

JSTASR: Joint Size and Transparency-Aware Snow Removal Algorithm Based on Modified Partial Convolution and Veiling Effect Removal

Wei-Ting Chen^{1*}, Hao-Yu Fang², Jian-Jiun Ding², Cheng-Che Tsai¹, and Sy-Yen Kuo^{2**}

¹ Graduate Institute of Electronics Engineering, National Taiwan University, Taipei 106, Taiwan

² Department of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan

<https://github.com/weitingchen83/JSTASR-DesnowNet-ECCV-2020>

{jimmy3505090, danielfang60609}@gmail.com, {jjding, r08943148, sykuo}@ntu.edu.tw

Abstract. Snow removal usually affects the performance of computer vision. Comparing with other atmospheric phenomenon (e.g., haze and rain), snow is more complicated due to its transparency, various size, and accumulation of veiling effect, which make single image de-snowing more challenging. In this paper, first, we reformulate the snow model. Different from that in the previous works, in the proposed snow model, the veiling effect is included. Second, a novel joint size and transparency-aware snow removal algorithm called JSTASR is proposed. It can classify snow particles according to their sizes and conduct snow removal in different scales. Moreover, to remove the snow with different transparency, the transparency-aware snow removal is developed. It can address both transparent and non-transparent snow particles by applying the modified partial convolution. Experiments show that the proposed method achieves significant improvement on both synthetic and real-world datasets and is very helpful for object detection on snow images.

Keywords: Transparency and size-aware snow removal, Snow detection, Differentiable dark channel prior

1 Introduction

Snow is an atmospheric phenomenon which usually obstructs the visibility and degrades the image quality severely. Therefore, snow removal is significant for several computer vision missions. Compared with other atmospheric phenomenon, due to the non-transparency property, snow removal is even more challenging because it may occlude the background, which leads to the loss of information. In

* Indicates equal contribution.

** Corresponding author: Sy-Yen Kuo.

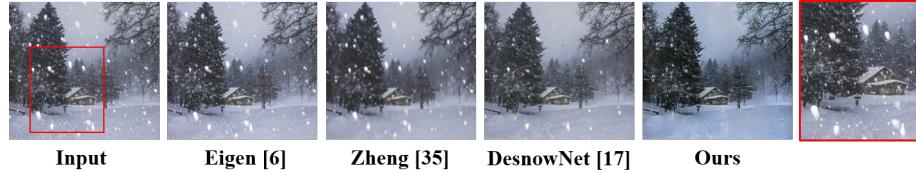


Fig. 1. Comparison of the proposed method with state-of-the-art methods [6,35,17]. The proposed method can achieve superior performance in both snow removal and veiling effect recovery. The red bounding box demonstrates the complicated snow scenario. (i.e. the snow particles with different sizes, transparencies, and the veiling effect)

the open literature, many snow removal algorithms [28,35,19,17,27,32] have been developed for decades. Xu *et al.* [28] proposed a snow flake removal algorithm using the guided filter without any pixel-based statistical information. Zheng *et al.* [35] proposed to use the multi-guided filter to separate the snow part from background. Pei *et al.* [19] utilized the features based on the color information for snow removal. In [17], a learning-based snow removal algorithm was proposed and the formation of snow was modeled by

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

where K is snow image acquired by the camera, J is the clean image, Z is the snow mask, C denotes the chromatic aberration map for snow images. Although these algorithms have good performance in some scenarios, they may have several problems which limit the performance. First, for the conventional snow removal strategies, the over-smoothness on the snow-free region may be produced because the snow region is not identified in prior. Second, owing to the complicated formation of the snow image, the conventional snow removal algorithms fail to address the crucial factor in real snow images, such as the veiling effect caused from haze veil and the accumulation of tiny snow particles. Third, unlike haze or rain, which are transparent, snow particles may be non-transparent and the background information is lost. Last, snow images usually consist of various size of particles, however, the existing methods do not address this problem, which leads to the limited recovered performance.

To address these limitations, we develop a novel snow removal model which can cope with the complex snow scenes. The main contributions of this paper are summarized as follows:

1. First, a novel de-snowing model called the **Joint Size and Transparency-Aware Snow Removal (JSTASR)** is proposed. It applies three stages: (i) veiling effect removal, (ii) the size-aware snow identifier, and (iii) transparency-aware snow removal. The proposed architecture can retrieve snow information (i.e., size, location and transparency), remove snow, and remove the veiling effect jointly.
2. A new snow formation model is proposed based on the observation of snow images in real world. The snow region mask and the veiling effect are included

in this model. In order to train our network in this scenario, a large scale snow dataset called **Snow Removal in Realistic Scenario (SRRS)** is proposed.

3. The size-aware snow identifier, discriminator and loss function is developed to address the complicated snow scenarios according to the snow size. With these modules, the performance of snow removal can be much improved.
4. In order to address particles with different transparency (i.e. transparent and non-transparent snow), the modified partial convolution operation and the transparency-aware snow removal algorithm are proposed.
5. To deal with the veiling effect and increase the visual quality of the recovered result, a differentiable dark channel prior layer and an atmospheric light prediction module are proposed and embedded into the network for fully end-to-end learning in the de-snowing pipeline.

