This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

ALL Snow Removed: Single Image Desnowing Algorithm Using Hierarchical Dual-tree Complex Wavelet Representation and Contradict Channel Loss

Wei-Ting Chen[1,3*], Hao-Yu Fang[2*],
Cheng-Lin Hsieh[1], Cheng-Che Tsai[1], I-Hsiang Chen[2], Jian-Jiun Ding[2], and Sy-Yen Kuo[2]

[1]Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
[2]Department of Electrical Engineering, National Taiwan University, Taiwan
[3]ASUS Intelligent Cloud Services, Taiwan

{jimmy3505090, danielfang60609}@gmail.com
{r08943150, r08943148, r09921058, jjding, sykuo}@ntu.edu.tw

Abstract

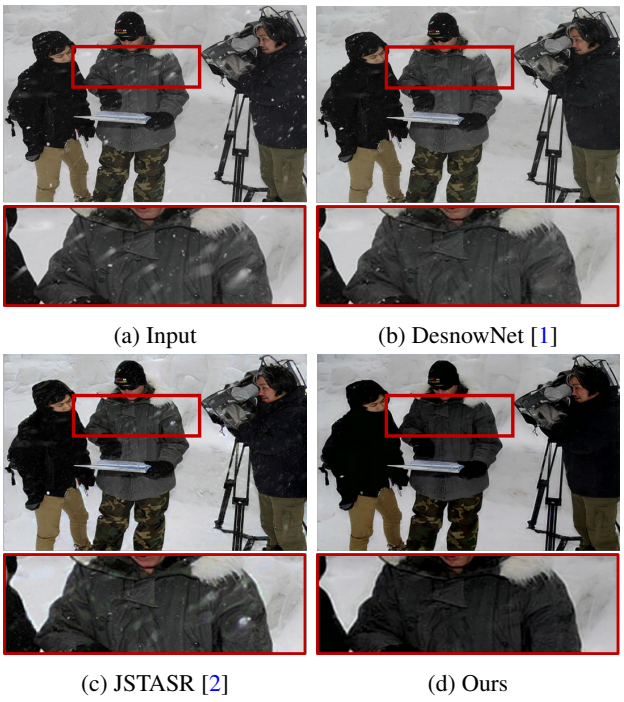
Snow is a highly complicated atmospheric phenomenon that usually contains snowflake, snow streak, and veiling effect (similar to the haze or the mist). In this literature, we propose a single image desnowing algorithm to address the diversity of snow particles in shape and size. First, to better represent the complex snow shape, we apply the dual-tree wavelet transform and propose a complex wavelet loss in the network. Second, we propose a hierarchical decomposition paradigm in our network for better understanding the different sizes of snow particles. Last, we propose a novel feature called the contradict channel (CC) for the snow scenes. We find that the regions containing the snow particles tend to have higher intensity in the CC than that in the snow-free regions. We leverage this discriminative feature to construct the contradict channel loss for improving the performance of snow removal. Moreover, due to the limitation of existing snow datasets, to simulate the snow scenarios comprehensively, we propose a large-scale dataset called Comprehensive Snow Dataset (CSD). Experimental results show that the proposed method can favorably outperform existing methods in three synthetic datasets and real-world datasets. The code and dataset are released in https://github.com/weitingchen83/ICCV2021-Single-Image-Desnowing-HDCWNet.

1. Introduction

Snow is an atmospheric phenomenon that usually affects the performance of computer vision such as object detection and semantic segmentation [3, 2]. According to the previous work [2], the formation of snow can be modeled as:

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

where $\mathbf{K}(x) = \mathbf{J}(x)(1 - \mathbf{Z}(x)\mathbf{R}(x)) + \mathbf{C}(x)\mathbf{Z}(x)\mathbf{R}(x)$, $\mathbf{I}$ denotes the image deteriorated by snow, $\mathbf{K}$ is the snowy image



(a) Input (b) DesnowNet [1]
(c) JSTASR [2] (d) Ours

Figure 1: **Comparison between state-of-the-art snow removal algorithms.** (a) Input; (b)-(d) Recovered results by the DesnowNet [1], the JSTASR [2], and our proposed method. One can see that the proposed method can achieve better performance on snow removal.

*Indicates equal contribution.

without the veiling effect, **A** is the atmospheric light, and **J** is the scene radiance. $\mathbf{T}(x) = e^{-\beta d(x)}$ is the media transmission where $d(x)$ is the depth and β is the scattering coefficient. **C** and **Z** are the chromatic aberration map for snow images and the snow mask, respectively. **R** is a binary mask which presents the snow location information. Snow is a very complex atmospheric phenomenon because it contains snowflakes, snow streaks, and the veiling effect. Existing snow removal strategies can be categorized into two classes: hand crafted feature-based and learning-based strategies. For hand crafted feature-based methods [4, 5, 6, 7, 8], Pei *et al*. [6] utilized the features based on the color information such as saturation and visibility for desnowing. Xu *et al*. [4] developed a guidance image based snow removal pipeline. Zheng *et al*. [5] extracted features by multi-guided filters to separate the snow part from the background. These methods are based on the features observed by a human. However, they may not hold for diverse real-world scenarios. With the rising of deep learning, several CNN-based approaches have been proposed [3, 2, 1]. Liu *et al.* [1] developed the first learning-based desnowing method called the Desnownet. They proposed the first snow dataset called Snow-100K. Chen *et al.* [2] proposed an algorithm by image inpainting and veiling effect recovery to perform snow removal and developed a dataset called SRRS to make the snow removal process more robust. Although these methods achieve good performance, there are still some issues to be noted:

(i) Robustness to real-world scenarios: The real-world snow scenes are highly complicated. For existing snow removal strategies, the diversity of snow shape and size is often ignored. First, existing methods lack an effective mechanism to capture the shape of snow. Moreover, conventional algorithms only consider snowflakes but neglect snow streaks. Second, because large snow particles are not easy to be modeled for training, they are difficult to be removed under real-world scenarios. Due to these issues, as shown in Figure 1, existing methods may have limited performance in real-world scenarios.

