

Snow Mask Guided Adaptive Residual Network for Image Snow Removal

Bodong Cheng, Juncheng Li*, Ying Chen*, Shuyi Zhang, and Tiejong Zeng

Abstract—Image restoration under severe weather is a challenging task. Most of the past works focused on removing rain and haze phenomena in images. However, snow is also an extremely common atmospheric phenomenon that will seriously affect the performance of high-level computer vision tasks, such as object detection and semantic segmentation. Recently, some methods have been proposed for snow removing, and most methods deal with snow images directly as the optimization object. However, the distribution of snow location and shape is complex. Therefore, failure to detect snowflakes / snow streak effectively will affect snow removing and limit the model performance. To solve these issues, we propose a Snow Mask Guided Adaptive Residual Network (SMGARN). Specifically, SMGARN consists of three parts, Mask-Net, Guidance-Fusion Network (GF-Net), and Reconstruct-Net. Firstly, we build a Mask-Net with Self-pixel Attention (SA) and Cross-pixel Attention (CA) to capture the features of snowflakes and accurately localized the location of the snow, thus predicting an accurate snow mask. Secondly, the predicted snow mask is sent into the specially designed GF-Net to adaptively guide the model to remove snow. Finally, an efficient Reconstruct-Net is used to remove the veiling effect and correct the image to reconstruct the final snow-free image. Extensive experiments show that our SMGARN numerically outperforms all existing snow removal methods, and the reconstructed images are clearer in visual contrast. All codes will be available.

Index Terms—Image snow removal, snow mask guided, self-pixel attention, cross-pixel attention, multi-level guidance.

I. INTRODUCTION

As a common atmospheric phenomenon, snow is often inevitably captured in images by camera lenses, which will affect the accuracy of high-level computer vision tasks such as image classification [1], [2], object detection [3], [4], and facial recognition [5]–[7]. Different from other weather phenomena, snow is more complex that includes opaque and translucent snowflakes and snow streaks, and will cause veiling effects. According to previous work [8], images affected by snow can be modeled as

$$I(x) = K(x)T(x) + A(x)(1 - T(x)), \quad (1)$$

*Corresponding author

B. Cheng is with the School of Computer Science and Technology, Xidian University, Xian, Shanxi 710071, China, and the Department of Cyberspace Security, Beijing Electronic Science and Technology Institute, Beijing, China. (E-mail: bdcheng@stu.xidian.edu.cn)

J. Li, and T. Zeng are with the Department of Mathematics, The Chinese University of Hong Kong, New Territories, Hong Kong. (E-mail: cvjunchengli@gmail.com, zeng@math.cuhk.edu.hk)

Y. Chen is with the Department of Cyberspace Security, Beijing Electronic Science and Technology Institute, Beijing, China. (E-mail: ychen@besti.edu.cn)

S. Zhang is with Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin 541004, China. (E-mail: zhangshuyi@stu.gxnu.edu.cn)

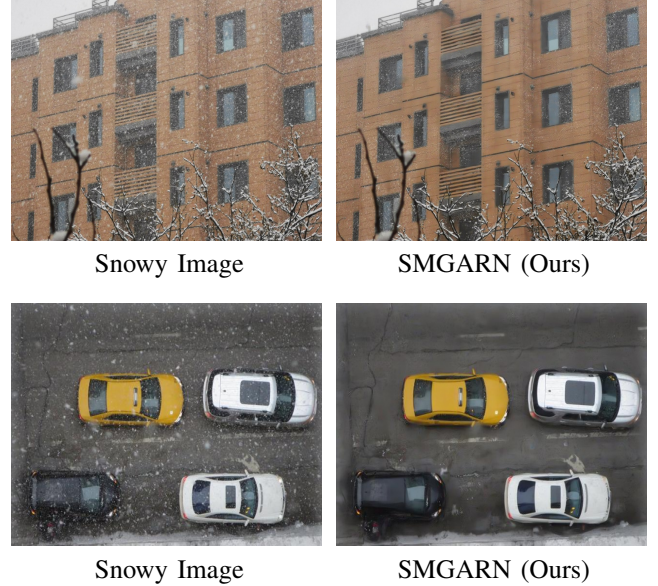


Fig. 1. Examples of real image snow removal by our SMGARN. All of these snowy images are real images captured by cameras, which contains different shapes and sizes of snowflakes / snow streaks. It can be seen that our SMGARN can reconstruct high-quality snow-free images in real scenes.

where $I(x)$ denotes the snowy image, $T(x)$ is the media transmission, and $A(x)$ is the atmospheric light. Meanwhile, $K(x)$ represents a snow scene image without veiling effect, which can be obtained by the following formula

$$K(x) = J(x)(1 - Z(x)R(x)) + C(x)Z(x)R(x), \quad (2)$$

where $J(x)$ is the scene radiance, $R(x)$ is a binary mask that presents the snow location information, $C(x)$ and $Z(x)$ are the chromatic aberration map for the snow image and the snow mask, respectively.

In the past, in order to better deal with the problem of single image desnowing, many methods [9]–[13] have been proposed with manually extracting features. However, most of them rely on human intuition and do not have the ability to learn the deep features of images, so the snow removal performance is limited and the generalization ability is weak. Recently, with the powerful feature extraction capabilities of convolutional neural networks (CNNs), more and more CNN-based image desnowing methods [8], [14], [15] have been proposed. Among them, Liu et al. [14] proposed the first CNN-based image snow removal method, called DesnowNet. It produces snow-free images by sequentially processing complex translucent and opaque snow particles. Chen et al. [15] took the removal

of the veiling effect as part of the image snow removal and inpainted the image. In order to better guide the model to learn the information of snow particles and pay more attention to the heavy snow. Jaw et al. [16] proposed an efficient modular snow removal network, and introduced a generative adversarial network (GAN) to further improve the snow removal ability. Although JSTASR also predicts snow mask as part of joint snow removal, its performance is not excellent and the number of model parameters is huge. Chen et al. [8] proposed a dual-tree complex wavelet transform-based image snow removal model, named HDCWNet, and used contradictory channel loss to improve the desnowing performance. Using Dual-Tree Complex Wavelet Transform (DTCWT) to locate high-frequency snow information in images enables HDCWNet to obtain better snow removal results. However, wavelet transform cannot accurately distinguish snow from other high-frequency information in the original image, so part of useful information will also be removed. Moreover, the loss of information in the up-sampling process of wavelet transform will inevitably result in the change of the information of these sub-bands, which will affect the quality of the reconstructed images.

According to our research on the generation mechanism of snow images, we found that the distribution of snow can be predicted through a specially designed network. Meanwhile, we believe that the predicted snow mask can accurately reflect the distribution of snow in the image and the reconstructed images guided by snow mask will have less snow residue. To achieve this, we propose a Snow Mask Guided Adaptive Residual Network (SMGARN) for image snow removal. Specifically, SMGARN adopts an end-to-end modular design, and the overall model consists of three parts: Mask-Net, Guidance-Fusion Network (GF-Net), and Reconstruct-Net. Among them, Mask-Net is proposed to predict the snow mask, GF-Net is specially designed to adaptively remove snow with the guidance of the learned snow mask, and Reconstruct-Net is used to suppress the veiling effect and reconstruct the final snow-free image. In summary, the main contributions of this paper are as follows

- We build an efficient Mask-Net to directly predict the snow mask from the snowy image. With the help of Self-pixel Attention (SA) and Cross-pixel Attention (CA), Mask-Net can capture the features of snowflakes and accurately localize the shape and location of snow.
- We design a multi-level Guidance-Fusion Network (GF-Net) to adaptively remove snow from the image with the guidance of the learned snow mask.
- We design a Reconstruct-Net to reconstruct the final snow-free image by the specially designed multi-scale aggregated residual blocks.
- We propose a novel Snow Mask Guided Adaptive Image Residual Network (SMGARN) for Image Snow Removal. Meanwhile, we construct a new real snow image test dataset, named SnowWorld24, which contains 24 different scenes of real snow images from all over the world.

The rest of this paper is organized as follows. Related works are reviewed in Section II. A detailed explanation of

the proposed SMGARN is given in Section III. The experimental results, ablation studies, and discussion are presented in Section IV, V, and VI respectively. Finally, we draw a conclusion in Section VI.

II. RELATED WORKS

Images are easily disturbed by atmospheric phenomena such as rain, fog, and snow during the acquisition process. This will contaminate and blur the acquired images, thus affecting subsequent analysis of the image. Therefore, designing efficient models to restore images distorted by extreme atmospheric phenomena into clear images is necessary.

A. Single Image Dehazing

Early image dehazing algorithms [17]–[21] took the physical properties of hazy images, such as contrast, reflectivity, etc., as research objects, to remove haze in images. For example, He et al. [22] analyzed a large number of haze-free images and proposed a well-known image dehazing algorithm guided by dark channel priors. After that, a considerable part of works take the dark channel as the research focus of image dehazing [23], [24]. However, due to the lack of generalization ability of manual setting priors, the generality of these methods is greatly limited. Recently, many deep learning-based methods [25]–[34] have been proposed for image dehazing. For instance, Cai et al. [26] proposed the first learning-based image dehazing model called Dehazenet. Zhang et al. [25] proposed a Densely Connected Pyramid Dehazing Network (DCPDN) that can simultaneously estimate the transmission map and atmospheric light intensity. Although these models have achieved promising results, they cannot play the role of image snow removal task since the haze phenomenon is quite different from the snow phenomenon. Image dehazing methods generally assume haze particles to be translucent and that there are no fully occluded opaque regions. This is not useful for snow removal since some snowflakes completely cover parts of the image and are difficult to remove.

