the 3×3 convolution to compress the channel of features.

$$\text{Conv with } 3 \times 3 : \mathbb{R}_s^{H \times W \times 64} \rightarrow \mathbb{R}_s^{H \times W \times 32}. \quad (19)$$

From Fig. 2, two cascaded MSTB and a 3×3 convolution are employed to provide the final clean image J from features.

Intermediate Layers: As depicted in Fig. 2, the intermediate layers as a transmitter and a receiver coordinates the complex information flow and enhances the coupling of each layer in the expert network. Specifically, the number of the channel of MSTB and DPAM is only 64 (32 in the tiny version) at this level, and CLAGM receives the $\frac{1}{2}$ down-sampled degenerated images to aggregate and refines features. In the proposed task-specific expert network, the intermediate layers serve as a critical link between various layers by gathering the most intricate features from different scales and refining them to explore deeper and more valuable features.

High-level Layers: The designed high-level layers are responsible for extracting more information in high dimension space from features with 128 channels (64 channels in tiny version) and $\frac{1}{4}$ size. In Fig. 2, with the help of $\frac{1}{4}$ down-sampled degenerated images and external features from the intermediate layers, the high-level layers help the designed expert network to generate more pleasant clean images from degraded winter images.

3 Degradation-Adaptive Neural Network

The MoE strategy [21, 25] is an effective long-standing solution that exploits the integration of multiple expert networks to improve single-task performance. A trainable gating network is usually employed to produce the gated weights for activating each expert network based on an explicit or implicit partition of the input data, such as labels, classes (explicit) or contextual content features (implicit). In the DAN-Net, the proposed adaptive gated neural learns how to generate the gated attention maps as the activating weight, according to the uncertain degradation of the input winter image without explicit signal. Besides, instead of the intent of improving single-task performance in previous methods [21, 26], the adaptive mechanism is adopted based on MoE and the rich inner knowledge of task-specific expert networks, which can not only address the uncertain multi-degradation problem but also boost the comprehensive performance of DAN-Net in winter nasty scenes. Such a novel image restoration pipeline is effective and efficient due to the degradation-adaptive mechanism and well-learned expert networks with lightweight architecture.

Specifically, an adaptive gated neural module is devised as the guider to adaptively control the contribution of two different pre-trained expert networks. For real-world winter

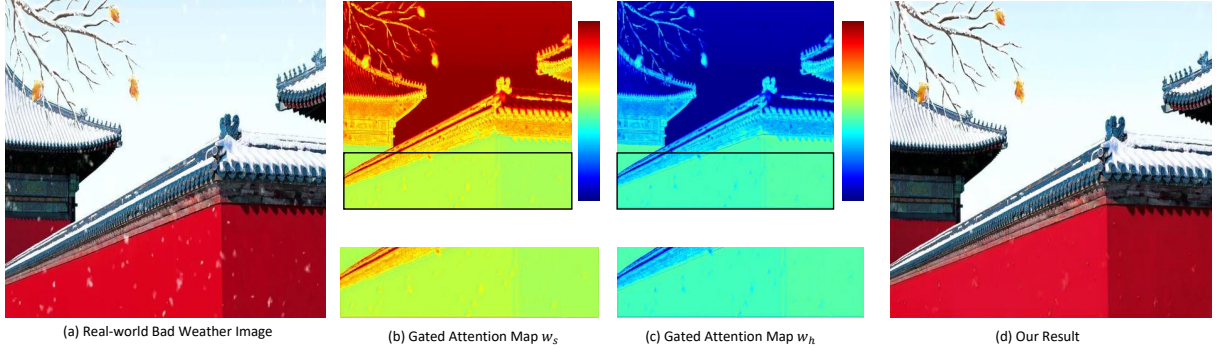


Fig. 4: Illustration of the gated attention maps. (a) and (d) are the real-world bad weather image and the corresponding restoration result. (b) and (c) are two gated attention maps w_s , w_h corresponding to the desnowing and dehazing expert activation intensity, respectively.

conditions, the captured degraded image usually suffers from the uneven spatial distribution and different types of degradations. To overcome this challenge, we introduce 2D attention maps that serves as a gated mechanism to address the degraded attributes and improve overall performance. In Fig. 1, the adaptive gated neural predicts the gated attention maps $w_s \in \mathbb{R}^{1 \times H \times W}$, $w_h \in \mathbb{R}^{1 \times H \times W}$ based on the input degraded images:

$$w_h, w_s = \text{AGN}(I_{in}), \quad (20)$$

where **AGN** denotes the adaptive gated neural network, and w_h, w_s denote the gated attention maps for dehazing and desnowing results, respectively. Then, we multiply the gated attention maps w_h and w_s with $J_{dehazing}$, $J_{desnowing}$, which are generated by dehazing expert and desnowing expert networks, respectively:

$$J_{out} = w_h \cdot J_{dehazing} + w_s \cdot J_{desnowing} \quad (21)$$

where $J_{out}^{3 \times H \times W}$ is the final result of the proposed DAN-Net.