(ii) Existing snow datasets cannot reflect the complicated constitution of real-world snow scenarios: In the Snow-100K dataset [1], the snowflakes are in different size are synthesized, but snow streaks and the veiling effect are not considered. In the SRRS dataset [2], although the veiling effect is considered, it does not include the snow streaks. The snow streak is similar to the rain streak, but it has stronger intensity and may be blurrier than other snow particles. The lack of the dataset with comprehensive snow features may degrade the performance of the network when handling real-world snow scenes.

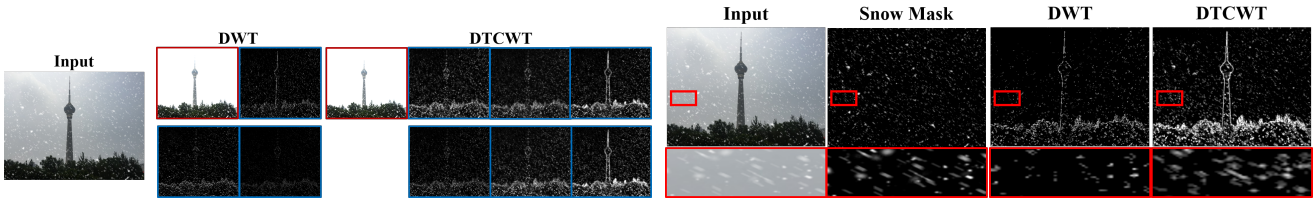
Therefore, in this paper, we propose a novel snow removal architecture which can handle the complicated snow scenes by embedding the hierarchical dual-tree complex wavelet transform [9] to the network. Based on this operation, the snow image is decomposed and recovered in each scale by the recurrent decomposition. Based on this operation, all subbands can be reconstructed at the small scales. Moreover, snow particles with large size which do not appear in the training data can be represented at small scale. Consequently, the snow particles in the decomposed subbands can be easier to be removed because the training set can cover and simulate their distribution. Besides, the snow streaks and snowflakes can be well represented with the multi-direction property of DTCWT. Thus, with this architecture, the size-free and more comprehensive snow removal can be achieved. Then, we design two sub networks for complex wavelet recovery, that is, the high-frequency reconstruction network and the low-frequency reconstruction network. In addition, we investigate snow and snow-free images and find a new feature called the contradict channel (CC). Based on the CC, we construct the contradict channel loss to improve the performance of snow removal. Moreover, we proposed a novel large-scaled snow dataset called **C**omprehensive **S**now **D**ataset (`CSD`) which can solve the limitations in existing datasets. The main contributions in this paper are summarized as follows:

- A novel hierarchical DTCWT desnowing network is proposed. Benefiting from the multi-direction and multi-scale properties of the hierarchical DTCWT, the proposed network can achieve better adaptability on complicated snow scenarios. Experiments on all existing synthetic datasets and real-world dataset show that the proposed method achieves the state-of-the-art performance on snow removal.
- We propose a new feature called the contradict channel (CC) based on investigating the difference of snow and clean images. With the discriminative feature, the contradict channel loss is proposed to benefit the optimization of the network. As far as we know, this is the first work proposing the CC.
- In order to address the limitation in existing snow datasets, we proposed a large-scale snow image dataset called `CSD` to provide the training data.

2. Related Works

2.1. Single Image Snow Removal

For snow removal in the single image [4, 5, 6, 1, 7, 8, 3, 10, 2], Zheng *et al.* [5] investigated the difference between snow streaks and clear background edges, and applied the multi-guided filter to remove snowflakes. Wang *et al*. [7] proposed a three-layer hierarchical scheme which combines image decomposition and dictionary learning. Voronin *et al*. [11] developed the anisotropic gradient in Hamiltonian quaternions to remove rain and snow. Li *et al*. [12] applied the generative adversarial network (GAN)



(a) Decomposed results by DWT and DTCWT

(b) Comparison of snow particles extraction

Figure 2: **Comparison of decomposed results and snow extraction by DWT and DTCWT.** From (b), one can see that with the better interpretation of multi-directions, the DTCWT can retrieve more information about snow particles, especially for snow streaks.

for snowflake removal. Liu *et al*. [1] proposed a two-staged snow removal architecture called the DesnowNet based on Inception-v4 [13]. Moreover, they proposed the first snow dataset called Snow100K. However, it cannot remove snow adequately because the veiling effect and non-transparent snow particles are not considered. Chen *et al.* [2] proposed a network called JSTASR to remove snow particles and the veiling effect by combining the modified partial convolution and the differentiable dark channel prior layer. They also proposed a snow dataset called SRRS which considers the veiling effect. Although this algorithm can address the veiling effect and non-transparency snow particles, sometimes this method may fail due to: i) lack of considering snow streaks; ii) inaccurate detection of snow particles, which may cause over/under desnowing; iii) over deveiling problem which causes the color distortion.

2.2. Computer Vision Applications Based on Dual-tree Complex Wavelet Transform

The Dual-tree Complex Wavelet Transform (DTCWT) [9] has been adopted to solve image processing and restoration problems [14, 15, 16, 17] for a long while. Singh *et al*. [18] applied the DTCWT with the ScatterNet to image classification. Li *et al.* [19] and Lu *et al.* [20] used the DTCWT for image segmentation. Chen *et al.* [21] leveraged the phase-shift property of the DTWCT to perform 3D registration. Sun *et al.* [22] and Jung *et al.* [23] utilized it together with the adaptive histogram equalization for low-light enhancement.