B. Single Image Deraining

Traditional image deraining methods usually use gradient features [35] or sparse coding [36]–[38] to reconstruct rain-free images. However, these methods have limited performance and applicability. Recently, deep learning-based [39]–[44] also promoted the development of image deraining. For example, Fu et al. [39] proposed a deep neural network-based image deraining method, called DerainNet, which train the model on the high-pass layer rather than in the image domain. Yang et al. [43] proposed a recurrent deep network to progressively remove rain streaks. Yi et al. [44] proposed a Structure Preserving Deraining Network (SPDNet) with the guidance of the residue channel prior. Although rain images are similar to snow images, directly using existing rain removal methods cannot completely remove them from images due to the diverse states of snow.

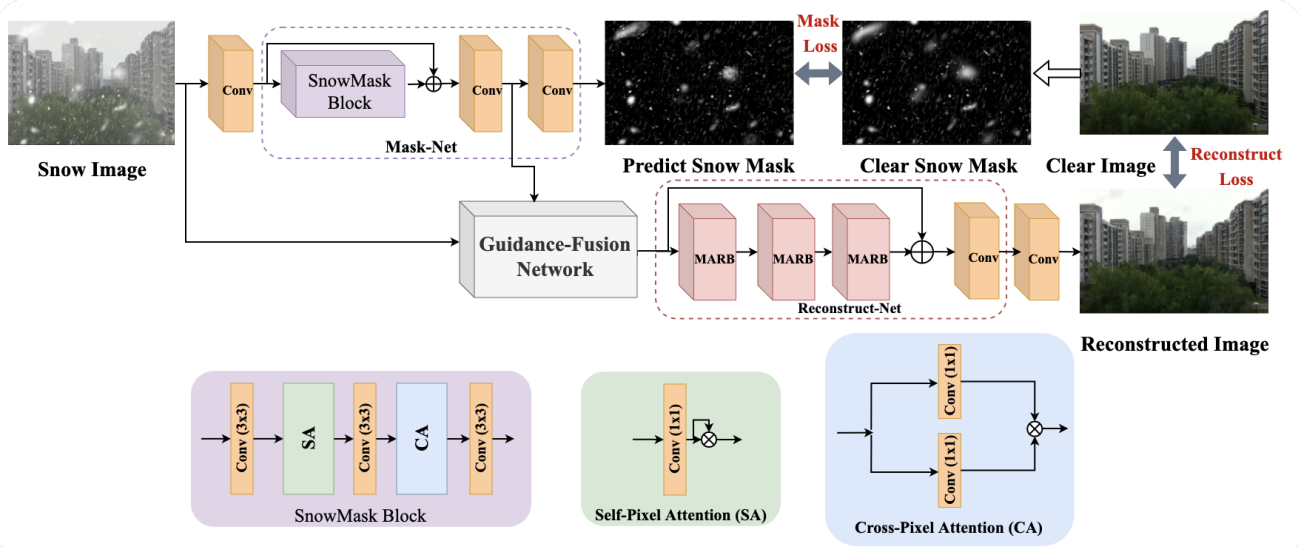


Fig. 2. The complete architecture of the proposed Snow Mask Guided Adaptive Residual Network (SMGARN). SMGARN consists of three parts: Mask-Net, GF-Net, and Reconstruct-Net.

C. Single Image Snow Removal

As another representation of atmospheric phenomena, the changes of snow are more complex, and the spatial states are more abundant. Traditional image snow removal methods still use feature priors to model snow particle information, such as histogram of gradients (HOG) [11], [45], frequency separation [46], etc. However, these methods cannot guarantee the generalization of snow removal and usually have poor performance. In order to further improve the ability of snow removal, Liu et al. [14] proposed the first deep neural network-based snow removal method, called DesnowNet. DesnowNet adopts the multi-scale pyramid model of Inception-v4 as the backbone and performs well in its proposed Snow100K dataset. Chen et al. [15] proposed a model named JSTASR that takes into account the veiling effect and opaque snow particles. It removes the effects caused by snow phenomena using convolution and slightly darker channel priors. After that, Chen et al. [8] proposed a hierarchical network based on the Dual-tree Complex Wavelet Transform (DTCWT), named HDCWNet, which solved the problem that JSTASR cannot remove snow patterns and color distortion.

Although HDCWNet has achieved state-of-the-art results, we observe that it will produce blurry images. This is because the downsampling unit is excessively used in HDCWNet to remove snowflakes and snow streaks. With the help of downsampling operations, snowflakes and snow streaks can be effectively removed, but also cause the loss of useful information, which is not conducive to reconstructing clear images. In this work, we aim to explore a more effective snow removal method that can remove snow from the image while preserving the quality of the image.

III. METHODOLOGY

Predicting an accurate snow mask can help the model to accurately locate and remove snow. More importantly, this

method will not mistakenly remove the original information of the image as snow, which greatly improves the quality of reconstructed image. To achieve efficient snow removal, we propose a Snow Mask Guided Adaptive Residual Network (SMGARN). As shown in Fig. 2, SMGARN consists of three parts: Mask-Net, Guidance-Fusion Network (GF-Net), and Reconstruct-Net. Among them, Mask-Net is specially designed to predict the snow mask of the snowy image. With the help of Self-pixel Attention (SA) and Cross-pixel Attention (CA) mechanisms, Mask-Net can quickly and accurately capture snowflakes and snow streaks in the image, thus can predict accurate snow mask. However, snow affects different regions of the image differently. Directly using snow masks and snow images to do residuals will generate a large number of dark areas, which will affect the quality of the reconstructed images. Therefore, we propose a multi-level Guidance-Fusion Network (GF-Net) to adaptively remove snow with the guidance of the predicted snow mask. Finally, in order to remove the haze phenomenon caused by the veiling effect, we design a multi-scale based Reconstruct-Net to achieve the effect of dehazing and further correct the area affected by snow. Each sub-network will be described in detail in the following chapters.

A. Mask-Net

As we all know, the key for image snow removal is to accurately capture snowflakes and snow streaks in the image. In HDCWNet [8], the authors use wavelet transform to decompose the high and low-frequency information to model snow. However, this method will confuse the information of non-snow targets in the image, and training with such features will affect the final result. To address this issue, we propose a mask prediction network (Mask-Net) to directly predict snow mask by using of the convolutional neural network. Specifically, Mask-Net (\mathcal{M}) takes the snowy image \mathcal{I}_{snowy} as input and

output the snow mask m and the mask feature map \mathcal{F}_{mask}

$$\mathcal{F}_{mask}, m = \mathcal{M}(\mathcal{I}_{snowy}). \quad (3)$$

As shown in Fig. 2, the backbone of Mask-Net consists of one SnowMaskBlock and two convolutional layers. In SnowMaskBlock, we design two kinds of pixel attention units, namely Self-pixel Attention (SA) unit and Cross-pixel Attention (CA) unit. It should be noted that the attention used in this paper is different from the operation performed by constructing a similarity matrix in the self-attention mechanism. The pixel attention we use here aims to enhance the activations of snow-covered regions in images with snow to help the model better extract snow features. Furthermore, we add residual connections on SnowMaskBlock to facilitate feature expression. Among them, SA is used to enhance the representation of important features and suppress unimportant features in the image. CA adopts two different encoding functions to improve the adaptability of the network. The operations of CA and SA can be defined as follows

$$\mathcal{SA}(\mathcal{X}) = f_0(\mathcal{X}) \odot f_0(\mathcal{X}), \quad (4)$$

$$\mathcal{CA}(\mathcal{X}) = f_1(\mathcal{X}) \odot f_2(\mathcal{X}), \quad (5)$$

where $\mathcal{SA}(\cdot)$ and $\mathcal{CA}(\cdot)$ denote the Self-pixel Attention and the Cross-pixel Attention, respectively. \mathcal{X} represents the feature matrix. $f_0(\cdot)$, $f_1(\cdot)$ and $f_2(\cdot)$ represent the encoding functions, and \odot represents the Hadamard product operation. To adequately encode features to extract deep information, we replace the encoding functions in Eq.(4) and Eq.(5) with convolutional layers

$$\mathcal{SA}(\mathcal{X}) = \text{Conv}_0(\mathcal{X}) \odot \text{Conv}_0(\mathcal{X}), \quad (6)$$

$$\mathcal{CA}(\mathcal{X}) = \text{Conv}_1(\mathcal{X}) \odot \text{Conv}_2(\mathcal{X}). \quad (7)$$

Specifically, we first square each element in the feature matrix, which will highlight regions with high activation value (regions with snow) and suppress regions with low activation (regions without snow). To further enhance the representation of features, we operate on the feature matrix through two parallel convolutional units, and then use the Hadamard product to intersect the two features.