Experiments demonstrate that the usage of the adaptive gated mechanism is beneficial for the performance of both dehazing and desnowing tasks and our DAN-Net achieves better performance on both tasks than single task-specific expert network. Fig. 4 depicts the obtained gated attention maps from a real-world bad weather image and the corresponding restoration result. From Fig. 4(b) and Fig. 4(c), the snowy degradation regions in w_s have higher activation intensity than w_h and other areas, which indicates the effectiveness of the proposed adaptive gated neural module.

4 Experiments

4.1 Datasets and Metrics

We choose the widely used PSNR and SSIM as experimental metrics to measure the performance of our networks.

The proposed task-specific expert networks are trained and tested on five large-scale datasets, namely RESIDE [27], Haze4k [14], CSD [3], SRRS [2] and Snow100K [15]. For fair comparison, we follow the latest dehazing and desnowing method [3, 14] for the benchmark setting. In addition, we randomly choose 1000 paired data from hazy or snowy datasets respectively to train the proposed DAN-Net.

4.2 Training Settings

We augment the training dataset with random rotation by 90, 180, 270 degrees and horizontal flip. The training image patches with the size 256×256 are extracted as input images of our networks. The Adam optimizer is used with initial learning rate of 2×10^{-4} , and we employ the CyclicLR to adjust the learning rate, where on the triangular mode, the value of gamma is 1.0, base momentum is 0.9, max learning rate is 3×10^{-4} and base learning rate is the same as initial learning rate. We adopt the PyTorch to implement all the networks with 4 Tesla V100 GPU with total batchsize of 40. The λ_{st} as trade-off weight is set as 0.2 in all experiments.

For the training details of the expert networks, each expert network, namely desnowing expert network or dehazing network, was trained on a single-task dataset. For instance, the desnowing expert network was trained on the corresponding desnowing benchmark training dataset, such as CSD, SRRS and Snow 100K. The training process used the same optimizer configuration as our proposed DAN-Net, and the number of training epochs for all expert networks was 500. Similarly, in terms of dehazing expert network, it also was trained and tested on the corresponding hazy benchmark training dataset, such as Haze4K and SOTS Outdoor. The optimizer configuration and the number of training epochs are identical to our proposed DAN-Net.

4.3 Loss Function

The proposed DAN-Net is trained in an end-to-end manner using a Charbonnier loss [28] as the basic reconstruction loss between the prediction and the ground truth, which is defined as follows:

$$\mathcal{L}_{\text{char}} = \frac{1}{N} \sum_{i=1}^N \sqrt{\|X^i - Y^i\|^2 + \epsilon^2}, \quad (22)$$

where ϵ is empirically set as 10^{-3} for all the experiments, X^i denotes the generated image of network and Y^i correspondingly denotes the ground truth. In addition, we also add the spectral transform to evaluate the difference of the reconstructed images and the ground truth images in the frequency domain:

$$\mathcal{L}_{st} = \frac{1}{N} \sum_{i=1}^N \|\mathcal{F}(X^i) - \mathcal{F}(Y^i)\|_1, \quad (23)$$

where $\mathcal{F}$ denotes the FFT operation. The spectral transform loss function is remarkably beneficial to the improvement of SSIM. The overall loss function are summarized as follows:

$$\mathcal{L}_{\text{expert}} = \mathcal{L}_{\text{char}} + \lambda_{st} \mathcal{L}_{st} \quad (24)$$

where λ_{st} is the trade-off weight.

5 Comparison with State-of-the-art Methods

Comparisons of Dehazing Results: In Table 1, we summarize the performance of our dehazing expert network (Tiny), DAN-Net (Tiny) and state-of-the-art dehazing methods [1, 6, 9–14, 29] on Haze4k [14] and SOTS [27]. By analyzing Table 1, the proposed DAN-Net achieves the best performance with 29.24 dB PSNR and 0.97 SSIM on the test dataset of Haze4k. Furthermore, the tiny version of DAN-Net also achieves 28.56 dB PSNR and 0.96 SSIM on the test dataset of Haze4k with impressive fewer computation and parameters. The quantitative comparisons verify that our proposed task-specific dehazing network (Dehazing Expert-Net and Dehazing Expert-Net-Tiny) and DAN-Net (DAN-Net and DAN-Net-Tiny) can guarantee superior performance with lower computational complexity for the dehazing task compared with other prevalent algorithms.