In Fig 1, the de-snowing results recovered by the proposed JSTASR are compared with other state-of-the-art algorithms. One can see that the proposed method can achieve more desirable results compared with other methods. More experimental results will be shown in Section 4.

2 Related Work

2.1 Single Image Snow Removal Algorithm

For snow removal in the single image [28,35,19,17,27,32], Zheng *et al.* [35] investigated the difference between snow streaks and clear background edges. By this statistical information, the multi-guided filter is applied to remove the snow flake. Wang *et al.* [27] proposed a three-layer hierarchical scheme which combines image decomposition and dictionary learning. Voronin *et al.* [26] developed the anisotropic gradient in Hamiltonian quaternions to remove rain and snow. Li *et al.* [15] applied the generative adversarial network (GAN) for snowflake removal. Liu *et al.* [17] proposed a learning-based snow removal architecture called the DesnowNet.

2.2 Single Image Haze Removal Algorithm

Although the proposed algorithm is related to snow removal, since it also adopts the dark channel prior (DCP), which is a dehazing technique, several existing image dehazing algorithms are also reviewed here. For the single image dehazing, several haze and smoke removal algorithms have been proposed for a decade [10,25,36,1,2,3,22,13,4,5]. He *et al.* [10] investigated haze-free images in nature and proposed the dark channel prior to compute the transmission value. Berman *et al.* [1] proposed a non-local image dehazing algorithm. For learning-based algorithms, Cai *et al.* [2] computed the transmission map by a learning architecture called DehazeNet. Chen *et al.* [3] proposed a new feature called the patch map to select the patch size adaptively and addressed the color distortion problem of the DCP. Ren *et al.* [22] proposed the multi-scale CNN to predict the transmission map.

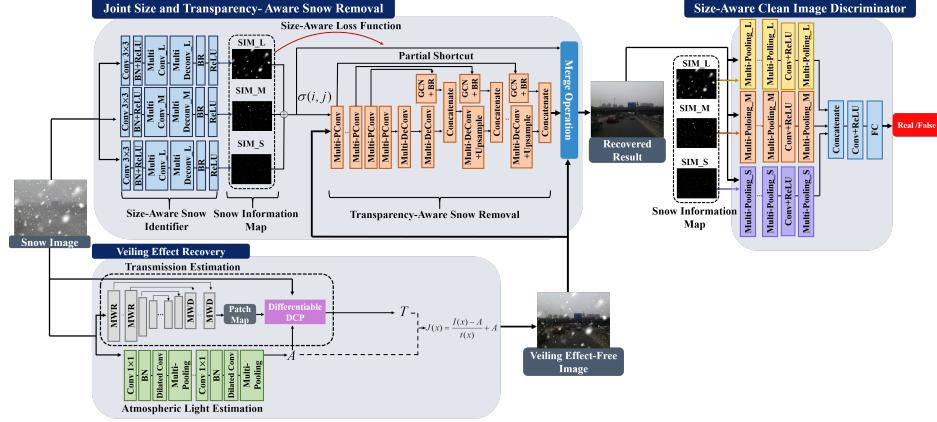


Fig. 2. The overall flow chart of the proposed de-snowing architecture. The proposed method consists three parts: snow removal, veiling effect removal, and clean image discriminator.

3 Proposed Method

The flowchart of the proposed snow removal algorithm is depicted in Fig. 2. Based on the observation of snow images in real world (see Fig. 1) and the limitations mentioned in section 1, several mechanisms are designed. First, the snow formation model is reformulated. Second, the **Joint Size and Transparency-Aware Snow Removal (JSTASR)** module is designed to remove both transparent and non-transparent snows. The proposed JSTASR consists of two parts: (i) the size-aware snow identifier and (ii) transparency-aware snow removal. The size-aware snow identifier will generate the snow information map according to different snow scales. Based on this information, the transparency-aware snow removal module can deal with the snow in different scales and transparencies separately. Moreover, to address non-transparent snow, the modified partial convolution is proposed. Third, the differentiable dark channel prior layer is proposed and embedded in the network to remove the veiling effect.

3.1 Snow Model Formulation

In this section, the snow model is reformulated due to the distribution and the veiling effect of snow images. First, unlike the global-distributed atmospheric phenomenon such as haze, mist, or rain, the snow may distribute locally. However, there is no region information in (1) and it may cause error in the snow removal process for snow-free pixels. By using existing snow removal algorithms, the recovered results tend to be over-smoothed and the detail may vanish in those pixels. Second, in snow images, the veiling effect, which is similar to haze and mist, usually occurs. There are two reasons that can explain this phenomenon. First, when snow starts to fall, the temperature on the ground decreases and the

vapor in the air may condense to haze or mist. Second, tiny snow particles may accumulate and overlap especially in the distant scene. This may also lead to the veiling effect in snow images. However, for the conventional snow formation model, the veiling effect is not considered and it may limit the performance of the recovered process. Therefore, inspired by the Koschmieder model, which is generally performed on the atmospheric particle turbulence problem, we proposed a new snow formation model as follows to take the veiling effect into consideration