The Mask-Net can be used as part of any image snow removal model to provide the snow mask or works independently to predict snow mask from the snowy image directly. The goal of Mask-Net is to learn a prediction function that can predict an accurate snow mask from the corresponding snowy input. To achieve this, we apply a mask loss \mathcal{L}_{mask} on Mask-Net

$$\mathcal{L}_{mask} = \|\mathcal{M}(\mathcal{I}_{snowy}) - \mathcal{I}_{mask}\|_1, \quad (8)$$

where $\mathcal{M}(\cdot)$ denotes the Mask-Net, $\mathcal{M}(\mathcal{I}_{snowy})$ represents the predicted mask, and \mathcal{I}_{mask} is the corresponding ground-truth snow mask. It is worth noting that we embed the Mask-Net as part of the SMGARN to provide the snow mask prior for snow-free images reconstruction in this work.

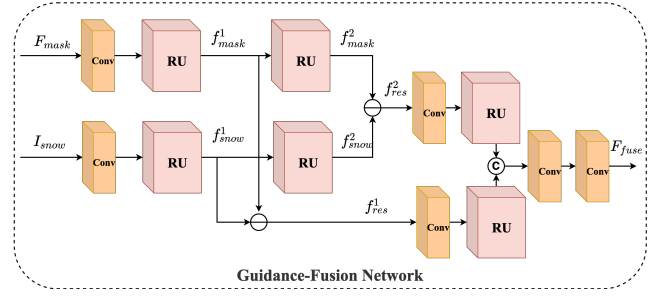


Fig. 3. The complete architecture of the proposed Guidance-Fusion Network.

B. Guidance-Fusion Network (GF-Net)

As an important bridge connecting snow images and snow masks, the proposed Guidance-Fusion Network (GF-Net, Fig. 3) plays a crucial role in improving the snow removal performance of the model. In GF-Net, we design a multi-level residual network to fully utilize the predicted snow mask to guide the model for snow removing. Specifically, we found that subtracting the snow image with the adaptively weighted snow mask (the \ominus symbol in Fig. 3 represents the residual operation) can effectively remove snow streaks. Meanwhile, we observe that it is impossible to remove all snow by directly applying the predicted snow mask to the snowy image as a residual. The reason is that the snowy image is not a simple addition of the snow mask and the clear image, and the predicted snow mask is just an enhanced schematic diagram of the position and shape of the snow in the image. Therefore, GF-Net adopts an adaptive way to guide the model to remove snow according to the information provided by the snow mask. In addition, to better preserve the features of the snow mask, we use the feature map of the snow mask as input rather than the snow mask itself.

As shown in Fig. 4, we also design a ResUnit (RU) for high-dimensional features encoding. Since snow has different effects in different regions, RU encodes these differences into features, making the model more adaptive. Meanwhile, we adopted a multi-level residual method to fuse the residual information of the shallow layer and deep layer to further improve the snow removal effect. Therefore, the adaptive residual block in GF-Net has two levels, and the residual of the i -th level can be expressed as

$$f_{res}^i = f_{snowy}^i - f_{mask}^i, \quad (9)$$

where f_{mask}^i and f_{snowy}^i represent the feature encoding of the snow mask and the snowy image at the i -th level, respectively.

The specific process of GF-Net is intuitively reflected in Fig. 3. Firstly, it performs the first-level adaptive coding on the input mask feature \mathcal{F}_{mask} and the snowy image \mathcal{I}_{snowy} to obtain f_{mask}^1 and f_{snowy}^1

$$f_{snowy}^1 = RU^{11}(\text{Conv}^1(\mathcal{I}_{snowy})), \quad (10)$$

$$f_{mask}^1 = RU^{12}(\text{Conv}^2(\mathcal{F}_{mask})), \quad (11)$$

where $RU^{ij}(\cdot)$ represents the j -th ResUnit of the i -th level. After that, the residuals of these two features are used to obtain the first-level residual f_{res}^1 . Secondly, the network performs

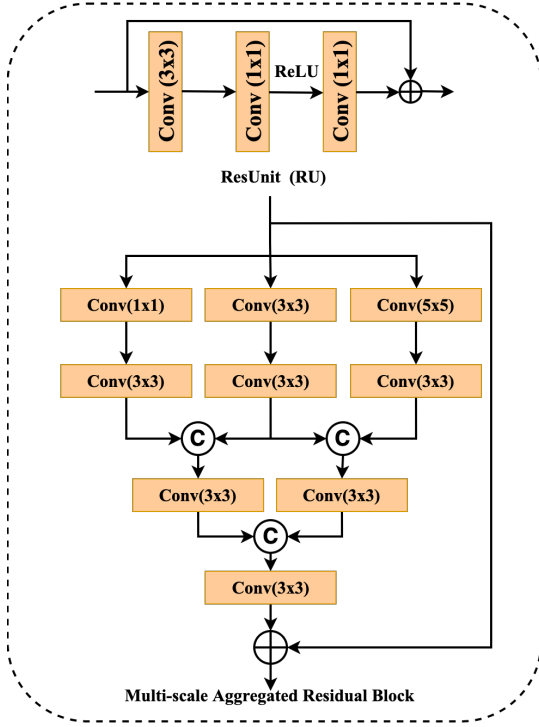


Fig. 4. The complete architecture of the proposed ResUnit (RU) and Multi-scale Aggregated Residual Block (MARB).

second-level adaptive coding on f_{mask}^1 and f_{snow}^1 to obtain f_{mask}^2 and f_{snow}^2

$$f_{snow}^2 = RU^{21}(f_{snow}^1), \quad (12)$$

$$f_{mask}^2 = RU^{22}(f_{mask}^1). \quad (13)$$

Similarly, the residuals of f_{mask}^2 and f_{snow}^2 are used to obtain the second-level residual f_{res}^2 . Finally, we encode and fuse f_{res}^1 and f_{res}^2 through two convolutional layers to obtain the final feature \mathcal{F}_{fuse} in a relatively snow-free state.

C. Reconstruct-Net

After GF-Net, we can obtain features \mathcal{F}_{fuse} that are close to snow-free. However, in addition to snow phenomena such as snowflakes and snow streaks, the veiling effect often appears in snow images. This phenomenon is similar to haze in that its effect on an object depends on its distance. Therefore, image snow removal must also include removal of the veiling effect. Not only that, there are also some imperfections and residual snow after GF-Net. To solve these problems, we also propose a Reconstruct-Net to correct these areas to ensure the quality of the reconstructed images.

As shown in Fig. 2, Reconstruct-Net consists of three Multi-scale Aggregated Residual Blocks (MARBs) and a convolutional layer. Among them, MARB (Fig. 4) is specially designed to capture the multi-scale features with different receptive fields, which is inspired by [47]–[49]. The haze effect caused by the veiling effect and the residual snow removal from the previous part of the network is unevenly distributed in the image. Using the multi-scale receptive field

of MARB can well remove and repair the pollution in these areas. Specifically, MARB consists of 1×1 , 3×3 , and 5×5 parallel convolutional layers, and all of them will pass through a 3×3 convolutional layer respectively. Then, the branch of the 3×3 kernel size will be connected with the branches of the 1×1 and 5×5 kernel size to form two new branches. After the 3×3 convolutional layer is performed on these two new branches, they will also be connected as a new branch, and a 3×3 convolutional layer will be performed again. Finally, the input of MARB is summed with output features to enhance the information flow. In Reconstruct-Net, a global residual connection is performed after the three MARBs are stacked to enhance feature representation. The operation of Reconstruct-Net can be defined as

$$\mathcal{I}'_{clear} = \mathcal{R}(\mathcal{F}_{fuse}), \quad (14)$$

where $\mathcal{R}(\cdot)$ denotes the Reconstruct-Net and \mathcal{I}'_{clear} is the reconstructed snow-free image.

D. Loss Function

During training, we use $L1$ loss as the reconstruct loss to minimize the difference between the reconstructed snow-free image and the corresponding clear image \mathcal{I}_{clear}

$$\mathcal{L}_{reconstruct} = \left\| \mathcal{I}'_{clear} - \mathcal{I}_{clear} \right\|_1. \quad (15)$$

In summary, Mask-Net, GF-Net, and Reconstruct-Net form the complete Snow Mask Guided Adaptive Residual Network (SMGARN). In this work, Mask-Net served as a sub-network to provide the snow mask and SMGARN adopts an end-to-end training mode. Therefore, the reconstruct loss $\mathcal{L}_{reconstruct}$ and mask loss \mathcal{L}_{mask} form the complete mask-assisted loss \mathcal{L}_{mas}

$$\mathcal{L}_{mas} = \mathcal{L}_{reconstruct} + \lambda \mathcal{L}_{mask}, \quad (16)$$

where λ is a hyper-parameter used to control the composition of the mask loss. To ensure that $\mathcal{L}_{reconstruct}$ and \mathcal{L}_{mask} receive equal attention, we set λ to 1 in this work.

IV. EXPERIMENTS

In this part, we provide relevant experimental details, descriptions, and results to verify the effectiveness and excellence of the proposed SMGARN.

A. Datasets

In this work, we trained our SMGARN with the CSD [8] train set (8000 images) for synthetic images and use the Snow100K [14] train set (50000 images) to train the model for real images. For evaluation, we use three benchmark test sets, including SRRS [15], CSD [8], and Snow100K [14]. Specifically, we adopted the last 2000 images in the SRRS test set for evaluation. For CSD, we use its 2000 test images for verification. In addition, we use the three test sets Snow100K-S, Snow100K-M and Snow100K-L provided by Snow100K for evaluation.