Comparisons of Desnowing Results: To reveal the desnowing performance of the proposed expert network and DAN-Net, the quantitative comparisons with state-of-the-art desnowing [2, 3, 15–17] and all-in-one network [17] methods are also conducted on the widely used benchmarks (i.e.

CSD [3], SRRS [2] and Snow100K [15]), as depicted in Table 2. From Table 2, the proposed expert network (Desnowing Expert-Net and Desnowing Expert Net-Tiny) and DAN-Net (DAN-Net and DAN-Net-Tiny) show the better desnowing performance for both PSNR and SSIM values.

Comparison of Visual Effects: We also conduct the visual comparisons with state-of-the-art dehazing and desnowing methods on synthetic and real-world hazy images. In order to verify the desnowing and dehazing performance of the proposed DAN-Net, the synthetic snowy image with haze interference are used for comparisons and the processed results by different algorithms are revealed in Fig. 5. In comparisons, four specialized dehazing methods (i.e. MSBDN [10], FFA [9], DA [12] and PSD [13]) only remove the haze and are unable to handle the snow particles with different size. The desnowing methods, JSTASR [2] and HDCW [3], can effectively alleviate the snow degradations, whereas the residual snowflakes still exist in these snow removal results due to the insufficient desnowing ability. Furthermore, the hazes cannot be removed thoroughly because JSTASR [2] and HDCW [3] are incapable of dealing with multiple degradations. Compared with these above methods, our proposed DAN-Net can simultaneously achieve image dehazing and desnowing and the restoration results are more closer to the ground truths. For the real-world winter degraded images, the visual comparisons are also performed in Fig. 6. As observed in Fig. 6, the dehazing approaches cannot address the snow degradations and two specialized desnowing methods cannot obtain the satisfactory results for real-world snowy images because the residual snow streaks still remain in the restoration results. In addition, the details influenced by the haze cannot also be recovered well, such as in the second row of Fig. 6.

For single image dehazing task, the visual comparisons are conducted on synthetic and real-world hazy images to verify the effectiveness of the proposed DAN-Net. In Fig. 7 and Fig. 8, previous state-of-the-art dehazing algorithms cannot effectively remove the uneven haze effects and the obvious hazes remain in the restoration results. In comparisons, the proposed DAN-Net outperforms other specialized dehazing methods in the aspect of visual effects for synthetic and real-world scenarios.

Furthermore, more degraded images acquired by the real-world winter scene and the corresponding restoration results are revealed in Fig. 9 to demonstrate the robustness of the proposed DAN-Net. From the bottom row of Fig. 9, our proposed network can provide haze-free and snow-free restoration images with more details.

Table 1: Quantitative comparisons of our framework with the state-of-the-art single image dehazing methods on Haze4k and SOTS outdoor dataset (PSNR(dB)/SSIM). The best results are **bold**, second best results are underlined.

Method	Haze4k [14]		SOTS Outdoor [27]		#Param	#GMacs
	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
(TPAMI'10)DCP [1]	14.01	0.76	15.09	0.76	-	-
(CVPR'16)NLD [29]	15.27	0.67	17.27	0.75	-	-
(TIP'16)DehazeNet [6]	19.12	0.84	22.46	0.85	0.01M	0.52G
(ICCV'17)GDN [11]	23.29	0.93	30.86	0.98	0.96M	21.49G
(CVPR'20)MSBDN [10]	22.99	0.85	23.36	0.87	31.35M	41.58G
(CVPR'20)DA [12]	24.03	0.90	27.76	0.938	1.64M	33.59G
(AAAI'20)FFA-Net [9]	26.97	0.95	<u>33.57</u>	<u>0.98</u>	4.6M	150.94G
(CVPR'21 Oral)PSD-FFA [13]	16.56	0.73	15.29	0.72	4.6M	150.94G
(ACMMM'21)DMT-Net [14]	<u>28.53</u>	<u>0.96</u>	-	-	54.9 M	80.71G
Dehazing Expert-Net	29.12	0.97	33.59	0.99	1.14M	12.00G
Dehazing Expert-Net-Tiny	28.18	0.96	32.09	0.98	288K	3.06G
DAN-Net	29.24	0.97	33.69	0.99	2.73M	31.36G
DAN-Net-Tiny	28.56	0.96	32.24	0.98	1.02M	13.47G

Table 2: Quantitative comparisons of our expert networks and DAN-Net(Tiny) with the state-of-the-art desnowing methods on CSD, SRRS and Snow 100K desnowing dataset (PSNR(dB)/SSIM). The best results are shown in **bold**, and second best results are underlined.

