Robust Computer Vision in Snow

Seo Young Kim
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
seoyoung@andrew.cmu.edu

Chen Zhao
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
chenzha3@andrew.cmu.edu

Chang Liu
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213
changl7@andrew.cmu.edu

1 Introduction

1.1 Project Motivation

Autonomous vehicles (AVs) can have trouble driving safely on snowy roads because of reduced visibility and slippery road surfaces. Reduced visibility limits the perception performance and slippery road surfaces are a big challenge for vehicle control. Due to these reasons, images captured in snowy days suffer from noticeable decrease of scene visibility, which decreases the performance of vision-based intelligent systems. Therefore, training robust networks that can detect objects in snowy weather is an important topic in computer vision.

Comparing with other atmospheric phenomenon such as haze and rain, snow is more complicated due to its transparency, size variability, and accumulation of veiling effect, which make image de-snowing more challenging [5].

1.2 Project Summary

Our project aims to build a robust Computer Vision in snowy weather. We plan to use the following datasets: Comprehensive Snow Dataset [5], Snow Removal in Realistic Scenario [4], and Snow100K [10]. The biggest challenge is the variability in data. It is important to differentiate patterns representing snow from other valid objects within an image. We plan to achieve this goal by applying snow removal techniques and implementing Mask R-CNN [7] for object detection.

1.3 Project Goals

As mentioned earlier, our goal is to improve the perception performance of AVs in snowy weather. We plan to use snow removal techniques to remove snow from image. We are trying different implementations to see which method fits our purpose the best. We are currently looking into Hierarchical Dual-tree Complex Wavelet Representation and Contradict Channel Loss approach based on the work proposed by Chen et al. [5] We are also looking into different post-processing techniques such as image de-noising and image quality enhancement. For object detection networks, we plan to use Mask R-CNN for the object instance segmentation.

To compare the performance, we will train the network with images with snow and images without snow (snow removal techniques applied). Then, we will run CARLA on both of these cases to simulate the comparison.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

1.4 Dataset

Comprehensive Snow Dataset (CSD): This dataset is proposed by Chen et al. [5] in order to consider veiling effect and 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. It consists of 10,000 synthesized snow images.

Snow Removal in Realistic Scenario (SRRS): This large scale snow dataset is proposed by Chen et al. [4]. It contains 15,000 synthesized snow images and 1000 snow images in real scenarios downloaded from the Internet. Moreover, in order to test the generalization of the proposed network, the test dataset proposed by Liu et al. [10] is applied.

Snow100K: This dataset is created by Liu et al. [10]. In order to model snowy images, they first produce 5,800 snowy masks and download 100K clean images. Then snowy images are synthesized based on the clean images and snowy masks. This dataset provides three kinds of snow, i.e., small, medium, and large particle sizes. It also provides 1,329 realistic snowy images to evaluate models in terms of generalization in the real world

Figure 1: Example visual results of snow removal. The goal of image de-snowing is to remove snow and restore high-quality snow-free images. The various scene, occlusion and illumination variations of objects make it difficult to accomplish this task.

1.5 Use Cases

There are many use cases of this technique. Some obvious applications are in image processing, computer vision, and autonomous driving. More specifically, it can be helpful for image restoration, object detection, object tracking, semantic segmentation, and scene analysis with possibly combinations of these applications such as restoring blurry snowy image and using it for scene analysis. We are most interested in applying the technique to the autonomous driving domain. We can use it for navigation in bad weather for taxi, truck, or delivery. We can also use it for path finding to avoid ice and detect lanes under snow.

2 Related Work

2.1 Snow Removal

For snow removal in the single image [19, 22, 11], Zheng et al. [22] 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. [18] proposed a three-layer hierarchical scheme which combines image decomposition and dictionary learning. Voronin et al. [17] developed the anisotropic gradient in Hamiltonian quaternions to remove rain and snow. Li et al. [9] applied the generative adversarial network (GAN) for snowflake removal. Liu et al. [10] proposed a learning-based snow removal architecture called the DesnowNet.

(a) Input and DesnowNet results

(b) JSTASR and ALL Snow Removed results

Figure 2: (a) Input and recovered results by the DesnowNet[10] (b) Recovered results by the JSTASR [4], and our method [5]. You can see that the proposed method can achieve better performance on snow removal.

2.2 Haze Removal

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. He et al. [8] 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 learningbased 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. [14] proposed the multi-scale CNN to predict the transmission map.