The training and testing datasets of the existing datasets (SRRS, CSD, and Snow100K) are all generated by synthetic techniques, so there are many abnormal images, such as



Fig. 5. Some sample images from the proposed SnowWorld24 dataset. All of these images were taken with different cameras on different snowy days.

snowflakes in summer or sunny days, and the effect of directly using them for evaluation is limited. Although Snow100K has proposed a real snow scene dataset, most of the images are extracted directly from web pages, and many images are not clear and have watermarks. More importantly, there are also a large number of post-processed artistic photos in this dataset. Since there are no labels and snow masks in the real scene, images need to be more clear and able to represent the complex situation of the real world. Therefore, Only the data provided by Snow100K cannot effectively evaluate the performance of the model. In order to further evaluate the performance and generalization of the snow removal model more comprehensively, we construct a new real snow image dataset, named **SnowWorld24**.

SnowWorld24. As shown in Fig. 5, SnowWorld24 contains 24 snow images of different scenes, which are directly obtained under snow weather without any processing, and there are no traces such as watermarks that are not related to the images themselves. Meanwhile, all of these images were taken with different cameras on different snowy days, and all of these scenes are from different countries and regions. The purpose of this is to increase the diversity of the snow images, so as to better verify the generalization of the model. In summary, SnowWorld24 is a real dataset specially proposed to verify model validity and generalization.

B. Implementation Detail

1) *Training Details*: Following previous works, we randomly extract 16 snowy patches with the size of 128×128 as inputs. Meanwhile, we augment the training data with flips and rotates operations to further improve the generalization ability of the model. The learning rate is initialized to 10^{-4} and halved every 100 epochs. We implement our model with the PyTorch framework and update it with the Adam optimizer.

2) *Model Details*: In our final model, we use three MARBs in the Reconstruct-Net. Meanwhile, we set the embedding dimension of the model to 112 and the kernel size of the convolutional layers in Fig. 2 are set to 3×3 . In addition, between the second and third convolutional layers of ResUnit,

TABLE I
QUANTITATIVE COMPARISON WITH OTHER ADVANCED METHODS ON SNOW100K-S, SNOW100K-M, AND SNOW100K-L. AMONG THEM, BEST RESULTS ARE **HIGHLIGHT** AND THE SECOND BEST RESULTS ARE *ITALIC*.

Method / Dataset	Snow100K-S	Snow100K-M	Snow100K-L
Zheng [10]	20.29 / 0.73	20.18 / 0.71	18.83 / 0.66
DehazeNet [26]	22.06 / 0.78	21.54 / 0.74	20.19 / 0.70
HDCWNet [26]	22.13 / 0.78	21.66 / 0.72	19.91 / 0.69
DerainNet [39]	25.74 / 0.86	23.36 / 0.84	19.18 / 0.74
DeepLab [50]	25.94 / 0.87	24.36 / 0.85	21.29 / 0.77
JORDER [43]	25.62 / 0.88	24.97 / 0.87	23.40 / 0.80
DuRN-S-P [51]	32.27 / 0.94	30.92 / 0.93	27.21 / 0.88
DeSnowNet [14]	32.23 / 0.95	30.86 / 0.94	27.16 / 0.89
DS-GAN [16]	33.43 / 0.96	31.87 / 0.95	28.06 / 0.92
All in One [52]	- / -	- / -	28.33 / 0.88
SMGARN(Ours)	34.46 / 0.95	33.22 / 0.95	29.44 / 0.93

we expand the feature embedding dimension by 2 times, and the channel of the final output will be reduced to the original dimension.

C. Comparisons with State-of-the-Art Methods

In this section, we compare our proposed SMGARN with other advanced methods to verify the validity of the proposed model. Since image snow removal is an emerging field, the test data and test method of some method are confusing. In this work, we unify multiple classic snow removal models, and all of which are trained on the same dataset and conditions. It is worth noting that we re-produced or re-tested some models whose experimental details were unclear. Besides, due to the commonality of image restoration and atmospheric phenomena, we also provide the results of classic models for image restoration (SRCNN [53]), image dehazing (DehazeNet [26]), and deraining (RESCAN [54]). All these methods are trained under the same experimental setting and training set. For the validation method, we adopt PSNR and SSIM to evaluate the final results.

1) *Quantitative Comparison*: In Tables I and II, we show the quantitative comparison between our method and other advanced methods. Among them, TABLE I shows the results

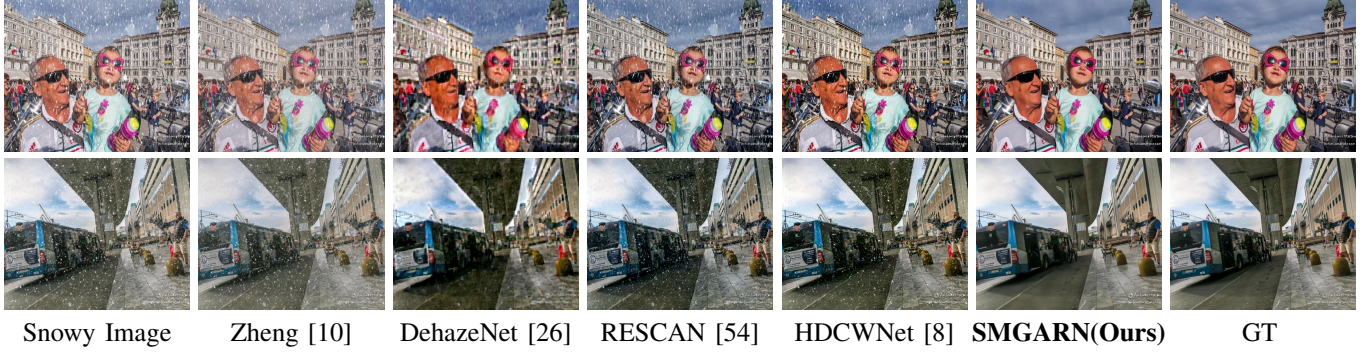


Fig. 6. Visual comparison with other advanced models on Snow100K. Obviously, our proposed method can reconstruct high-quality snow-free images.

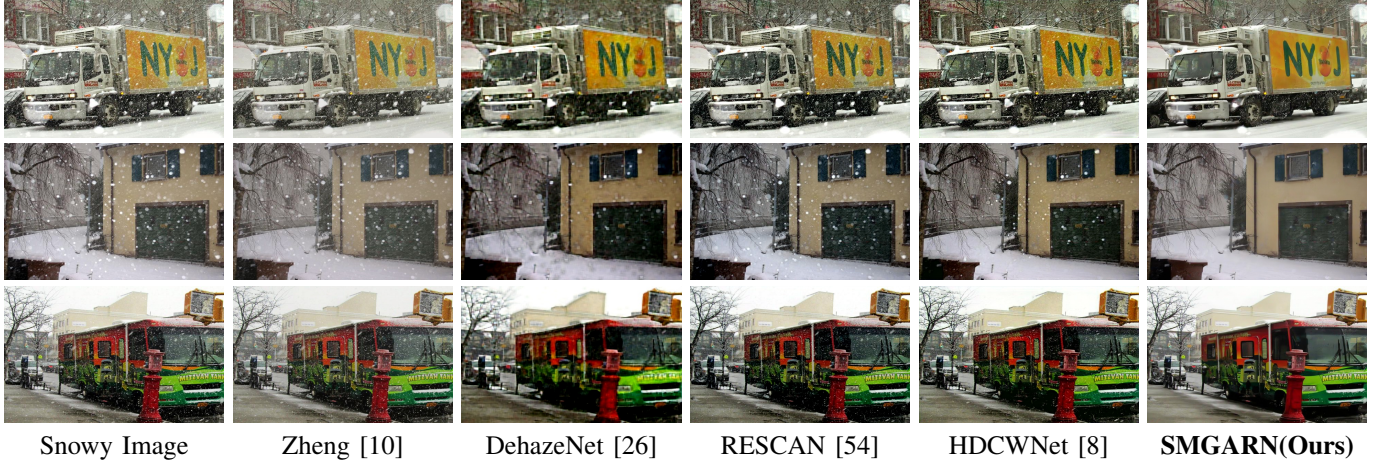


Fig. 7. Visual comparison on real snow images from Snow100K. Obviously, our proposed method can reconstruct clear images with less snow residue.

TABLE II
QUANTITATIVE COMPARISON WITH OTHER ADVANCED METHODS ON CSD AND SRRS. AMONG THEM, BEST RESULTS ARE **HIGHLIGHT** AND THE SECOND BEST RESULTS ARE *ITALIC*.