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

where $K(x) = J(x)(1 - Z(x)R(x)) + C(x)Z(x)R(x)$, I denotes the image acquired by the camera, K is the veiling effect-free but snowy image. $T(x) = e^{-\beta d(x)}$ is media transmission where $d(x)$ is the distance from the camera to the object and β is the scattering coefficient. A is the atmospheric light of the veiling effect. J is the clean image. C and Z are the same definition as (1). R is a binary mask to denote the snow location information. By this reformulation, the veiling effect problem in snow scenario can be well addressed. Moreover, different from previous snow removal algorithms, with the introduction of R , the local reconstruction is performed on the snow pixels. Therefore, with the veiling effect recovery and the local reconstruction, the recovered image can achieve sharper result containing more fruitful edge comparing with other methods.

3.2 Joint Size and Transparency-Aware Snow Removal

In this section, we illustrate the proposed JSTASR module in detail. It can address the aforementioned snow removal problem via the fully end-to-end learning system. As discussed before, the conventional snow removal algorithms may tend to over or under reconstruction because they do not consider the variety of snow particles. Compared with haze and rain, snow is more complicated because snow particles have different sizes and transparencies and are distributed locally. Thus, in the proposed JSTASR module, the information of (i) size, (ii) transparency, and (iii) localization of snow particles are incorporated to the network design. The flowchart of JSTASR module is shown in upper side of Fig. 2.

Size-Aware Snow Identifier: In the proposed architecture, the size-aware snow identifier is designed. That is, given a snow image, one can acquire three snow information maps in different scales (i.e., large, medium, and small). These maps are gray-scale image consisting of the location and the intensity information. For the design of the proposed snow identifier, three different networks are designed to predict large, medium, and small snow particles and the multi-scale convolution and deconvolution layers are adopted. In these layers, several convolution and deconvolution filters with different scales are connected in parallel. (i.e., $L = L_1 \parallel L_2 \parallel L_3$ for large snow particles, $M = M_1 \parallel M_2 \parallel M_3$ for medium snow particles, $S = S_1 \parallel S_2 \parallel S_3$ for small snow particles, where \parallel denotes the parallel connection and L_n , M_n , and S_n present the convolution kernels with different kernel sizes). In general, the kernel size in L is largest and that in S is smallest. With this architecture, the size and the location information can be

acquired. Moreover, the shape information of snow particles can also be estimated by this operation. Then, the accurate snow information map (SIM) can be computed and it will be used in the snow removal process in next subsection.

Transparency-Aware Snow Removal: For snow particles, unlike haze and rain, its transparency may be more divergent. From Fig. 1, one can observe that, in a snow image, some snow particles are non-transparent while some are transparent. Recovering transparent particles may be easier because it has been developed for a while in [33,30]. However, recovering non-transparent particles may be challenging because the background information is totally occluded by invisible snow, which means that these regions can be only estimated by the neighboring area. Thus, inspired by other computer vision applications [16,18,31], the technique of image inpainting is adopted to mitigate this problem. Image inpainting aims to fill the holes (broken pixels) in the image, which is similar to the problem in non-transparent snow removal. In this paper, the modified partial convolution which can perform the transparency-aware snow removal is proposed. This method is inspired by the inpainting strategy in [16] which applies the partial convolution to inpaint irregular area. The proposed snow removal operation can be expressed as

$$y' = \begin{cases} H^T (S \odot M) \frac{\text{sum}(1)}{\text{sum}(U)} + b, & \text{if } \text{sum}(U) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where y' denotes the output of inpainting, H presents the convolution filter weight; b is the corresponding bias, S is the input feature map of the snow image for the current convolutional kernel, U is the corresponding binary mask which records the useful (transparent or clean) information for the inpainting process, \odot means element-wise multiplication, and $\text{sum}(1)$ presents summation of the matrix which has the same dimension as U but all values are one. $\text{sum}(1)/\text{sum}(U)$ is a scaling factor to adjust the varying amount of unmasked inputs and prevent the feature map from vanishing. After the operation in (3) is performed for one round, the elements of the original snow mask in U will be updated to U' by (4)

$$U'(m,n) = \begin{cases} 1, & \text{if } \sum_{(p,q) \in \Omega_{m,n}} (U(p,q)) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $\Omega_{m,n}$ is a patch centered at (m, n) . By (4), the pixel value at (m, n) will be updated as a useful information if there is at least one useful pixel (transparent or clear pixels) within the patch $\Omega_{m,n}$. Finally, after the inpainting process repeated for sufficient rounds, the whole snow mask U will be set to 1, which means that all non-transparent snow particles are filled and transparent pixels are also recovered to clear pixels. To better remove the snow with different transparencies, different from the original partial convolution, the transparency-aware mechanism is proposed to initialize the snow mask U in (3) according to the snow information maps predicted by the size-aware snow identifier. This operation is important because it can determine whether this information is useful

in the inpainting process. The transparency-aware mechanism in this architecture can help us to select proper reference pixels for inpainting. The operation can be expressed as

$$\sigma(i, j) = \begin{cases} 1, & \text{if } N(i, j) \geq \gamma \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $\sigma(i, j)$ denotes the transparency-aware snow region map (i.e., initial U in (3)), $N(i, j)$ presents the summation of snow information maps in three scales generated from the size-aware snow identifier, and γ is the transparent threshold which determines whether the pixel is visible or not. With this operation, snow pixels can be recovered according to their transparencies, that is, non-transparent pixels are recovered only based on the neighboring pixels while the transparent pixels are recovered partially by itself and partially by neighbor pixels.