2.3 Semantic Segmentation

In contrast to the classification tasks, a label is assigned to each pixel in an image in semantic segmentation. Several pre-trained networks can predict semantic labels based on the coarsely desnowed images. After that, they can learn a semantic-aware representation under the guidance of semantic labels. The network processes the input features and semantic labels. The output is inputted into a softmax function to transfer a set of attention weights, each of them corresponds to a type of objects. After that, group convolution is used to process the semantic-aware representation and merge all the features from different groups via a 1×1 CNN layer [21].

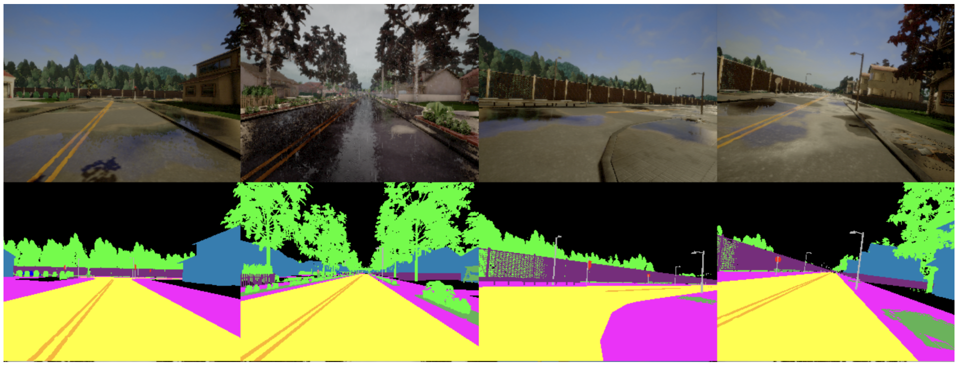

Figure 3: Example visual results of the proposed architecture predicting the semantic segmentation image on CARLA simulator.

3 Methodology

3.1 ALL Snow Removed

In order to remove snow from images, we plan to use hierarchical dual-tree complex wavelet representation and contradict channel loss [5]. In this literature, the authors propose a single image de-snowing algorithm to address the diversity of snow particles in shape and size. First, to better represent the complex snow shape, they apply the dual-tree wavelet transform and propose a complex wavelet loss in the network. Second, they propose a hierarchical decomposition paradigm in the network for better understanding the different sizes of snow particles. Last, they propose a novel feature called the contradict channel (CC) for the snow scenes. They find that the regions containing the snow particles tend to have higher intensity in the CC than that in the snow-free regions. Therefore, they leverage this discriminative feature to construct the contradict channel loss for improving the overall performance of snow removal. Comparison between state-of-the-art snow removal algorithms including the work we are implementing is shown in Figure 2.

3.2 Mask R-CNN

Mask R-CNN is a convolutional neural network (CNN) [7] and state-of-the-art in terms of image segmentation. This variant of a deep neural network detects objects in an image and generates a high-quality segmentation mask for each instance.

Traditional CNN consists of a convolutional layer, a pooling layer, and a fully connected layer. The convolutional layer helps to abstract the input image as a feature map via the use of filters and kernels. The pooling layer helps to downsample feature maps by summarizing the presence of features in patches of the feature map. The fully connected layer connects every neuron in one layer to every neuron in another layer. Although CNN is powerful, the traditional CNN can only predict pictures of a specific object.

Region-based Convolutional Neural Network (R-CNN) is built on the traditional CNN. It adds a pre-processing step by extracting all the regions of interest (ROI) and classifying them independently using CNN. This neural network can identify different objects in a single image. R-CNN significantly increases the applicability of CNN. Based on the R-CNN, many improvements have been made and new frameworks are derived.

Fast R-CNN[13] is an improved version of the R-CNN architecture with two stages: region proposal network and R-CNN. Region Proposal Network (RPN) is a network trained to generate "proposals" for the region where the object lies. Fast R-CNN uses it to extract ROI efficiently. The performance of R-CNN is also improved by performing convolution operation only once per image, which significantly increases its performance. Faster R-CNN is an optimized form of R-CNN because it uses the selective search [16] for generating regions of interest, while Faster R-CNN uses a "Region Proposal Network"(RPN).