3. Methodology

3.1. Dual-tree Complex Wavelet Transformation

The DTCWT [9] is an improved form of the discrete wavelet transformation (DWT) [24]. Although the DWT has been adopted in several image and signal processing tasks [25, 26] because it can represent the signal in a more sparser way, it may have some limitations: i) poor tolerance of noise; ii) poor directional selectivity which makes the processing of geometric features such as edges, ridges, and diagonal features more challenging. The DTCWT can overcome these problems by introducing complex wavelets and the tree structure. In the DTCWT, a signal is decomposed by analytic filters based on a scaling function, complex shifting, and a dilated mother wavelet:

$$\psi(t) = \psi_h(t) + j\psi_g(t), \tag{2}$$

where $j = \sqrt{-1}$, $\psi_h(t)$ and $\psi_g(t)$ are wavelet basis. They are the real part and imaginary part of the mother wavelet, respectively. For the 2-D DTCWT, as shown in Figure 2a, six high-pass subbands (blue bounding box) and a low-pass subband (red bounding box) are generated. These high-pass subbands contain more detailed and texture information in different directions (i.e., $\pm 15°, \pm 45°, \pm 75°$) and diagonal shapes. This property is beneficial for extracting the snow information because snow particles generally consist of diagonal snowflakes and bevelled snow streaks. As shown in Figure 2b, one can see that the high-frequency subbands in DTCWT can achieve better ability on representing the snow information, especially snow streaks. Therefore, in this paper, we embed the DTCWT to the network for better retrieving snow information.

3.2. Hierarchical DTCWT-based Desnowing

In this section, we illustrate the idea about handling the problem of the snow particles in various sizes and shape. The overall network is presented in Figure 3. Initially, the input snow image is decomposed to high-frequency components and a low-frequency component (i.e., $\{HF_1^1, ..., HF_1^k\}$ and LF_1 where k is the number of the subbands) via the DTCWT in the first level. Then, we decompose the low-frequency component to the next level, that is, the LF_1 is decomposed to $\{HF_2^1, ..., HF_2^k\}$ and LF_2, respectively. Following this operation, we decompose the low-frequency component in each level recurrently to the i^{th} level. We apply the high-frequency reconstruction network in each scale to remove the residual snow (small and medium size) and recover the detailed information while the low-frequency reconstruction network is applied to conduct structural information recovery and the large snow particle

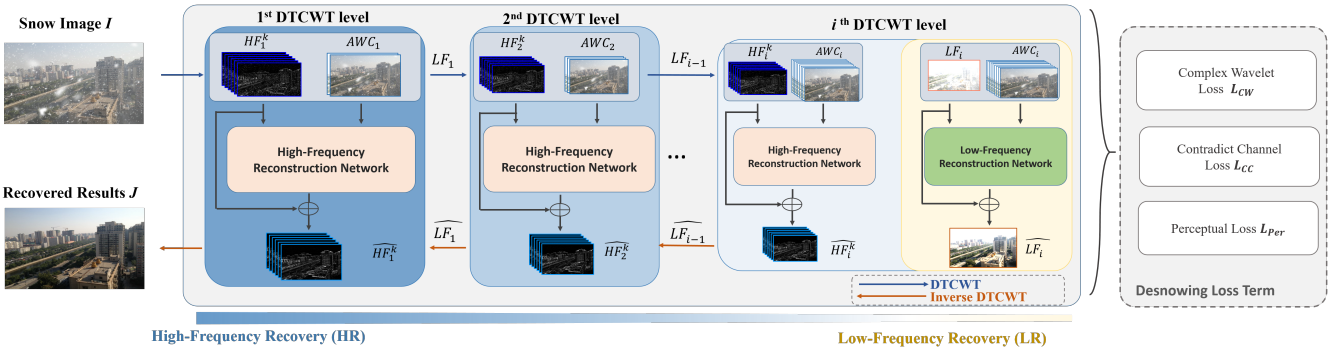


Figure 3: **The overview of the proposed architecture.** The proposed network is based on hierarchical DTCWT decomposition. HF_i^k and $\widehat{HF}_i^k$ denote the input and predicted high-frequency components with k subbands at the i^{th} level.

removal. Then, the reconstructed low-frequency subband ($\widehat{LF}$) and high-frequency subbands ($\widehat{HF}$) are combined by the inverse DTCWT recurrently. By this decomposition and reconstruction paradigm, each subband may contain shrunk spatial resolution, which means that snow images are reconstructed at small scales. Thus, removing snow streaks and snowflakes with larger size may become easier. Moreover, with the better geometrical representation in multi-directions, the snow information can be retrieved in the high-frequency subband in each scale appropriately. Therefore, based on the combination of these two mechanisms, the snow particles with larger size and various shape can be addressed effectively.

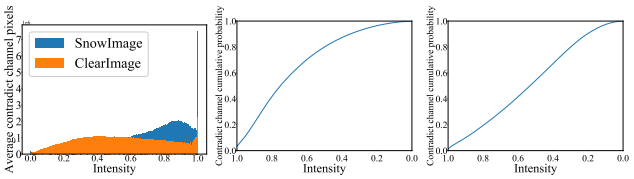
Network Architecture. The hierarchical dual-tree complex wavelet representation network is divided into two parts: i) high-frequency reconstruction (HR) and ii) low-frequency reconstruction (LR). The two sub-networks leverage the Res2Net [27] as the backbone. We further introduce the multi-deconvolution [28, 29], the global convolution [30], and the boundary refinement [30] into the network to improve the performance. The detailed architecture is presented in the Supplementary Material. To improve the performance of the sub-networks, we propose the aggregate wavelet component (AWC) to provide the multi-scale spatial information for two sub-networks.