The proposed transparency-aware snow removal architecture is inspired by U-Net [23]. For the encoder part, different from the traditional partial convolution which applies single convolution kernel size, the multi-scale architecture is adopted. With it, the encoder can extract features in various levels adaptively according to the different sizes of snow. For the decoder part, different from the original partial convolution which applies up-sampling and convolution in the decoder, we adopt both multi-scale deconvolution and up-sampling operations to prevent the recovered feature from blurring. Moreover, the partial shortcut as follows is applied to improve the quality of the reconstructed images :

$$k_i = \rho [(y'_i \odot U_i) \cup k_{i-1}] \quad (6)$$

where k_i denotes the output of the decoder in the layer i , ρ is a combination of the traditional convolution operation, the global convolutional network (GCN) [20] module, and the boundary refinement (BR) [20]. y'_i and U_i indicate the output of the partial convolution and the snow mask in the layer i , which has been presented in (3), and \cup presents the concatenate operation. The proposed idea is that, for the original design, the decoded information usually tends to be blurring. Inspired by the U-Net [23], the shortcut of original information is applied to increase the recovered resolution. However, instead of passing the original information directly, we pass the partial information to the decoder because the background information in non-transparent region is occluded in the original data and this information may damage the recovered process. Therefore, we discard the information on non-transparent pixels and only adopt useful information. Moreover, the adopted GCN and BR modules are helpful for enhancing the boundary and semantic information. With these operations, high-quality and high-resolution de-snowing results can be acquired.

Size-Aware Loss Function: In order to enhance the performance of the generator, we leverage the snow information maps which are predicted by the snow identifier and proposed size-aware loss function in (7).

$$L_{size-aware} = (w_s + 0.5w_m) [\lambda_1 L_{pixel} + \lambda_2 L_{TV}] + (w_l + 0.5w_m) [\lambda_3 L_S + \lambda_4 L_P] \quad (7)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are constants, w_s, w_m and w_l are the composition percentages of snow pixels in the small, medium, and large snow information maps, respectively. L_{pixel} [16], L_{TV} [16], L_S [16] and L_P [11] denote the pixel loss, the total variation loss, the style loss and perceptual loss, respectively. The idea is based on that, for the region recovered from small snow particles, the total variation loss and the pixel loss should have more weight because the results are generally suffered from noise and the under-smoothed problem. On the other hand, the area damaged by the large snow should put more emphasis on the style loss and the perceptual loss because the recovered region may lose the texture and global information [16,11]. Moreover, for the pixels damaged by medium snow, since the recovered results are between large and small snow, the weights should be shared with w_s and w_l . By designing the size-aware loss function, the error can be calculated according to different snow scenarios.

Size-Aware Discriminator: To improve the performance of the transparency-aware snow removal module, the size-aware discriminator which can identify whether the recovered result looks like a real image is proposed. Inspired from [21], which applied the attention map to improve the performance of the discriminator in rain drop removal. Based on the observation of snow above, to further improve the recovered process in the snow scenario, three snow information maps based on different sizes (i.e, L, M, S) are applied in the proposed network. The features are extracted with snow information map in different scales by multi-pooling as follows

$$l = \psi \underset{e \in s}{M_e^k}(x). \quad (8)$$

where x represents the features determined from the previous layer, M_e^k denotes the stride convolution operation with the kernel size e and the dilated level k , ψ denotes the concatenate operation, and s is the scale range for the stride convolution where $s \in \{2, 3, 5\}$. In this work, k is set to 2. By the operation of multi pooling, the extracted features can be well preserved in different sizes and scales. Moreover, the concatenating operation is performed instead of addition because we hope to preserve the properties at each scale. The idea of size-aware discriminator is based on that, compared to rain drop, snow is more complicated. It may be non-transparent and the size variation is large. By extracting the features from different scales and focusing on different transparencies, the performance of the discriminative process will be improved effectively. Then, these features will be concatenated together at the end of network and the fully connected operation is performed to identify the recovered result.

Overall Loss Function: The overall loss function of the proposed JSTASR module can be presented as

$$L_{JSTASR} = L_{size-aware} + \lambda_5 L_{Identifier} + \lambda_6 L_{Adv} \quad (9)$$

$$L_{Adv} = \min_G \max_D E_x[\log D(x)] + E_x[\log(1 - D(G(x')))] \quad (10)$$

where λ_5 and λ_6 are constants. x and x' are snow-free image and snow image. $L_{Identifier}$ presents the MSE between the predicted snow information maps and their corresponding ground truths. L_{JSTASR} and L_{Adv} are the generative loss and the adversarial loss, respectively.