Method	CSD(2000) [3]		SRRS (2000) [2]		Snow 100K (2000) [15]		#Param	#GMacs
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
(TIP'18)Desnow-Net [15]	20.13	0.81	20.38	0.84	30.50	0.94	15.6M	-
(CVPR'20)All-in-One [17]	26.31	0.87	24.98	0.88	26.07	0.88	44 M	12.26G
(ECCV'20)JSTASR [2]	27.96	0.88	25.82	0.89	23.12	0.86	65M	-
(ICCV'21)HDCW-Net [3]	29.06	<u>0.91</u>	27.78	<u>0.92</u>	<u>31.54</u>	<u>0.95</u>	6.99M	9.78G
Desnowing Expert-Net	30.56	0.95	29.07	0.95	32.14	0.96	1.1M	12.00G
Desnowing Expert-Net-Tiny	29.06	0.92	28.20	0.94	31.67	0.95	288K	3.06G
DAN-Net	30.82	0.95	29.34	0.95	32.48	0.96	2.73M	31.36G
DAN-Net-Tiny	29.12	0.92	28.32	0.94	31.93	0.95	1.02M	13.47G

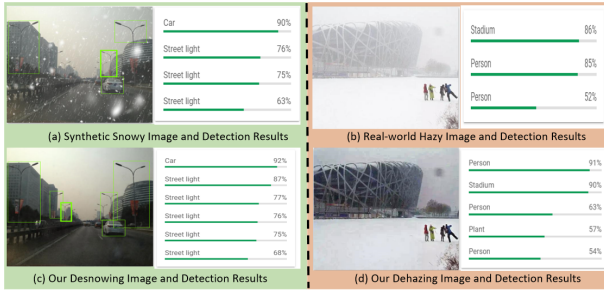


Fig. 10: Synthetic and real-world winter degenerated images (top) and corresponding restoration results (bottom) of the proposed DAN-Net with labels and confidences by Google Vision API.

Comparison of Computational Complexity: For real-time deployment of models, the computational complexity is an important factor to consider. To check the computational complexity, we summarize the number of GMacs and parameters of previous state-of-the-art desnowing and dehazing networks in Table 1 and Table 2. It is evident that the tiny desnowing expert network achieves best parameter-performance trade-off compared with the latest state-of-the-art desnowing method HDCW-Net (ICCV'21) and the num-

ber of GMacs is only 3.06G for the input size of 256×256 .

Quantifiable Performance for High-level Vision Tasks:

To verify the effectiveness of the proposed algorithm for high-level vision tasks, we adopt Google Vision API¹⁾ to produce subjective but quantifiable examples, as viewed in Fig. 10. In Fig. 10, the degradations can impede the performance of high-level vision tasks and our restoration results have higher visibility and confidences for multiple labels. These results suggest that our proposed DAN-Net can help improve the performance of high-level tasks.

6 Ablation Study

To demonstrate the contribution of the proposed modules, the ablation studies are performed. We make use of the widely used desnowing and dehazing datasets (i.e. CSD [3] and Haze4k [14]) for training and evaluation. Next, the effectiveness of three components and the loss function are verified.