Method / Dataset	CSD [8]	SRRS [15]
Zheng [10]	14.21/0.61	16.34/0.69
SRCNN [53]	22.25/0.82	22.46/0.85
DehazeNet [26]	20.91/0.72	21.22/0.75
RESCAN [54]	22.11/0.81	22.79/0.86
DeSnowNet [14]	20.13/0.81	20.38/0.84
CycleGAN [55]	20.98/0.80	20.21/0.74
DAD [56]	24.33/0.85	24.31/0.86
All in One [52]	26.31/0.87	24.98/0.88
JSTASR [15]	27.96/0.88	-
HDCWNet [8]	28.62/0.89	25.03/0.89
SMGARN (Ours)	29.94/0.94	25.43/0.92

of these methods on three subsets of the Snow100K test set (Snow100K-S, Snow100K-M, Snow100K-L). According to the table, we can observe that the PSNR performance of our method are significantly outperforms previous methods. Compared with DS-GAN [16], SMGARN improves PSNR by **1.03dB** and **1.35dB** in Snow100K-S and Snow100K-M, respectively. Notably, on the hardest Snow100K-L, our method achieves a PSNR improvement of **1.11dB** over the previous best model (All in One [52]). It can be seen that our model

improves the PSNR results more than **1dB** compared to the current SOTA model on these three datasets. This fully demonstrated that the proposed SMGARN can significantly improve the snow removal performance.

Since Snow100K does not take into account the influence of the veiling effect, we tested it on the other snow removal datasets, CSD and SRRS, to further evaluate the proposed method. In TABLE II, we provide PSNR and SSIM results of these methods. It is clear that SMGARN outperforms all previous methods. Especially on the CSD test set, our proposed method improves the PSNR performance by **1.32dB**, which is a huge improvement. All these results further verify the effectiveness of the proposed SMGARN.

2) *Visual Comparison*: In Figs. 6 and 8, we provide the visual comparison with other advance models on Snow100K [14], CSD [8], and SRRS [15], respectively (please zoom in to see details). According to Fig. 6, we can clearly observed that our SMGARN can reconstruct high-quality snow-free image very close to GT. According to Fig. 8, we can found that the snow-free images reconstructed by our SMGARN are clearer and have less snow residue. Specifically, the first 4 rows of results in Fig. 8 show that SMGARN can effectively eliminate the interference of snow streaks and fine snowflakes on the image. Rows 5 and 6 show that our method is more helpful in removing large snowflakes without creating large areas of shading.

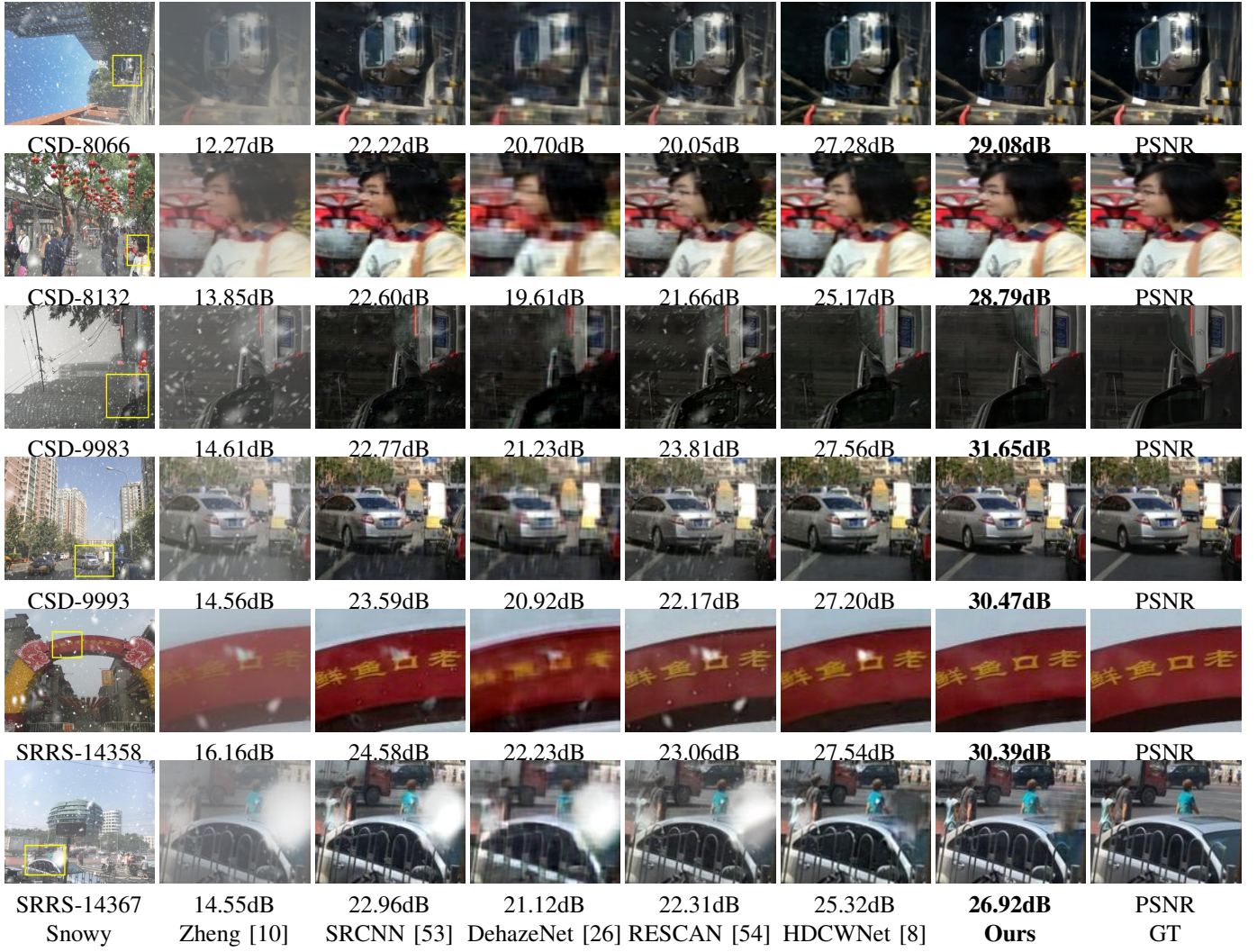


Fig. 8. Visual comparison with other advanced models on CSD and SRRS. Obviously, our SMGARN can reconstruct high-quality snow-free images.

Further, we demonstrate the snow removal capability of SMGARN in real scene in Fig. 7 to verify the generalization ability of the proposed method (please zoom in to see details). From the figure, we can clearly see that SMGARN can remove both large particles of snow and dense snowflakes well. This is benefit from the proposed Mask-Net, which can predict accurate location and shape of the snow. With the help of the predicted snow mask, the images reconstructed by our SMGARN have almost no snow and texture details can be better preserved.

3) *Model Size and Performance Comparison:* We also provide the trade-off analysis of the proposed SMGARN and other classic snow removal methods in the number of parameters and performance in Fig. 9. Among them, JSTASR also uses the information of the snow mask for snow removal, but its parameters quantity is 10 times that of our SMGARN. Meanwhile, our method improves the performance of PSNR and SSIM by 1.98dB and 0.06 respectively compared to JSTASR. Compared to HDCWNet, our method improves snow removal performance by more than 1dB with fewer parameters. The above results fully demonstrate that the pro-

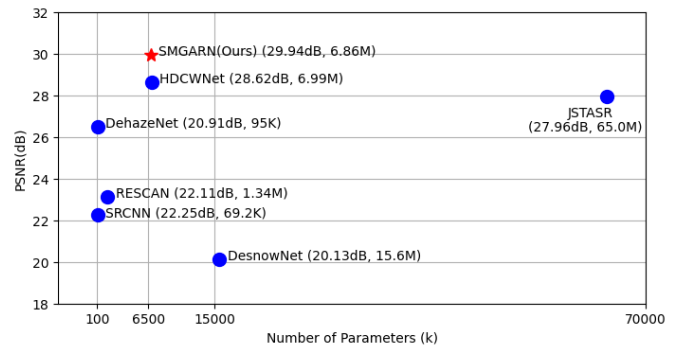


Fig. 9. Model Size and Performance Comparison. The results in the figure show that SMGARN achieves a more perfect balance between model parameters and performance.

posed SMGARN can achieve better performance with few parameters. Therefore, we can draw a conclusion that the proposed SMGARN achieves a good balance between the model performance and parameter quantity.

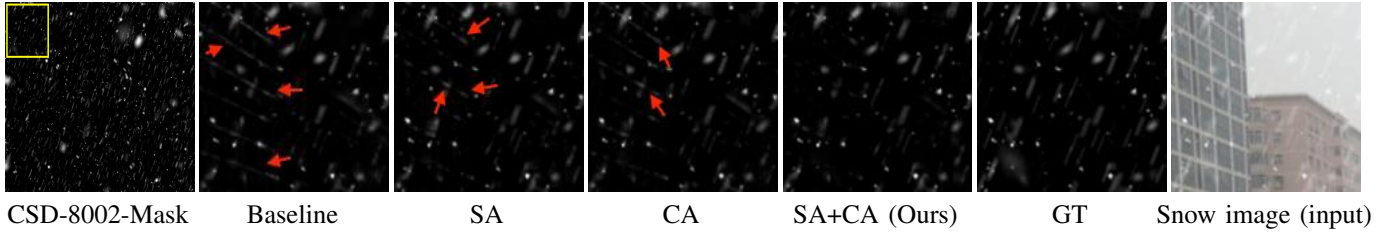


Fig. 10. The effect of different attention mechanism on the quality of the snow mask predicted by Mask-Net. Obviously, lacking either SA or CA, some areas (marked by red arrows) will be disturbed by the edge of the building in the original image, resulting in the wrong snow pattern in the predicted snow mask.