3.3 Veiling Effect Removal

In veiling effect removal, we aim to address the inverse problem in (2), that is, estimating the accurate transmission and atmospheric light values. For the transmission value, we apply the differentiable DCP layer with the patch map [3]. The flowchart of veiling effect removal is shown in the bottom of Fig. 2.

Differentiable Dark Channel Prior: To cope with the veiling effect problem in the snow removal process, in the proposed method, we apply the DCP [10] to estimate the transmission because it has been proved to be an effective way to deal with the veiling effect in other works about atmospheric particle removal [19,34]. However, this method has been indicated several limitations such as the color degradation in the white and bright scene [3,29]. Although in PMS-Net [3], the patch map can address the limitation of the dark channel prior effectively, sometimes the performance is limited in color fidelity and recovered quality. The reason is that the PMS-Net cannot train the whole system including the atmospheric light estimation and refined network because the patch-map based dark channel cannot be embedded into this method. Moreover, due to the complicated scenario in our proposed method, end-to-end training and optimization is necessary to achieve better performance. Therefore, in this paper, we develop the patch-map based differentiable dark channel prior layer to improve the performance of this method and further embed into the proposed snow removal process to achieve fully end-to-end learning. This layer is extended with a learnable patch map. The inputs are the image with veil, the predicted patch map, and the predicted atmospheric light. The output is the estimated transmission map. First, the minimum operation in the local patch is performed in each color channel. The operation can be expressed as follows

$$H(x, y, z, p) = \min_{\alpha, \beta \in [1, p]} [I(x + \alpha, y + \beta, z)] \quad (11)$$

where I is the input image. x and y are the indexes of pixel location. z indicates the different color channel. p is the index of the patch size axis, and $H(x, y, z, p)$ is the result of the minimum operation. Then, $H'(x, y, p)$ is formulated by applying the minimum filter along the color channel axis z on $H(x, y, z, p)$. The patch map is then projected to the patch map box $PMB(x, y, p)$:

$$PMB(x, y, p) = \begin{cases} 1, & \text{if } H'(x, y, p) = p \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where $P(x, y)$ is the predicted patch map. Then, $H'(x, y, p)$ will be multiplied with $PMB(x, y, p)$ and the coarse transmission map will be computed by summing the value along the patch size axis (i.e., p -axis). With this process, the fully differentiable DCP layer can be achieved in the end-to-end veiling effect removal network and the performance of recovered quality and color distortion can be improved effectively.

Learning Atmospheric Light: To estimate the global atmospheric light \mathbf{A} from input images, we adopt the learning architecture to predict the air light intensity because we want to achieve end-to-end learning with the differentiable DCP layer. For this sub-network, we apply the VGG-16 architecture [24] as the

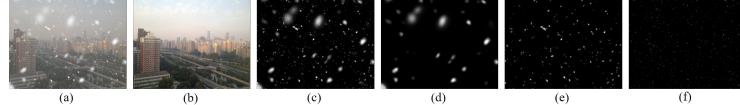


Fig. 3. The example in SRRS: (a) Snow image; (b) ground truth; (c) combined snow mask (d) large snow mask; (e) medium snow mask; (f) small snow mask. Note that the snow masks are gray-scale images which contain the information of snow location, intensity, and size.

backbone, but each convolution operation is replaced by the multi-level pooling operation mentioned in subsection 3.2.

Loss Function: The loss functions of the veiling effect removal module is formulated as

$$L_{Veil} = L_{PatchMap} + \lambda_7 L_A \quad (13)$$

where L_{Veil} is the overall loss of veiling effect removal $L_{PatchMap}$ is the loss of the patch map selection network proposed in [3]. L_A is the atmospheric light loss, which is defined as the MSE between the predicted atmospheric light and the ground truth.

The overall loss function of the proposed whole network can be expressed as

$$L_{Overall} = L_{JSTASR} + L_{Veil} \quad (14)$$

4 Experimental Result

4.1 Dataset Generation

Although there is one large-scale synthetic dataset [17] for image de-snowing, it does not contain the veiling effect. In this work, in order to address various snow scenarios, a novel dataset called the Snow Removal in Realistic Scenario (SRRS)¹ is proposed. It contains 15000 synthesized snow images and 1000 snow images in real scenarios downloaded from the Internet. We synthesize the SRRS by two steps. First, we apply the popular haze benchmark dataset called RESIDE dataset [14] to synthesize the image with veiling effect based on the procedure in their paper. We set $\beta \in [0.4, 1.6]$ and $A \in [0.5, 1]$. Then, for each snow image, various types of snow are synthesized by Photoshop and the corresponding snow information (i.e., transparency, size, and location) is labeled. The example of SRRS is shown in Fig. 3. For the training procedure, we randomly pick 2500 snow images as the training dataset. For the test dataset, 1000 snow images are picked randomly from the proposed SRRS dataset and we call it as **Test A**. Moreover, in order to test the generalization of the proposed network, the test dataset proposed by [17] is applied. We call this dataset as **Test B** which contains 1000 snow images only with snow particles but without veiling effect.