¹⁾ Google Vision API: <https://cloud.google.com/vision/>

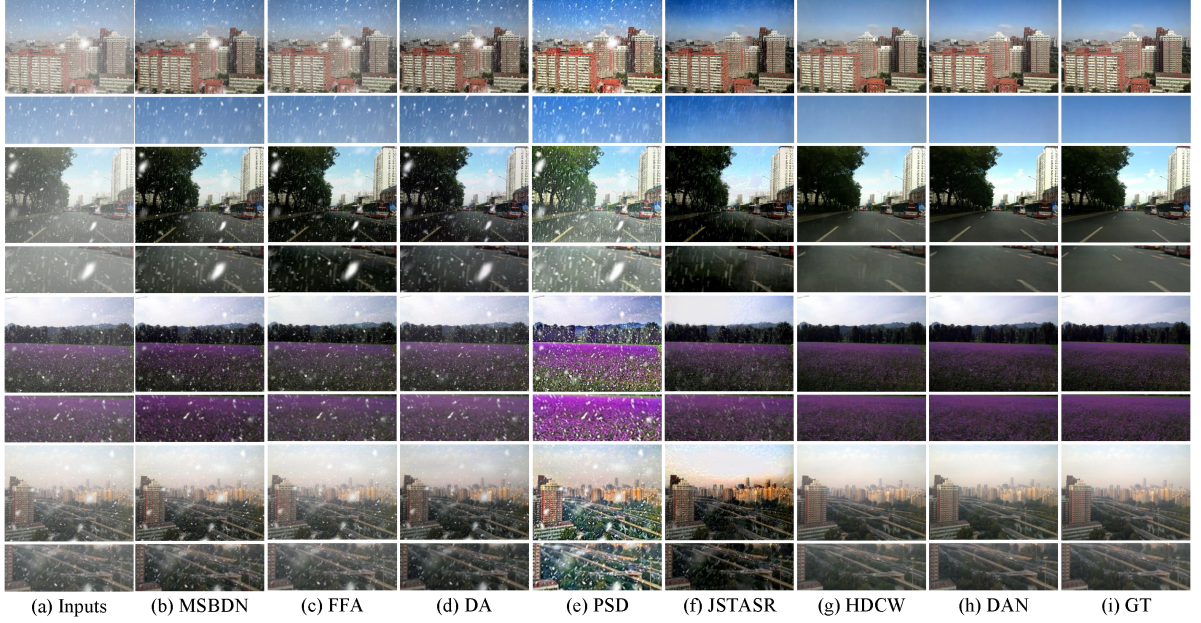


Fig. 5: Visual comparisons of desnowing results on synthetic snowy images from CSD dataset [3]. (a) and (i) are the input snowy images with haze interference and the ground truths. (b)-(g) are processed results respectively produced by MSBDN [10], FFA [9], DA [12], PSD [13], JSTASR [2] and HDCW [3]. (h) is our results provided by DAN-Net.

Effectiveness of MSTB: To prove the superior performance of spectral transform for network training, we replace the global branch with plain Conv-ELU-Conv branch in the proposed network. Specifically, in Table 3, we respectively adopt the vanilla convolution with kernel size 3×3 (**VC**), the Local-branch (**LB**), the Global-branch (**GB**), the MSTB without spectral transform operation (**MSTBwoST**) and the complete MSTB (**Ours**) to evaluate the dehazing and desnowing performance of the proposed network with different MSTB configurations. From Table 3, the designed spectral transform for the degradation feature extraction technique can achieve better recovered results compared with other configurations.

Table 3: Comparisons of the performance of the proposed expert network with different MSTB configurations.

Metric	VC	LB	GB	MSTBwoST	Ours
Desnowing PSNR/SSIM	27.13/0.91	27.43/0.92	27.92/0.91	28.21/0.92	30.56/0.95
Dehazing PSNR/SSIM	26.53/0.93	26.69/0.93	26.73/0.93	27.05/0.94	29.12/0.97

Effectiveness of DPAM: To verify the effectiveness of the proposed DPAM, we construct several settings for comparison, that is, (i) The proposed expert network without DPAM (**wo DPAM**); (ii) Remove the Channel Attention Part (**CA**) of the attention module in the proposed DPAM. (iii) Remove the Spatial Attention Part (**SA**) of the attention module in the proposed DPAM. The quantitative results in Table 4 indicate that the performance of the expert networks with the

proposed complete DPAM is better compared with other settings.

Table 4: Comparisons of the performance of the proposed expert networks with different DPAM configurations.

Metrics	wo DPAM	CA	PA	Ours
Desnowing PSNR/SSIM	28.76/0.93	29.31/0.94	29.67/0.94	30.56/0.95
Dehazing PSNR/SSIM	28.15/0.94	28.57/0.96	28.68/0.96	29.12/0.97

Effectiveness of CLAGM: In the proposed CLAGM, the external information from the down-sampled original image is adopted to regulate the input features for better performance. The ablation experiments are performed to demonstrate the effectiveness of the regulation mechanism. As observed in Fig. 11, we provide six configurations of CLAGM. The performance of different configurations is listed in Table 5. It is evident that the proposed expert networks with the complete CLAGM exhibit highest PSNR and SSIM values, which indicates the effectiveness of the proposed CLAGM.

Table 5: Comparisons of the performance of the proposed expert networks with different CLAGM configurations.

Configurations	A	B	C	D	E	Ours
Desnowing PSNR/SSIM	29.29/0.93	29.48/0.93	29.97/0.94	30.29/0.94	30.32/0.95	30.56/0.95
Dehazing PSNR/SSIM	27.73/0.94	28.11/0.95	28.30/0.95	28.49/0.96	28.92/0.97	29.12/0.97