Fig. 11. The effect of snow mask guidance on snow-free images reconstructed by the model. Due to the lack of guidance by the snow mask, there are residual snow patterns in Case1 and Case2 (marked by red arrows).

V. INVESTIGATION

This section conducts a comprehensive ablation study on all modules of SMGARN to verify the effect of different modules on snow removing. We use PSNR and SSIM as evaluation metrics, all experiments are evaluated on the CSD dataset, and a patch size of 64×64 is used for training. In addition to the ablation studies, we also investigate multiple aspects of the model to fully validate the effectiveness of the model.

A. Study on Mask-Net

As a lightweight snow mask prediction network, Mask-Net plays an important role in extracting snow information. In order to verify the rationality of the Mask-Net, we evaluate the effect of SA (Self-Pixel Attention) and CA (Cross-Pixel Attention) in Mask-Net from both visual and quantization perspectives. In Fig. 10, we compare the effect of different attention mechanisms on the snow mask predicted by Mask-Net. From the figure we can clearly see that the results predicted by Mask-Net with both SA and CA are very close to the ground-truth snow mask. On the contrary, if the Mask-Net without any attention unit or composed of only one attention unit, the generated snow mask will be disturbed by other objects in the image. Meanwhile, we also provide the quantitative comparison in TABLE III. According to the table, we can observed that SA and CA can significantly promote the final snow removal performance of the model. In addition, the performance of the model can be further improved when both SA and CA are adopted. This fully demonstrated the rationality of the Mask-Net.

Moreover, we provide the effect of snow mask on model performance in Fig. 11 and TABLE IV. Specifically, we use the model without Mask-Net as the baseline model, denoted as Case 1. Case 2 represents the model after removing \mathcal{L}_{mask} and Case 3 is the final SMGARN model. For Case 4, we directly use the ground truth snow mask as a guide to further verify

TABLE III
STUDY OF TWO PIXEL ATTENTION MECHANISMS IN MASK-NET.

Model Case	SA	CA	PSNR
Mask-Net-baseline	×	×	28.83
Mask-Net-SA	✓	×	28.88
Mask-Net-CA	×	✓	28.90
Mask-Net-CASA	✓	✓	29.01

TABLE IV
STUDY ON THE IMPORTANCE OF SNOW MASK. AMONG THEM, BEST RESULTS ARE **HIGHLIGHT** AND THE SECOND BEST RESULTS ARE *ITALIC*.

Metric	Case1	Case2	Case3 (Ours)	Case4
PSNR/SSIM	28.62/0.93	28.92/0.93	<i>29.01/0.93</i>	30.07/0.95

the effect of snow mask on the snow removal ability of the model. These results show that the snow mask plays a crucial role in model performance, and the quality of the snow mask will determine the effectiveness of snow removal. The higher the quality of the provided snow mask, the better the model performance. **This experiment proves the decisive role of the snow mask in the image snow removal task, making up for the deficiency of the previous research.**

B. Study on GF-Net

As the key structure of SMGARN, GF-Net plays an important role in guiding snow mask to remove snow. In TABLE V, we investigate the effectiveness of GF-Net. Specifically, in Case 1, we concatenate the snow mask with the features of the snow image and replace the GF-Net with 8 convolutional layers. In Case 2, we replace the feature connection operation with the residual operation based on Case 1. It can be seen that since the snow image is a weighted combination of the snow mask and the clean image, the performance of Case 2 is significantly higher than that of Case 1. To further verify the

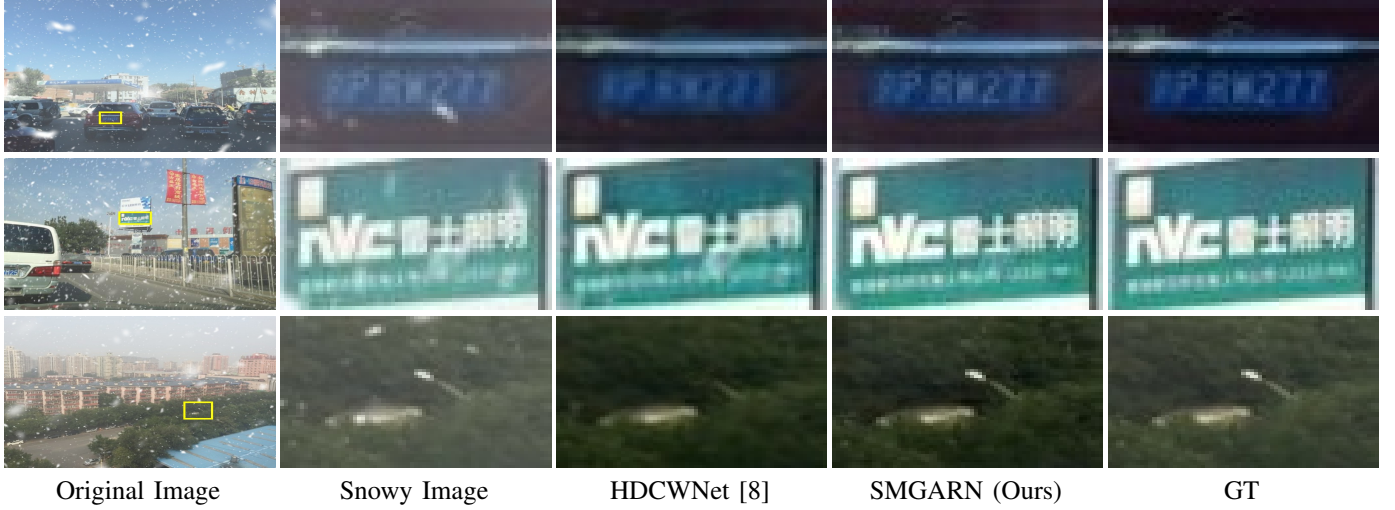


Fig. 12. Comparison with HDCWNet on the quality of snow-free images. Obviously, our SMGARN can reconstruct more clear images.

TABLE V
INVESTIGATION OF GF-NET.

Metric	Case1	Case2	Case3	Case4
PSNR/SSIM	27.22/0.91	27.31/0.91	28.39/0.92	29.01/0.93

TABLE VI
COMPARISON OF DIFFERENT NETWORK STRUCTURES IN THE PROPOSED MARB.

Model Case	Single-Scale	Multi-Scale	Single-Agg	Multi-Agg	PSNR
MARN-SS-SA	✓	×	✓	×	28.64
MARN-MS-SA	×	✓	✓	×	28.75
MARN-SS-MA	✓	×	×	✓	28.72
MARN-MS-MA	×	✓	×	✓	29.01

rationality of GF-Net, we replace the adaptive residual in the original GF-Net with the connection operation and represent it as Case 3. Compared with the original model Case 4, we can find that the performance of Case 3 is severely degraded. This means that the residual operation of adaptive coding is more in line with the snow image generation mechanism, and has better snow removal ability in high-dimensional feature space, which is very beneficial to performance improvement. ***In summary, Case 1 and Case 2 prove the effectiveness of residual operation for snow removal. Case 3 and Case 4 demonstrate that adaptive residual structure is more beneficial to reconstruct snow-free images than feature connections.*** This experiment fully illustrate the effectiveness and necessity of the proposed GF-Net.

C. Study on Reconstruct-Net

In this work, we propose a Reconstruct-Net to reconstruct the final snow-free image by the specially designed MARBs. In In TABLE VI, we provide the results for several different MARB designs. Single-Scale (SS) means that the three input branches use 3×3 convolutional layers for feature extraction in MARB. Multi-Scale (MS) means that MARB uses convo-

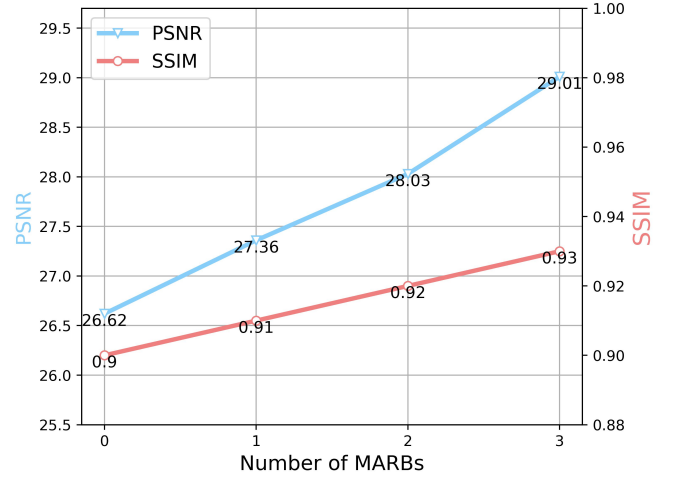


Fig. 13. Ablation study on different number of MARBs.

lutional layers with kernels of 1×1 , 3×3 , and 5×5 to process features in parallel. Single-Agg (SA) means that only one join operation is applied in MARB. Multi-Agg (MA) means that MARB aggregates features from different branches multiple times. It can be seen from the table that the multi-scale and multi-aggregation design can significantly improve the model performance, which proves the effectiveness of MARB.