¹ The dataset can be downloaded from our project page

Table 1. Quantitative evaluation for comparison with other state-of-the-art methods on the Test A and Test B. (PSNR/SSIM)

		Eigen [6]	Zheng [35]	DesnowNet [17]	Ours
Test A	w/o MSCNN [22]	14.51/0.58	15.48/0.729	16.32/0.80	25.62/0.89
	w MSCNN [22]	17.36/0.57	18.02/0.76	18.42/0.82	-
Test B	-	18.57/0.43	23.72/0.83	25.58/0.85	25.69/0.86

Table 2. The evaluation of sharpness in terms of JNBM [7] and e [8] on 1000 real-world images. Note: Larger values of JNBM and e means sharper and more newly visible edges on recovered image

	Zheng [35]	Eigen [6]	DesnowNet [17]	Ours
JNBM/ e	2.19/-0.08	3.04/-0.31	2.95/-0.09	3.92/0.25

Table 3. Run time comparison with state-of-the-art snow removal methods.

	Zheng (C) [35]	Eigen (G) [6]	DesnowNet (G) [17]	Ours (G)
Time (s)	1.4	2.81	1.38	0.36

Table 4. Comparing the performances of object detection when using the proposed algorithm and other state-of-the-art methods for de-snowing in prior.

	Baseline	Zheng+MSCNN	Eigen+MSCNN	DesnowNet+MSCNN	Ours
Accuracy (IoU)	0.23	0.43	0.33	0.50	0.72

4.2 Training Detail

The proposed de-snowed network mainly consists of two sub-networks: the JS-TASR module and veiling effect removal. First, we train the veiling effect removal network with 2500 hazy images based on the RESIDE [14] dataset. Second, for JSTASR, initially, we train the size-aware snow identifier to predict the snow information. Then, the identifier is fixed to train the transparency-aware snow removal module. These two processes are pre-trained with the fixed veiling effect removal network. After the pre-trained process, two sub-networks are trained together in the fine-tuned state. All these modules and the fine-tuned process are trained based on the training set of *SRRS* dataset. The learning rate is set to e^{-4} and the Adam [12] optimizer is adopted in both JSTASR and veiling effect removal. In this network, the pre-trained process is adopted. For the parameter setting in the proposed network, we set $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$, $\lambda_6 = 0.5$, and $\lambda_7 = 10^{-4}$. We set the threshold value $\gamma = 0.1$. For each epoch, we cut 15% of the training data as the validation set. Finally, two pre-trained sub-networks will be combined and trained together to fine-tune the entire network for several thousand iterations. The proposed method is implemented on a workstation with a 3.5 GHz CPU, 64G RAM, and Nvidia GTX 1080 Ti GPU. The number of the parameter in the proposed network is 6.5×10^7 . Moreover, the whole training process of this network takes five days.

Table 5. Quantitative evaluation for the ablation study in the proposed modules of snow removal on Test A (PSNR/SSIM)

Module	J	J+P	J+P+M	J+P+M+T	All w/o EVER	All w EVER
PSNR/SSIM	24.01/0.81	24.86/0.87	23.02/0.88	24.27/0.89	20.32/0.77	25.85/0.89

Table 6. Quantitative evaluation for the ablation study of the proposed differential dark channel prior layer on the indoor dataset in SOTS [14]. The lower value of CIEDE 2000 means less color distortion.

	DCP [10]	PMS-Net [3]	Ours
PSNR/SSIM/CIEDE 2000	18.2/0.83/9.42	21.14/0.88/5.88	24.68/0.90/4.25

4.3 Comparison with State-of-the-art Methods

Analysis in Synthetic Snow Dataset. In Table 1, the proposed de-snowed algorithm is evaluated on the synthetic snow dataset. We apply two metrics: the peak-to-peak signal to noise ratio (PSNR) and the structural similarity (SSIM). We apply state-of-the-art snow removal algorithms(i.e., Zheng [35], and DesnowNet [17]) and atmospheric particle removal methods (i.e., Eigen [6]) for comparison. To compare with these methods fairly, we apply the conventional haze removal method (i.e., MSCNN [22]) as the veiling effect removal strategy because the images in Test A contain the veiling effect. The quantitative comparison is presented in Table 1. We can find that the proposed method outperforms state-of-the-art methods on both Test A and B in all metrics no matter the veiling removal strategy is applied or not. Note that, the proposed method is trained on our proposed dataset but still can achieve better performance on that proposed by the DesnowNet [17]. These results can prove that our proposed method for snow removal is guaranteed and effective.