Effectiveness of loss function: For the loss function, we

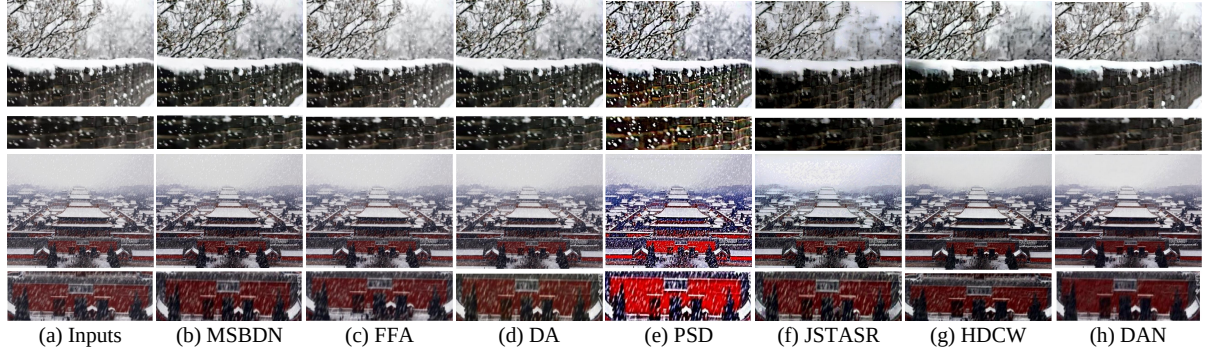


Fig. 6: Visual comparisons of desnowing results on real-world snowy images. (a) is the input snowy images. (b)-(g) are processed results respectively produced by MSBDN [10], FFA [9], DA [12], PSD [13], JSTASR [2] and HDCW [3]. (h) is our results provided by DAN-Net.

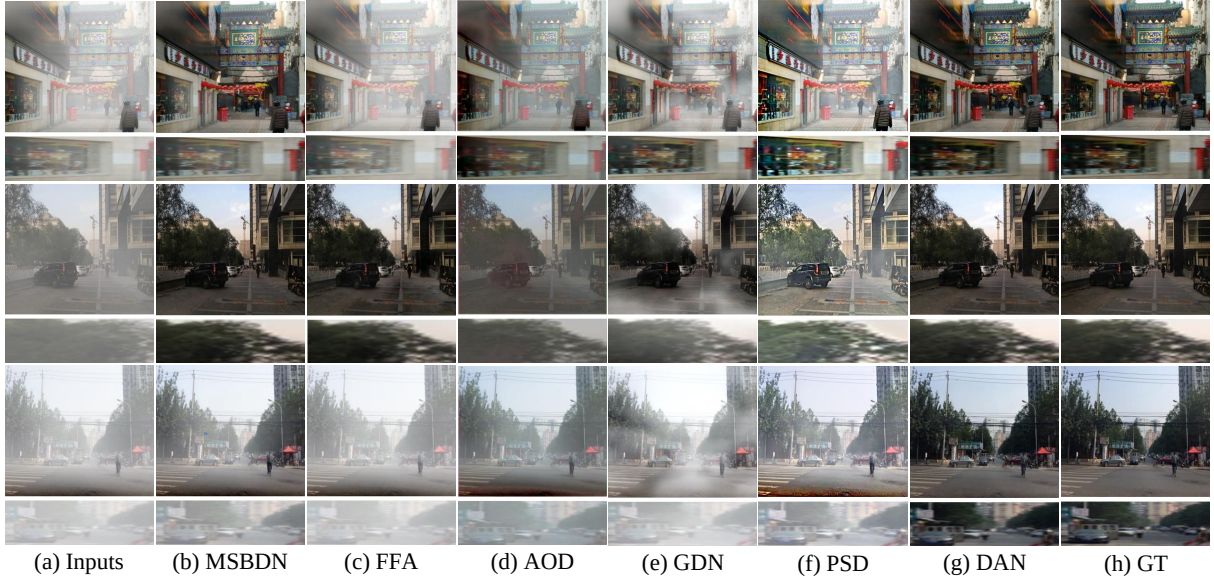


Fig. 7: Visual comparisons of dehazing results on synthetic hazy images from Haze4K [14]. (a) and (h) are the input hazy image and the corresponding ground truths. (b)-(f) are the dehazed results respectively produced by MSBDN [10], FFA [9], AOD [8], GDN [11] and PSD [13]. (g) is the dehazed results provided by the proposed DAN-Net.