To verify the effect of MARB on the model performance, we also tested the model with different numbers of MARBs. According to Fig. 13, we can clearly observe that the performance of the model increases as the number of MARBs increases. This illustrates the effectiveness of MARB and the potential of Reconstruct-Net. However, we also noticed that as the number of MARBs increases, the number of parameters of the model also increases. Therefore, we used three MARBs in the final model to achieve a good balance between the model performance and size.

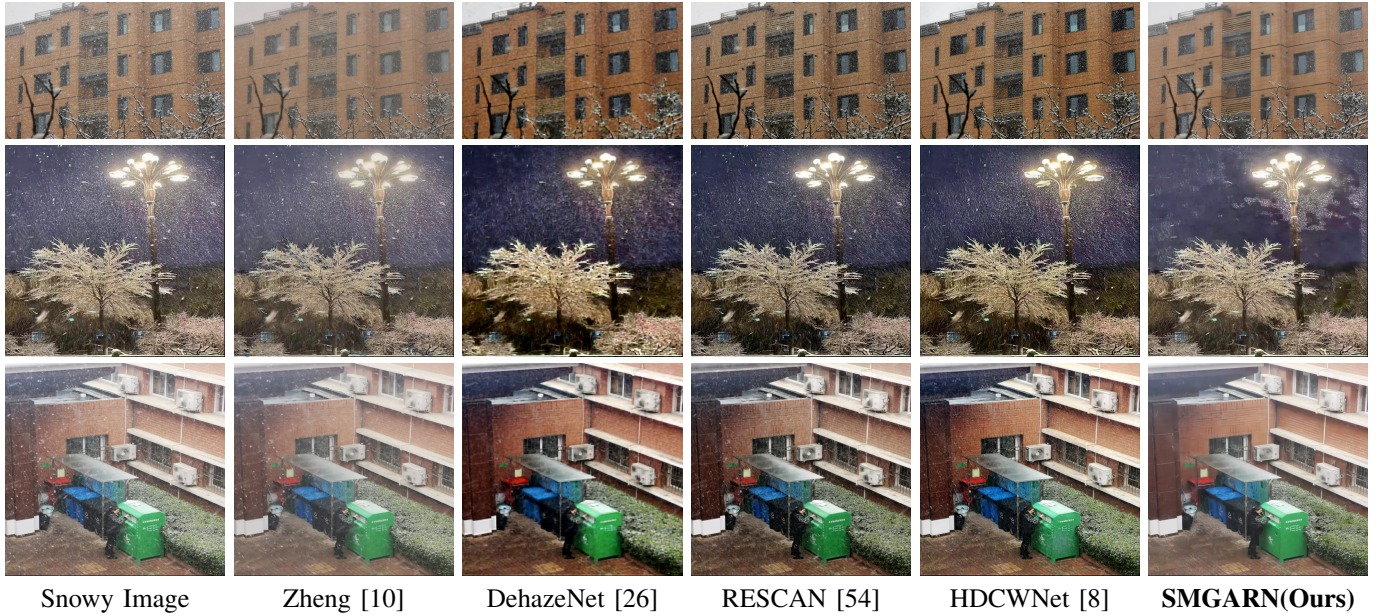


Fig. 14. Visual comparison on real snow images from SnowWorld24. Obviously, our proposed method can reconstruct clear images with less snow residue.

D. Image Quality Comparison

For the image snow removal task, it is not only necessary to eliminate the interference of snow in the image, but also to retain the information in the original image to the maximum extent possible for the reconstructed image. In this part, we evaluate the image quality reconstructed by the proposed method to further verify the effectiveness of SMGARN. In TABLE VII, we show the NIQE results of our SMGARN and other advanced methods. NIQE is a fully blind image quality analyzer that can be used to assess the authenticity of images. ***It is worth noting that lower NIQE values represent better perceptual quality.*** According to the table, we can found that our method has a lower NIQE, proving that it can generate higher quality snow-free images. In addition, we also evaluated the image quality from the visual aspect. According to the first two lines of images in Figure 12, it can be clearly seen that the images generated by SMGARN well retain the clear text information in the original images. In contrast, the images generated by HDCWNet blur the details of text regions, which affects the image quality. In addition, the last line of Figure 12 shows that SMGARN can accurately distinguish the original information of the image and the snow, and will not remove other objects as snow particles. This further demonstrated that our proposed SMGARN can generate high-quality images and have better robustness to be compatible with more scenes.

E. Study on SnowWorld24

To further verify the snow removal ability of the model in the real snow scene, we test SMGARN on our proposed SnowWorld24. SnowWorld24 is real snow image dataset that can fully verify the model performance in the real scene. According to Fig. 14, we can clearly see that HDCWNet can only remove part of the snow with large particles, and can do nothing for dense snowflakes. Although HDCWNet

TABLE VII
PERFORMANCE COMPARISON ON REAL-WORLD DATASET. ↓ INDICATES THAT THE LOWER THE NIQE VALUE, THE HIGHER THE QUALITY OF THE RECONSTRUCTED IMAGE.

Model	SnowWorld24 (NIQE↓)	Snow100K (NIQE↓)
Zheng [10]	4.3762	3.9886
DehazeNet [26]	4.6954	4.5731
RESCAN [54]	3.0795	3.3893
HDCWNet [8]	3.8623	4.0448
SMGARN(Ours)	2.9371	3.0751

TABLE VIII
QUESTIONNAIRE RESULTS. LARGE SNOWFLAKE, DENSE SNOWFLAKES, AND QUALITY SCORES RANGE FROM 1 TO 5, WITH HIGHER SCORES INDICATING BETTER IMAGES. THE FIRST AND SECOND METRICS REPRESENT THE ABILITY OF THE MODEL TO REMOVE LARGE AND DENSE SNOWFLAKES, RESPECTIVELY. QUALITY REPRESENTS THE CLARITY AND FIDELITY OF THE IMAGE.

Model Name	Big Snowflake	Dense Snowflakes	Quality
DehazeNet [26]	2.38	1.74	0.46
RESCAN [54]	2.75	2.06	1.21
HDCWNet [8]	3.41	2.96	3.16
SMGARN(Ours)	4.28	3.84	3.92

can improve the contrast of the image, it does not achieve the original intention of removing snow from the image. In contrast, our SMGARN can remove almost snow in the image and fully retain the texture information of the image. All the above results further illustrate the effectiveness and robustness of the proposed SMGARN.

F. User Study

To evaluate the proposed model more objectively from the visual aspect, we conduct a user study on real snow images. Specifically, we collected 80 images from the Snow100K real

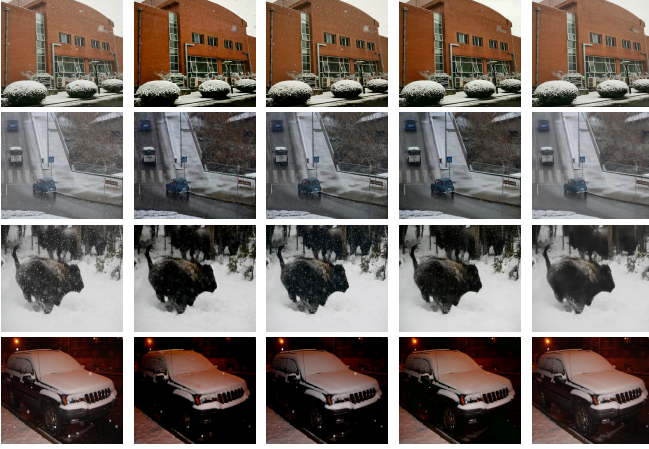


Fig. 15. Examples of real snow image categories in the questionnaire survey and the snow removal results corresponding to SMGARN. Column 1 represents snow images, and columns 2 to 5 correspond to DehazeNet, RESCAN, HDCWNet and our proposed SMGARN, respectively.

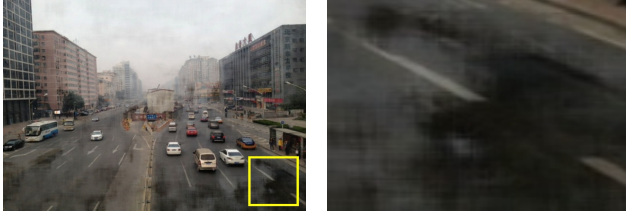


Fig. 16. Example of our reconstructed image with artifacts.

snow dataset and our proposed SnowWorld24 dataset, and invited 25 participants for evaluation. We set three evaluation indicators as the basis for the participants to score, namely, the ability to remove large snowflakes, the ability to remove dense snowflakes, and the image quality. The 80 test images cover various scenes in the real world, including buildings, streets, crowds, and animals. In Fig. 15, we shows four sets of examples. In addition, we evaluate DehazeNet, RESCAN, HDCWNet, and our SMGARN as objects. Specific results are listed in TABLE VIII. In the survey, more than half of the participants believed that the snow-free images reconstructed by SMGARN are closer to the real situation. Some participants believed that the snow-free image of SMGARN will not destroy the scene in the original image, such as the fourth row in Fig. 15. Meanwhile, the snow-free image generated by the four methods only SMGARN retains the details of the wheel hub in the lower right corner of the image. Therefore, in TABLE VIII, our method achieves the highest score. This proves that SMGARN has excellent performance in real scenes and can greatly preserve the original information of the image.