Analysis in Real-world Snow Dataset. The qualitative comparison of the proposed method with other state-of-the-art de-snowing algorithms is performed. In Fig. 4, some examples of our real-world dataset are shown. One can notice that, the results recovered by Eigen [6] and Zheng [35] may still contain snow particles and those recovered by the later one may have severe blurring problem. Although the results recovered by DesnowNet [17] can achieve better performance comparing with Eigen [6] and Zheng [35], the large and non-transparent snow particles cannot be removed clearly (see the 2nd, 3rd and 4th rows). Moreover, all other methods cannot address the veiling effect problem (see the 1st, 3rd and 6th and the Fig. 1). By contrast, the proposed algorithm can well remove snow particle with different sizes and transparencies, and veiling effect because we propose the size-aware snow identifier, transparency-aware snow removal and differentiable dark channel prior. Furthermore, with the snow information maps predicted by the size-aware snow identifier, the proposed method can prevent the recovered images from blurring effectively because the snow location is considered during the snow removal procedure. To evaluate the effectiveness of this mechanism, we apply the just noticeable blur metric (JNBM) [7] and the newly visible edge ratio (e) in VLD [8] to evaluate the performance of detail and sharp-

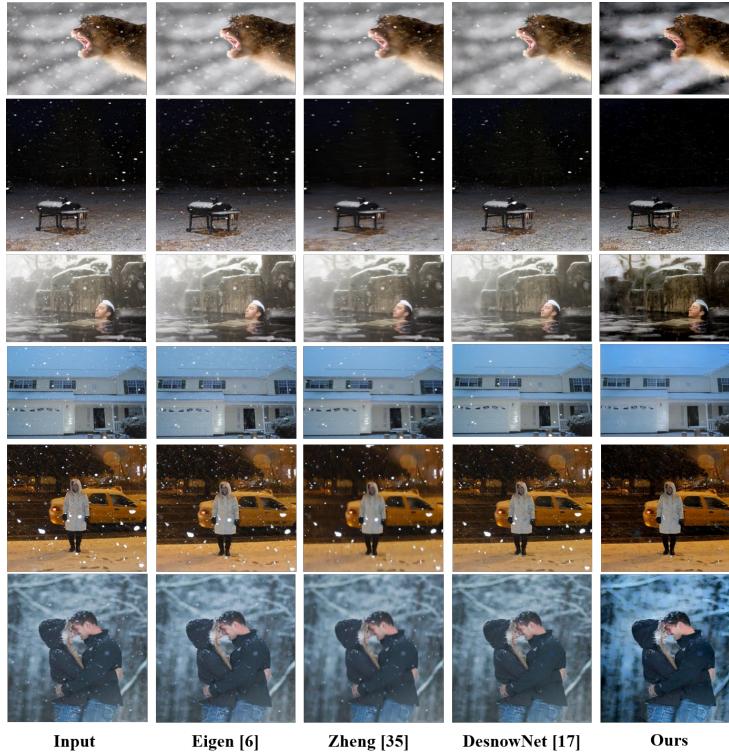


Fig. 4. A qualitative comparison of the proposed method with state-of-the-art algorithms on real-world images.(Zoom-in to view in detail.)

ness preservation. The results shown in Table 2 demonstrate that the proposed method achieve the best performance on retaining both the sharpness and edge comparing with other methods. It can prove the effectiveness of the proposed method in the real-world images.

Run Time Comparison. We further compare the runtime in Table 3. The results show that the proposed snow removal method can achieve much better performance in computational time². It proves that the proposed method has superior performance comparing to state-of-the-art methods in terms of both the run time and the recovered quality.

Improvement in High-Level Vision Application of Object Detection. In Table 4, we verify that the proposed snow removal algorithm can benefit to the high-level vision tasks such as object detection in the snow scenario. We apply the RTTS dataset in RESIDE [14], which contains the hazy images with annotation and conduct the snow synthesis process in subsection 4.1 to generate snow images with annotation. We compare the proposed algorithm with

² The input size of DesnowNet in this experiment is 480×480 .

three desnowing methods. The object detection algorithm is fixed to the Mask R-CNN [9]. From Table 4, one can see that, with the utility of the proposed de-snowing algorithm, the accuracy of object detection can be improved by 44% and 210% comparing to the DesnowNet and baseline. It proves that the proposed approach is helpful for achieving better performance for the object detection.

4.4 Ablation Study

To verify the effectiveness of the proposed modules, the ablation study is performed. The evaluation consists of two part: the proposed JSTASR module and the differentiable DCP. For the former one, six component combinations are constructed, that is, 1) JSTASR with traditional partial convolution (**J**); 2) the **J** module with partial shortcut (**J+P**); 3) the **J+PS** module with multi-convolution (**J+P+M**); 4) the **J+P+M** module with the transparency-aware mechanism (**J+P+M+T**); 5) the **J+P+M+T** module with the size-aware mechanism (i.e., the size-aware loss and the size-aware discriminator) but without end-to-end veiling effect removal (**All w/o EVER**); 6) the JSTASR with end-to-end veiling effect removal (**(All) w EVER**). The result is shown in Table 5. One can see that all the proposed modules are effective for the snow removal process. Furthermore, for the veiling effect removal, the results illustrate that the image quality achieves great improvement with the proposed differentiable DCP. To prove the effectiveness of the proposed differentiable dark channel prior, we adopt the ablation study on the famous haze image benchmark called SOTS dataset without the snow removal process in Table 6. In this experiment, we only apply indoor dataset to evaluate the improvement. The results show that applying the differentiable DCP layer can achieve higher performance on dehazing comparing to the PMS-Net [3] and DCP [10].