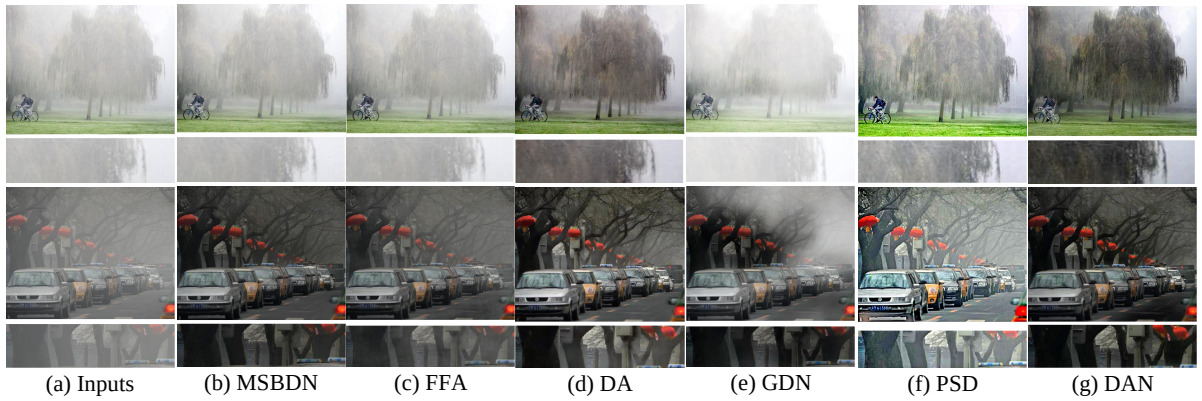


Fig. 8: Visual comparisons of dehazing results on real-world hazy images. (a) is the input hazy image. (b)-(f) are the dehazed results respectively produced by MSBDN [10], FFA [9], DA [12], GDN [11] and PSD [13]. (g) is the dehazed results by the proposed DAN-Net.

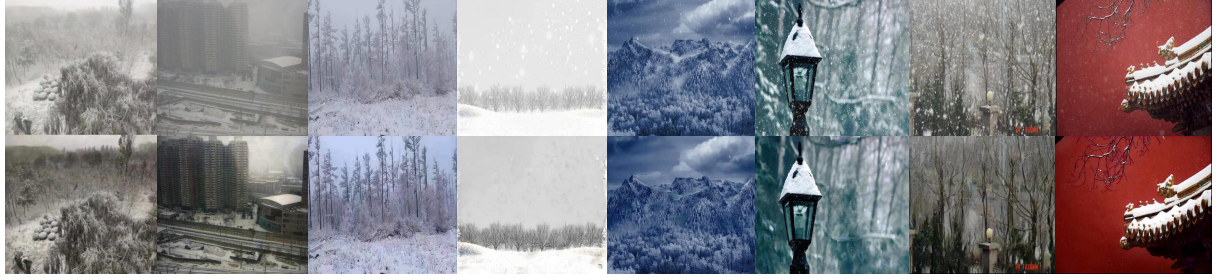


Fig. 9: Real-world winter images (top) and corresponding restoration results (bottom) using the proposed DAN-Net.

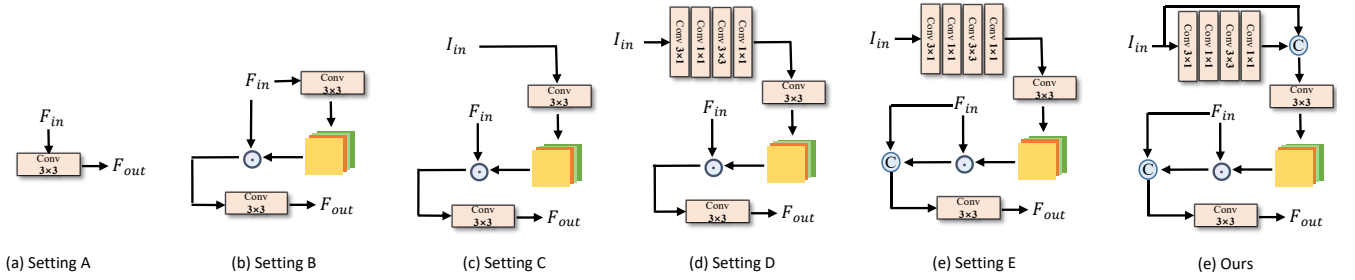


Fig. 11: Different configurations of CLAGM for ablation study.

also perform the ablation experiments to demonstrate the effectiveness of the used loss function in the proposed network, as shown in Table 6. The ‘**Baseline**’ means using the classic $L1$ loss on the final recovered result. In addition, we also respectively consider the Charbonnier loss function $\mathcal{L}_{char}$ and $\mathcal{L}_{char} + \mathcal{L}_{per}$ loss function to train our network. From Table 6, our designed overall loss function reveals the best dehazing and desnowing performance.

Table 6: Effectiveness of the proposed loss functions.