$$AWC_i = \begin{cases} \mu(\mathbf{I}), & i = 1 \\ [\mu(LF_1), ..., \mu(LF_{i-1}), \mu(\mathbf{I})], & \text{otherwise} \end{cases}, \quad (3)$$

where AWC_i and LF_i denote the AWC and the low-frequency component at the i^{th} level, respectively. $[,]$ presents the concatenate operation, and μ is the multi-pooling architecture [2] as:

$$\mu(q) = \Big\|_{\rho \in \phi} \kappa_\rho^\tau(q), \quad (4)$$

where $\kappa_\rho^\tau()$ denotes the stride convolution operation with the kernel size ρ and the dilated level τ, and $\phi \in \{2, 3, 5\}$ is the scale range for the stride convolution. In this work, the τ is set as 2. The idea of adopting the AWC is as follows. Compared to down-scale the input directly, the AWC contains more semantic and contextual information from the low-frequency components in different scales which can benefit the reconstruction process. Furthermore, to prevent the information loss in down-sampling, multi-pooling is applied. Note that, the architectures of both HR and LR sub-networks are similar. However, more parallel kernels (i.e., kernel sizes with 2, 3, 5, 7, and 9) in the Res2Net and the multi-pooling are used in LR, and its filter depth is wider than that of HR because low-frequency component recovery involves more complex semantic information.



(a) (b) (c)

Figure 4: **Statistical investigation of the contradict channels.** (a) Histogram of the pixel intensities on snow and clean images in 789 images; (b) & (c) Cumulative distributions of contradict channel in snow images and clean images, respectively. We adopt patch size of 35×35. The scale in the x-axis is reversed for the better understanding.

3.3. Contradict Channel Loss

In this paper, we proposed a novel operation which can identify the difference between snow and snow-free images. We call it the contradict channel (CC), which is:

$$\mathbf{I}_{Contradict}(x) = \max_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} \mathbf{I}^c(x) \right), \quad (5)$$

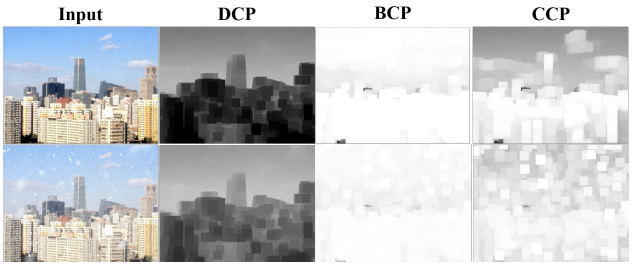


Figure 5: **The comparison of the dark channel, the bright channel, and the contradict channel operation with** 35×35 **patch size on snow-free (upper row) and snow (lower row) images.**

where $\mathbf{I}^c(x)$ is the intensity of the image in the color channel c and $\Omega(x)$ is a local patch with a fixed size centered at x. Then, we apply the CC operation on snow images and clean images, and investigate the histograms of them. The results are presented in Figure 4. From this figure, we can find an interesting phenomenon. Compared to clean images, the CC of snow images may have much more pixels close to one. The regions with snow particles have higher intensity compared with snow-free regions. This phenomenon is due to that: 1) snowflakes, 2) snow streaks, and 3) the veiling effect. These materials will result in high intensity in the contradict channel. On the other hand, compared to snow images, clean images are usually full of objects with vivid color, such as blue water, green grass/trees, and red or yellow flowers, which may make the contradict channel have lower intensity. We define this property as contradict channel prior (CCP).

Compared with existing image priors such as the dark channel prior (DCP) [31] and the bright channel prior (BCP) [32], the CC can be more discriminative in the snow scenario. The operation of these two priors can be expressed as:

$$\mathbf{I}_{Dark}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} \mathbf{I}^c(x) \right), \tag{6}$$

$$\mathbf{I}_{Bright}(x) = \max_{y \in \Omega(x)} \left(\max_{c \in \{r,g,b\}} \mathbf{I}^c(x) \right), \tag{7}$$

where $\mathbf{I}_{Dark}$ and $\mathbf{I}_{Bright}$ demonstrate the dark channel and the bright channel operations. In Figure 5, we present the dark channel, the bright channel, and the contradict channel of snow images, and their corresponding ground truths. One can see that, although the dark channel in snow images can retrieve the veiling effect information, the snow cannot be extracted effectively because the dark channel operation cannot detect the snow particles due to the minimum operation in the local patch. Regarding the bright channel, though some snow information can be acquired, the representation of snow may be limited due to the maximum operation in the color channel, which may force the snow-free regions with red, blue and green color to become one. It degrades the discriminative ability of the bright channel for the snow scene. Based on the analysis above, we can conclude that the contradict channel can be a discriminative feature for snow removal. The intensity of the contradict channel can be regarded as a natural metric to distinguish snow-free images from snow images. Motivated by this property, we construct a novel loss called the contradict channel loss to benefit the training process of desnowing. The contradict channel loss $\mathcal{L}_{CC}$ can be defined as:

$$\mathcal{L}_{CC} = ||CC(\mathbf{J}) - CC(\mathbf{J}_{GT})||_1, \tag{8}$$

where $CC()$ represents the contradict channel operation. $\mathcal{L}_{CC}$ indicates the contradict channel loss, $\mathbf{J}$ denotes the predicted snow-free image, and $\mathbf{J}_{GT}$ is the corresponding ground truth.

3.4. Loss Function

The proposed network adopts three losses: (i) the complex wavelet loss $\mathcal{L}_{CW}$, (ii) the perceptual loss $\mathcal{L}_{Per}$ [33], and (iii) the contradict channel loss $\mathcal{L}_{CC}$.

Complex Wavelet Loss. The complex wavelet loss $\mathcal{L}_{CW}$ is proposed and defined as follows:

$$\mathcal{L}_{CW} = \sum_i \left[\sqrt{\left| (\widetilde{\eta}_i^{\pi} - \eta_i^{\pi}) + j(\widetilde{\eta}_i^{\pi} - \eta_i^{\pi}) \right|^2 + \epsilon^2} \right], \tag{9}$$

where $\pi \in \{\mathrm{H}, \mathrm{L}\}$. 'H' and 'L' denote the wavelet coefficients in high-frequency and low-frequency domains. η_i^{π} and $\widetilde{\eta}_i^{\pi}$ represent estimated wavelet coefficients and the corresponding ground truths at the i^{th} level, respectively. ϵ denotes the slack value which can prevent the high-frequency wavelet coefficients from being zero to keep the texture details. Thus, we set $\epsilon = 0.02$ when calculating the loss in the high-frequency domain and set $\epsilon = 0$ in the low-frequency domain.