5 Conclusion

In this paper, a novel de-snowed algorithm based on joint size and transparency aware filters and veiling effect removal is proposed. The differentiable DCP is proposed to remove the veiling effect. In the snow removal procedure, first, the size-aware snow identifier is proposed to identify snow particles according to their sizes. In each snow information map, the intensity, location, and size information are recorded to perform adaptive snow removal. Then, the transparency-aware snow removal process based on the modified partial convolution is developed to address various snow types. Also, a binary mask is applied to select the useful information for snow pixel recovery. Moreover, to further improve the performance of the modified partial convolution, the partial shortcut and the multi-scale encoder and decoder are proposed. Last, to better optimize the snow recovered results in different sizes of snow particles, size-aware loss functions and the snow-free discriminator are designed. Experimental results showed that the proposed method can achieve better performance even in the complicated snow scenarios compared to other methods.

References

1. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1674–1682 (2016)
2. 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* **25**(11), 5187–5198 (2016)
3. Chen, W.T., Ding, J.J., Kuo, S.Y.: PMS-Net: Robust Haze Removal Based on Patch Map for Single Images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11681–11689 (2019)
4. Chen, W.T., Fang, H.Y., Ding, J.J., Kuo, S.Y.: Pmhld: Patch map based hybrid learning dehazenet for single image haze removal. *IEEE Transactions on Image Processing* (2020)
5. Chen, W.T., Yuan, S.Y., Tsai, G.C., Wang, H.C., Kuo, S.Y.: Color channel-based smoke removal algorithm using machine learning for static images. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 2855–2859. IEEE (2018)
6. Eigen, D., Krishnan, D., Fergus, R.: Restoring an image taken through a window covered with dirt or rain. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 633–640 (2013)
7. Ferzli, R., Karam, L.J.: A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). *IEEE transactions on image processing* **18**(4), 717–728 (2009)
8. Hautière, N., Tarel, J.P., Aubert, D., Dumont, E.: Blind contrast enhancement assessment by gradient ratioing at visible edges. *Image Analysis & Stereology* **27**(2), 87–95 (2008)
9. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence* **33**(12), 2341–2353 (2010)
11. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711 (2016)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. 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. pp. 4770–4778 (2017)
14. 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* **28**(1), 492–505 (2018)
15. Li, Z., Zhang, J., Fang, Z., Huang, B., Jiang, X., Gao, Y., Hwang, J.N.: Single Image Snow Removal via Composition Generative Adversarial Networks. *IEEE Access* **7**, 25016–25025 (2019)
16. Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 85–100 (2018)
17. 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* **27**(6), 3064–3073 (2018)

18. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2536–2544 (2016)
19. Pei, S.C., Tsai, Y.T., Lee, C.Y.: Removing rain and snow in a single image using saturation and visibility features. In: 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). pp. 1–6 (2014)
20. Peng, C., Zhang, X., Yu, G., Luo, G., Sun, J.: Large Kernel Matters-Improve Semantic Segmentation by Global Convolutional Network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4353–4361 (2017)
21. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2482–2491 (2018)
22. 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. pp. 154–169 (2016)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241 (2015)
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
25. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 2201–2208 (2009)
26. Voronin, V., Semenishchev, E., Zhdanova, M., Sizyakin, R., Zelenskii, A.: Rain and snow removal using multi-guided filter and anisotropic gradient in the quaternion framework. In: Artificial Intelligence and Machine Learning in Defense Applications. vol. 11169, p. 111690 (2019)
27. Wang, Y., Liu, S., Chen, C., Zeng, B.: A hierarchical approach for rain or snow removing in a single color image. *IEEE Transactions on Image Processing* **26**(8), 3936–3950 (2017)
28. Xu, J., Zhao, W., Liu, P., Tang, X.: An improved guidance image based method to remove rain and snow in a single image. *Computer and Information Science* **5**(3), 49 (2012)
29. Yang, D., Sun, J.: Proximal dehaze-net: a prior learning-based deep network for single image dehazing. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 702–717 (2018)
30. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1357–1366 (2017)
31. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5505–5514 (2018)
32. Yu, S., Zhao, Y., Mou, Y., Wu, J., Han, L., Yang, X., Zhao, B.: Content-adaptive rain and snow removal algorithms for single image. In: International Symposium on Neural Networks. pp. 439–448 (2014)
33. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3194–3203 (2018)
34. Zhang, Z., Liu, W., Ma, H., Liu, X.: Going Clear from Misty Rain in Dark Channel Guided Network

35. Zheng, X., Liao, Y., Guo, W., Fu, X., Ding, X.: Single-image-based rain and snow removal using multi-guided filter. In: International Conference on Neural Information Processing. pp. 258–265 (2013)
36. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing* **24**(11), 3522–3533 (2015)