Metrics	Baseline	$\mathcal{L}_{char}$	$\mathcal{L}_{char} + \mathcal{L}_{per}$	Ours
Desnowing PSNR/SSIM	29.26 / 0.92	29.44 / 0.92	30.13 / 0.93	30.56/0.95
Dehazing PSNR/SSIM	28.42/0.94	28.56/0.95	28.73/0.96	29.12/0.97

References

- He K, Sun J, Tang X. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 2011, 33(12): 2341–2353
- Chen W T, Fang H Y, Ding J J, Tsai C C, Kuo S Y. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In: *European Conference on Computer Vision*. 2020, 754–770
- Chen W T, Fang H Y, Hsieh C L, Tsai C C, Chen I, Ding J J, Kuo S Y, others. 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*. 2021, 4196–4205
- Ju M, Ding C, Ren W, Yang Y, Zhang D, Guo Y J. Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE Transactions on Image Processing*, 2021, 30: 2180–2192
- Ju M, Ding C, Ren W, Yang Y. Idhp: Image dehazing using blended priors including non-local, local, and global priors. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4867–4871
- Cai B, Xu X, Jia K, Qing C, Tao D. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 2016, 25(11): 5187–5198
- Ren W, Liu S, Zhang H, Pan J, Cao X, Yang M H. Single image dehazing via multi-scale convolutional neural networks. In: *European conference on computer vision*. 2016, 154–169
- Li B, Peng X, Wang Z, Xu J, Feng D. Aod-net: All-in-one dehazing network. In: *Proceedings of the IEEE international conference on computer vision*. 2017, 4770–4778
- Qin X, Wang Z, Bai Y, Xie X, Jia H. Ffa-net: Feature fusion attention network for single image dehazing. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020, 11908–11915
- Dong H, Pan J, Xiang L, Hu Z, Zhang X, Wang F, Yang M H. Multi-scale boosted dehazing network with dense feature fusion. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 2157–2167
- Liu X, Ma Y, Shi Z, Chen J. Griddehazenet: Attention-based multi-scale network for image dehazing. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, 7314–7323
- Shao Y, Li L, Ren W, Gao C, Sang N. Domain adaptation for image dehazing. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, 2808–2817
- Chen Z, Wang Y, Yang Y, Liu D. Psd: Principled synthetic-to-real

- dehazing guided by physical priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, 7180–7189
14. Liu Y, Zhu L, Pei S, Fu H, Qin J, Zhang Q, Wan L, Feng W. From synthetic to real: Image dehazing collaborating with unlabeled real data. In: Proceedings of the 29th ACM International Conference on Multimedia. 2021, 50–58
 15. Liu Y F, Jaw D W, Huang S C, Hwang J N. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 2018, 27(6): 3064–3073
 16. Engin D, Genç A, Kemal Ekenel H. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018, 825–833
 17. Li R, Tan R T, Cheong L F. All in one bad weather removal using architectural search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 3175–3185
 18. Lin C, Rong X, Yu X a. Msaff-net: Multiscale attention feature fusion networks for single image dehazing and beyond. *IEEE Transactions on Multimedia*, 2022
 19. Chen W T, Huang Z K, Tsai C C, Yang H H, Ding J J, Kuo S Y. 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 (CVPR). June 2022, 17653–17662
 20. Liu Y, Yan Z, Tan J, Li Y. Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(4): 1643–1657
 21. Pavlitskaya S, Hubschneider C, Weber M, Moritz R, Huger F, Schlicht P, Zollner M. Using mixture of expert models to gain insights into semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020, 342–343
 22. Yu H, Zheng N, Zhou M, Huang J, Xiao Z, Zhao F. Frequency and spatial dual guidance for image dehazing. In: Avidan S, Brostow G, Cissé M, Farinella G M, Hassner T, eds, *Computer Vision – ECCV 2022*. 2022, 181–198
 23. Shen H, Wang C, Deng L, He L, Lu X, Shao M, Meng D. Fddn: frequency-guided network for single image dehazing. *Neural Computing and Applications*, 2023, 1–16
 24. Schiapparelli L M, Sharma P, He H Y, Li J, Shah S H, McClatchy D B, Ma Y, Liu H H, Goldberg J L, Yates III J R, others . Proteomic screen reveals diverse protein transport between connected neurons in the visual system. *Cell Reports*, 2022, 38(4): 110287
 25. Jacobs R A, Jordan M I, Nowlan S J, Hinton G E. Adaptive mixtures of local experts. *Neural computation*, 1991, 3(1): 79–87
 26. Aljundi R, Chakravarty P, Tuytelaars T. Expert gate: Lifelong learning with a network of experts. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, 3366–3375
 27. Li B, Ren W, Fu D, Tao D, Feng D, Zeng W, Wang Z. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 2019, 28(1): 492–505
 28. Charbonnier P, Blanc-Feraud L, Aubert G, Barlaud M. Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings of 1st International Conference on Image Processing. 1994, 168–172
 29. Berman D, Avidan S, others . Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, 1674–1682