Overall Loss. In the proposed network, the overall loss function is expressed as:

$$\mathcal{L}_{Overall} = \mathcal{L}_{CW} + \lambda_1 \mathcal{L}_{CC} + \lambda_2 \mathcal{L}_{Per}, \tag{10}$$

where $\mathcal{L}_{Per}$ denotes the perceptual loss [33]. In this paper, we set λ_1 and λ_2 as 2 and 0.1, respectively.

4. Experiments

4.1. Data Generation and Implementation Detail

To reflect the complicated constitution in real-world snow scenarios, we construct a large-scale dataset called the Comprehensive Snow Dataset (CSD). It consists of 10000 synthesized snow images. In the synthesis process, first, we apply the famous haze dataset called RESIDE [39] and follow the synthesized procedure in [2] to simulate the veiling

Table 1: **Quantitative analysis on three existing snow datasets.** The proposed method can achieve the best results on all datasets, which verifies the robustness of the proposed method. 'wDH' and 'woDH' denote the cases with and without dehazing by the [34], respectively. Red numbers indicate the best results, and blue numbers are the second best results.

Dataset (size)		Methods (PSNR/SSIM/CIEDE 2000)							
		Zheng [5]	Eigen [35]	DesnowNet [1]	CycleGAN [36]	DAD [37]	All in One [38]	JSTASR [2]	Ours
Snow100K (2000)		23.86/0.82/3.90	20.12/0.72/6.40	30.50/0.94/2.31	26.81/0.89/5.73	24.93/0.84/7.61	26.07/0.88/5.53	23.12/0.86/6.31	31.54/0.95/2.18
SRRS (2000)	wDH	19.12/0.80/8.25	17.36/0.66/9.69	20.38/0.84/7.53	20.21/0.74/8.20	24.31/0.86/6.41	24.98/0.88/4.98	25.82/0.89/7.18	27.78/0.92/3.64
	woDH	15.78/0.74/12.28	14.25/0.62/14.46	16.22/0.79/12.14					
CSD (2000)	wDH	20.21/0.79/7.56	18.63/0.66/8.62	20.13/0.81/6.01	20.98/0.80/6.38	24.33/0.85/7.38	26.31/0.87/5.13	27.96/0.88/6.47	29.06/0.91/3.22
	woDH	13.98/0.67/15.69	12.71/0.55/17.96	17.63/0.75/9.43					

(a) (b) (c)

Figure 6: **Examples of the proposed CSD dataset.** (a) Input; (b) Ground truth; (c) Combined snow mask

effect. Then, we synthesize snowflakes and snow streaks with different properties (i.e., transparency, size, and location) based on the snow synthesis tutorial of Photoshop. Furthermore, we adopt Gaussian blurring on snow particles to better simulate the real-world snow scenarios. An example of the snow image in the CSD dataset is shown in Figure 6.

For the implementation detail, the learning rate is set as 10^{-4} and the Adam [40] optimizer is applied. We adopt the 4-level DTCWT. The proposed network was trained on a Nvidia RTX Titan GPU and trained with 1000 epochs with batch size 8. The proposed network is implemented on Tensorflow platform. For the training data, we use 8000 images from the CSD dataset. In each epoch, we split 30% training data as the validation set.

4.2. Quantitative Evaluations

In this section, four existing methods which are designed for the snow removal are adopted for comparison: Zheng *et al.* [5], Eigen *et al.* [35], DesnowNet [1], and JSTASR [2] (GAN-based method). Moreover, we adopt three methods (i.e., DAD [37], All in One [38], and CycleGAN [36]) which can address the degradation by different inclement weathers. For an apples to apples comparison, we retrain each model (if the original training code is provided) based on our training dataset and report the best result. We evaluate the proposed method on three datasets: Snow-100K [1], SRRS [2], and CSD. The results are shown in Table 1. The number of images in each dataset is shown after the name.

Image quality analysis. For quantitative evaluation, we apply three metrics: the structural similarity (SSIM), the peak signal to noise ratio (PSNR), and the CIEDE2000 color difference. From Table 1, one can see that the proposed method outperforms other desnowing methods in all metrics. Compared with the second best algorithm, the PSNR, the SSIM, and the CIEDE 2000 of the proposed method are **3.9%** higher, **3.4%** higher, and **37.2%** lower in the CSD dataset, respectively. It verifies that the proposed method can achieve the best performance in image quality and color fidelity.

Table 2: **Comparison of run time and the number of required parameters between the proposed method and other learning-based methods for snow removal.**

	DesnowNet [1]	JSTASR [2]	Ours
Time (s)	1.38	0.87	**0.14**
Parameters	1.56×10^7	6.5×10^7	$\mathbf{6.99 \times 10^6}$

Run time analysis. In Table 2[1], the run time and the number of required parameters analysis are presented. We compare our method with other learning-based snow removal algorithms. The results indicate that the proposed method requires much less computational resources than other desnowing methods. The proposed method saves at least **83%** of running time and **55%** of parameters.

4.3. Qualitative Evaluations

In Figure 7, some desnowing results from the real-world dataset by using the proposed algorithm and other algorithms are compared. One can see that the proposed method can remove more snow particles compared to other methods. For the images recovered by other methods, the snow particles and the veiling effect (see the 2^{nd} and the 5^{th} row) cannot be removed clearly. Moreover, the results in the 3^{th} row indicate that the proposed method can remove the snow streaks effectively compared with other methods. Therefore, the results prove that the proposed method has better generalization ability on real-world scenarios.