G. Application on High-Level Tasks

Image snow removal aims to restore clear images from images disturbed by snow phenomenon. It is not only a part of image processing, but also serves as a data preprocessing method for other vision tasks. To demonstrate that our proposed SMGARN can benefit high-level computer vision tasks, we use Google Vision API to evaluate the snow removal results. As a new evaluation method, it has been applied to

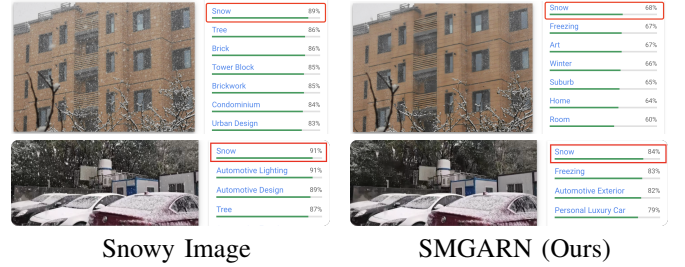


Fig. 17. The desnowing results are tested on the Google Vision API (snow classification). Please zoom in to see the details.



Fig. 18. The desnowing results are tested on the Google Vision API (object detection). Please zoom in to see the details.

image rain and snow removal to verify model performance from an application point of view. From Fig. 17, it can be seen that the probability of SMGARN-processed images being judged as snow drops significantly. Meanwhile, according to Fig. 18, the images reconstructed by our SMGARN can improve the accuracy of the object detection model. The confidence represents the probability of snowy weather and the probability of target, respectively. According to these results, we can observe that the results generated by our SMGARN effectively promote the performance of high-level vision tasks, which further demonstrates the effectiveness of SMGARN.

VI. DISCUSSION

Although our SMGARN has obvious improvement in both quantification performance and visual performance, there are artifacts that cannot be removed in some images heavily affected by the veiling effect, as shown in Fig. 16. This is because SMGARN lacks sufficient constraints for the haze phenomenon caused by the veiling effect. However, if too many constraints are imposed, the phenomenon of image blurring in HDCWNet will occur. Therefore, we believe that image desnowing should be regarded as an image restoration task covering a variety of extreme weather phenomena, which needs to be modeled by combining the characteristics of image dehazing and image deraining. Meanwhile, a new training dataset is also critical to the image snow removal field. In future works, we will make more contributions to this field.

VII. CONCLUSION

In this paper, we proposed a Snow Mask Guided Adaptive Residual Network (SMGARN) for image snow removal. Specifically, we build a Mask-Net to directly predict the snow mask from the snowy image with the help of Self-pixel

Attention (SA) and Cross-pixel Attention (CA) mechanisms. Meanwhile, a multi-level Guidance-Fusion Network (GF-Net) is designed to adaptively remove snow from the image with the guidance of the predicted snow mask. In addition, a Reconstruct-Net is proposed to reconstruct the final snow-free image by the specially designed Multi-scale Aggregated Residual Blocks (MARBs). Extensive experimental results show that the proposed SMGARN achieves excellent results on multiple datasets, and also shows great potential in practical applications. Although the current models and strategies have achieved excellent results, further improvements can be made in terms of visual presentation and training methods. In the future, we will explore better image prior guidance strategies to further improve model performance.

REFERENCES

- [1] Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, and Jon Shlens. Stand-alone self-attention in vision models. *NeurIPS*, 32, 2019.
- [2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV*, pages 213–229, 2020.
- [4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.
- [5] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *CVPR*, pages 5203–5212, 2020.
- [6] Guangwei Gao, Yi Yu, Jian Yang, Guo-Jun Qi, and Meng Yang. Hierarchical deep cnn feature set-based representation learning for robust cross-resolution face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [7] Weipeng Hu and Haifeng Hu. Orthogonal modality disentanglement and representation alignment network for nir-vis face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6):3630–3643, 2021.
- [8] 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 *ICCV*, pages 4196–4205, 2021.
- [9] Jing Xu, Wei Zhao, Peng Liu, and Xianglong Tang. An improved guidance image based method to remove rain and snow in a single image. *Computer and Information Science*, 5(3):49, 2012.
- [10] Xianhui Zheng, Yinghao Liao, Wei Guo, Xueyang Fu, and Xinghao Ding. Single-image-based rain and snow removal using multi-guided filter. In *ICONIP*, pages 258–265, 2013.
- [11] Soo-Chang Pei, Yu-Tai Tsai, and Chen-Yu Lee. Removing rain and snow in a single image using saturation and visibility features. In *ICME Workshop*, pages 1–6, 2014.
- [12] 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.
- [13] Shujian Yu, Yixiao Zhao, Yi Mou, Jinghui Wu, Lu Han, Xiaopeng Yang, and Baojun Zhao. Content-adaptive rain and snow removal algorithms for single image. In *International Symposium on Neural Networks*, pages 439–448, 2014.
- [14] 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.
- [15] 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 *ECCV*, pages 754–770, 2020.
- [16] Da-Wei Jaw, Shih-Chia Huang, and Sy-Yen Kuo. Desnowgan: An efficient single image snow removal framework using cross-resolution lateral connection and gans. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(4):1342–1350, 2020.
- [17] Robby T Tan. Visibility in bad weather from a single image. In *CVPR*, pages 1–8, 2008.
- [18] Bo-Hao Chen and Shih-Chia Huang. Edge collapse-based dehazing algorithm for visibility restoration in real scenes. *Journal of Display Technology*, 12(9):964–970, 2016.
- [19] Raanan Fattal. Single image dehazing. *ACM Transactions on Graphics*, 27(3):1–9, 2008.
- [20] Raanan Fattal. Dehazing using color-lines. *ACM Transactions on Graphics*, 34(1):1–14, 2014.
- [21] Pengyu Wang, Hongqing Zhu, Hui Huang, Han Zhang, and Nan Wang. Tms-gan: A twofold multi-scale generative adversarial network for single image dehazing. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5):2760–2772, 2021.
- [22] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.
- [23] Shih-Chia Huang, Bo-Hao Chen, and Yi-Jui Cheng. An efficient visibility enhancement algorithm for road scenes captured by intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 15(5):2321–2332, 2014.
- [24] Bo-Hao Chen and Shih-Chia Huang. An advanced visibility restoration algorithm for single hazy images. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 11(4):1–21, 2015.
- [25] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *CVPR*, pages 3194–3203, 2018.
- [26] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.
- [27] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169, 2016.
- [28] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, pages 1674–1682, 2016.
- [29] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018.
- [30] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *CVPR*, pages 8202–8211, 2018.
- [31] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.
- [32] Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In *ACCV*, pages 203–215, 2018.
- [33] Cosmin Ancuti, Codruta O Ancuti, and Radu Timofte. Ntire 2018 challenge on image dehazing: Methods and results. In *CVPR Workshops*, pages 891–901, 2018.
- [34] Qiaosi Yi, Juncheng Li, Faming Fang, Aiwen Jiang, and Guixu Zhang. Efficient and accurate multi-scale topological network for single image dehazing. *IEEE Transactions on Multimedia*, 2021.
- [35] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4):1742–1755, 2011.
- [36] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *ICCV*, pages 3397–3405, 2015.
- [37] Chang-Hwan Son and Xiao-Ping Zhang. Rain removal via shrinkage of sparse codes and learned rain dictionary. In *ICME Workshops*, pages 1–6, 2016.
- [38] Duan-Yu Chen, Chien-Cheng Chen, and Li-Wei Kang. Visual depth guided color image rain streaks removal using sparse coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(8):1430–1455, 2014.
- [39] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017.
- [40] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *CVPR*, pages 8405–8414, 2019.
- [41] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *CVPR*, pages 3877–3886, 2019.

- [42] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *CVPR*, pages 8346–8355, 2020.
- [43] Wenhao Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *CVPR*, pages 1357–1366, 2017.
- [44] Qiaosi Yi, Juncheng Li, Qinyan Dai, Faming Fang, Guixu Zhang, and Tiejong Zeng. Structure-preserving deraining with residue channel prior guidance. In *ICCV*, 2021.
- [45] Jérémie Bossu, Nicolas Hautiere, and Jean-Philippe Tarel. Rain or snow detection in image sequences through use of a histogram of orientation of streaks. *International Journal of Computer Vision*, 93(3):348–367, 2011.
- [46] Dhanashree Rajderkar and PS Mohod. Removing snow from an image via image decomposition. In *ICECCN*, pages 576–579, 2013.
- [47] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [48] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision (ECCV)*, pages 517–532, 2018.
- [49] Juncheng Li, Faming Fang, Jiaqian Li, Kangfu Mei, and Guixu Zhang. Mdcn: Multi-scale dense cross network for image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(7):2547–2561, 2020.
- [50] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
- [51] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019.
- [52] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *CVPR*, pages 3175–3185, 2020.
- [53] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015.
- [54] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, pages 254–269, 2018.
- [55] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 825–833, 2018.
- [56] Zhengxia Zou, Sen Lei, Tianyang Shi, Zhenwei Shi, and Jieping Ye. Deep adversarial decomposition: A unified framework for separating superimposed images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12806–12816, 2020.