[1]This result is tested on Nvidia GTX 1080ti GPU.

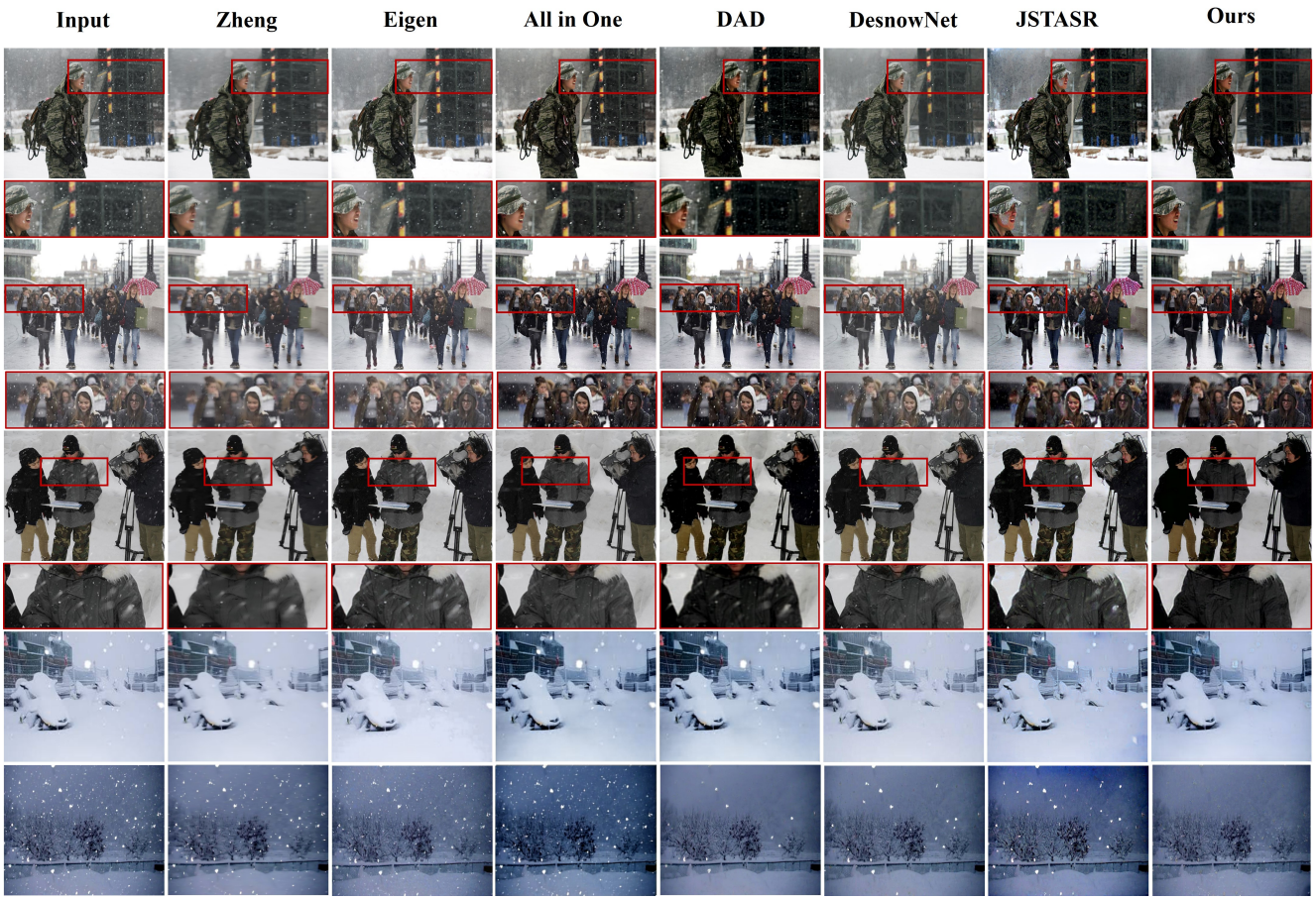


Figure 7: **Visual comparison between the proposed method and existing snow removal strategies.**

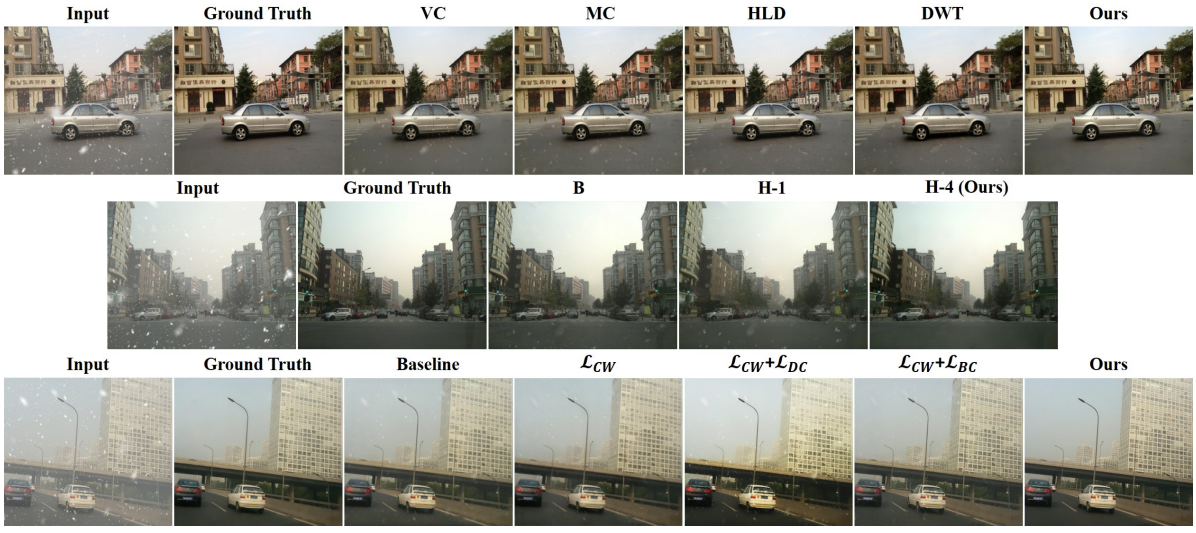


Figure 8: **Visual comparison for ablation study.** We prove the effectiveness of DTCWT (1^{st} row), hierarchical architecture (2^{nd} row), and loss functions (3^{rd} row). (Zoom-in for better visual quality.)

Table 3: **Comparing the results that apply different feature extraction techniques.**

Metric	VC	MC	HLD	DWT	Ours
PSNR/SSIM	22.18/0.81	24.31/0.83	26.13/0.86	27.92/0.88	**29.06/0.91**

Table 4: **Ablation study of the proposed hierarchical architecture.** Note that, **H-4** is the proposed architecture in this paper.

Metric	H-1	H-2	H-3	B	H-4	H-5
PSNR/SSIM	25.31/0.83	27.51/0.88	28.49/0.89	27.33/0.87	**29.06/0.91**	28.81/0.90

Table 5: **Verification for effectiveness of proposed loss functions.**

Metric	Baseline	$\mathcal{L}_{CW}$	$\mathcal{L}_{CW}$+$\mathcal{L}_{DC}$	$\mathcal{L}_{CW}$+$\mathcal{L}_{BC}$	Ours
PSNR/SSIM	26.13/0.87	27.33/0.88	27.43/0.87	27.53/0.88	**29.06/0.91**

4.4. Ablation Study

To verify the effectiveness of each proposed module, we divided the ablation study into three parts: effectiveness of the DTCWT, the hierarchical architecture, and the loss functions. All the experiments are evaluated and retrained on the CSD dataset.

Effectiveness of DTCWT. To prove that leveraging DTCWT can achieve the better performance for network training, we replace DTCWT with four conventional feature extraction techniques in the proposed network. In Table 3 and Figure 8, we adopt the vanilla convolution with kernel size 3×3 (**VC**), the multi-convolution (**MC**) [41], image-smoothing based high-low frequency decomposition (**HLD**) [39, 42], and the DWT [24]. For a fair comparison, except for the feature extraction technique, the rest part of the network is the same architecture. From Table 3 and the first row of Figure 8, we can find that using the DTCWT for snow feature extraction can achieve much better recovered results compared with other methods. Moreover, we present the visual comparison for different settings. One can observe that, with the embedding of the DTCWT, the desnowed results can be more desirable than others.

Effectiveness of hierarchical architecture. To verify the effectiveness of the hierarchical architecture, we construct several settings for comparison, that is, (i) the proposed hierarchical architecture which only decomposes the low-frequency component in each level but with different level of the DTCWT decomposition (**H-1** to **H-5**); (ii) the proposed hierarchical architecture (4-level decomposition) but both high and low-frequency component are decomposed in each level (**B**). The results are shown in Table 4 and the second row of Figure 8. One can see that, the proposed architecture (**H-4**) achieves the superior performance on snow removal. The reason is that, for (i), one-level decomposition cannot well decompose the snow particles into the small scaled ones adequately. However, too much level of decomposition may deteriorate the final result. For example, for the input image size (480×640), the subband in the bottom level may lose spatial information after the 5-level decomposition. In addition, for the setting in (ii), the performance compared to our method is limited because the number of recovered subbands is approximately four times greater than that of **H-4** and recovering excessive subbands may degrade the quality of the results.

Effectiveness of loss functions. We evaluate the effectiveness of the proposed loss functions, that is, $\mathcal{L}_{CW}$ and $\mathcal{L}_{CC}$. Moreover, we compared the $\mathcal{L}_{CC}$ with the dark channel loss ($\mathcal{L}_{DC}$) and the bright channel loss $\mathcal{L}_{BC}$. The comparison is demonstrated in Table 5 and the third row of Figure 8. 'Baseline' indicates the $L2$ loss on the final recovered result. Each combination in this experiment is trained with the perceptual loss. The results indicate that with the proposed two loss functions, the performance of the network can be improved greatly compared with the baseline. Moreover, compared with other existing loss functions (i.e., $\mathcal{L}_{DC}$ and $\mathcal{L}_{BC}$), the CC loss can acquire more significant improvement because the contradict channel is a discriminative feature for snow scenarios.

5. Conclusion

In this paper, a novel desnowing network to address the limitations in current desnowing methods was proposed. First, to handle the complicated snow scenario, a hierarchical network based on the DTCWT was developed. With the better representation of snow features, the snowflakes, snow streaks, and the veiling effect can be removed effectively. Second, we develop a novel prior in snow images called the contradict channel prior (CCP). Based on it, the contradict channel loss was designed to improve the desnowing performance. Last, a large-scale snow dataset called CSD was proposed. It can interpret the real-world snow scenarios effectively compared with previous datasets. Experimental results showed that the proposed method outperforms state-of-the-art desnowing algorithms in desnowing and can benefit high-level vision tasks.

6. Acknowledgement

We thank to National Center for High-performance Computing (NCHC) for providing computational and storage resources.

References

[1] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "Desnownet: Context-aware deep network for snow removal," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3064–3073, 2018. 1, 2, 3, 6

[2] W.-T. Chen, H.-Y. Fang, J.-J. Ding, C.-C. Tsai, and S.-Y. Kuo, "Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal," in *European Conference on Computer Vision*, 2020. 1, 2, 3, 4, 5, 6

[3] Z. Li, J. Zhang, Z. Fang, B. Huang, X. Jiang, Y. Gao, and J.-N. Hwang, "Single image snow removal via composition generative adversarial networks," *IEEE Access*, vol. 7, pp. 25 016–25 025, 2019. 1, 2

[4] J. Xu, W. Zhao, P. Liu, and X. Tang, "An improved guidance image based method to remove rain and snow in a single image," *Computer and Information Science*, vol. 5, no. 3, p. 49, 2012. 2

[5] X. Zheng, Y. Liao, W. Guo, X. Fu, and X. Ding, "Single-image-based rain and snow removal using multi-guided filter," in *International Conference on Neural Information Processing*. Springer, 2013, pp. 258–265. 2, 6

[6] S.-C. Pei, Y.-T. Tsai, and C.-Y. Lee, "Removing rain and snow in a single image using saturation and visibility features," in *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2014, pp. 1–6. 2

[7] Y. Wang, S. Liu, C. Chen, and B. Zeng, "A hierarchical approach for rain or snow removing in a single color image," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3936–3950, 2017. 2

[8] S. Yu, Y. Zhao, Y. Mou, J. Wu, L. Han, X. Yang, and B. Zhao, "Content-adaptive rain and snow removal algorithms for single image," in *International Symposium on Neural Networks*, 2014, pp. 439–448. 2

[9] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE signal processing magazine*, vol. 22, no. 6, pp. 123–151, 2005. 2, 3

[10] T. Yan, Y. Ding, F. Zhang, N. Xie, W. Liu, Z. Wu, and Y. Liu, "Snow removal from light field images," *IEEE Access*, vol. 7, pp. 164 203–164 215, 2019. 2

[11] V. Voronin, E. Semenishchev, M. Zhdanova, R. Sizyakin, and A. Zelenskii, "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, 2019, p. 111690. 2

[12] Z. Li, J. Zhang, Z. Fang, B. Huang, X. Jiang, Y. Gao, and J.-N. Hwang, "Single Image Snow Removal via Composition Generative Adversarial Networks," *IEEE Access*, vol. 7, pp. 25 016–25 025, 2019. 2

[13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261*, 2016. 3

[14] M. Ghorat, G. B. Gharehpetian, H. Latifi, and M. A. Hejazi, "A new partial discharge signal denoising algorithm based on adaptive dual-tree complex wavelet transform," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 10, pp. 2262–2272, 2018. 3

[15] P. Luo, X. Qu, X. Qing, and J. Gu, "Ct image denoising using double density dual tree complex wavelet with modified thresholding," in *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*. IEEE, 2018, pp. 287–290. 3

[16] X. Qu, J. Dong, and S. Niu, "shallowcnn-le: A shallow cnn with laplacian embedding for face anti-spoofing," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–8. 3

[17] L. Wang, A. Mohammad-Djafari, and N. Gac, "Bayesian method with sparsity enforcing prior of dual-tree complex wavelet transform coefficients for x-ray ct image reconstruction," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 478–482. 3

[18] A. Singh and N. Kingsbury, "Efficient convolutional network learning using parametric log based dual-tree wavelet scatternet," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 1140–1147. 3

[19] D. Li, L. Zhang, C. Sun, T. Yin, C. Liu, and J. Yang, "Robust retinal image enhancement via dual-tree complex wavelet transform and morphology-based method," *IEEE Access*, vol. 7, pp. 47 303–47 316, 2019. 3

[20] H. Lu, H. Wang, Q. Zhang, D. Won, and S. W. Yoon, "A dual-tree complex wavelet transform based convolutional neural network for human thyroid medical image segmentation," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2018, pp. 191–198. 3

[21] H. Chen and N. Kingsbury, "Efficient registration of nonrigid 3-d bodies," *IEEE transactions on image processing*, vol. 21, no. 1, pp. 262–272, 2011. 3

[22] T. Sun and C. Jung, "Readability enhancement of low light images based on dual-tree complex wavelet transform," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 1741–1745. 3

[23] C. Jung, Q. Yang, T. Sun, Q. Fu, and H. Song, "Low light image enhancement with dual-tree complex wavelet transform," *Journal of Visual Communication and Image Representation*, vol. 42, pp. 28–36, 2017. 3

[24] B. L. Sturm, "Stéphane Mallat: A Wavelet Tour of Signal Processing," 2007. 3, 8

[25] H.-H. Yang and Y. Fu, "Wavelet u-net and the chromatic adaptation transform for single image dehazing," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2736–2740. 3

[26] H.-H. Yang, C.-H. H. Yang, and Y.-C. J. Tsai, "Y-net: Multi-scale feature aggregation network with wavelet structure similarity loss function for single image dehazing," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2628–2632. 3

[27] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE TPAMI*, 2020. 4

[28] W.-T. Chen, J.-J. Ding, and S.-Y. Kuo, "PMS-Net: Robust Haze Removal Based on Patch Map for Single Images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 681–11 689. 4

[29] W.-T. Chen, H.-Y. Fang, J.-J. Ding, and S.-Y. Kuo, "Pmhld: Patch map based hybrid learning dehazenet for single image haze removal," *IEEE Transactions on Image Processing*, 2020. 4

[30] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large Kernel Matters-Improve Semantic Segmentation by Global Convolutional Network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4353–4361. 4

[31] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010. 5

[32] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4003–4011. 5

[33] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*, 2016, pp. 694–711. 5

[34] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2808–2817. 6

[35] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 633–640. 6

[36] D. Engin, A. Genç, and H. K. Ekenel, "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018. 6

[37] Z. Zou, S. Lei, T. Shi, Z. Shi, and J. Ye, "Deep adversarial decomposition: A unified framework for separating superimposed images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 806–12 816. 6

[38] R. Li, R. T. Tan, and L.-F. Cheong, "All in one bad weather removal using architectural search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3175–3185. 6

[39] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018. 5, 8

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 6

[41] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European conference on computer vision*. Springer, 2016, pp. 154–169. 8

[42] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Fast end-to-end trainable guided filter," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1838–1847. 8