Mask R-CNN derives from Faster R-CNN and improves its performance. Mask R-CNN performs two main types of image segmentation: semantic segmentation and instance segmentation. Semantic segmentation classifies each pixel into a fixed set of categories without differentiating object instances, while instance segmentation could precisely segment each instance. It is the combination of object detection, object localization, and object classification.

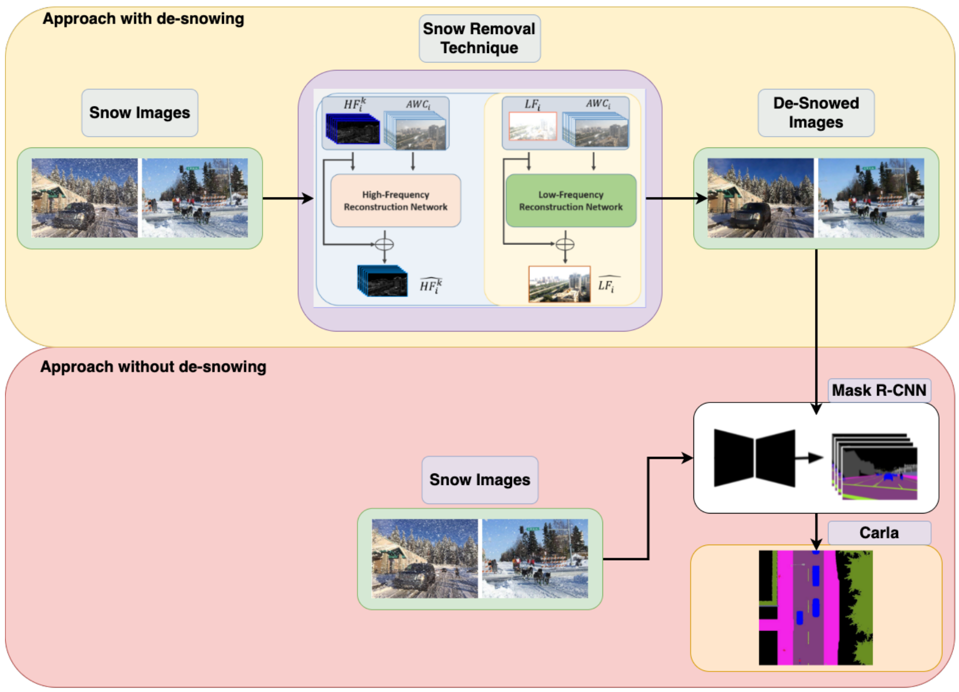

Figure 4: The overall flow chart of the proposed approach to tackle the challenging weather problem in perception. The proposed method has two different approaches: approach with de-snowing and approach without de-snowing. For the approach with de-snowing, the method consists three parts: snow removal, semantic segmentation, and CARLA simulation. For the apporach without de-snowing, the method consists two parts: semantic segmentation and CARLA simulation

3.3 CARLA

For the final simulation, we will use CARLA [6]. CARLA is an open-source autonomous driving simulator grounded on Unreal Engine. It serves as a modular and flexible API to address a range of tasks involved in the problem of autonomous driving. CARLA supports flexible specification of sensor suites and environmental conditions. In our project, CARLA will be used for our final simulation and demonstration. Figure 3 shows an example visual results of our proposed method on CARLA.

3.4 Extra Credit: YOLO

YOLO [12], “You Only Look Once”, is a famous framework for object detection. In comparison to Mask R-CNN, YOLO takes a totally different approach. It applies a single neural network to the full image, then divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities. YOLO is popular because it achieves high accuracy while also being able to run in real-time. Compared with Mask R-CNN, it runs faster, but lacks semantic segmentation.

If we have time to extend our ablation study, we plan to include the YOLO in our architecture instead of Mask R-CNN and run an end-to-end training and testing. We will then compare this result with the result with Mask R-CNN to analyze which method is better.

4 System Design and Evaluation

4.1 Approach with Snow Removal

In this approach, we plan to first process the dataset using a hierarchical dual-tree complex wavelet representation approach for snow removal. Then, the dataset will be used as an input for the Mask R-CNN network in order to generate the semantic segmentation predictions. Then, with the applications of some post-processing techniques such as image denoising, the segmented data will be sent to CARLA for the final simulation. The top block of the Figure 4 shows the block diagram of this approach.

4.2 Approach without Snow Removal

In this approach, we plan to generate semantic segmentations based on the raw images with snow using the mask R-CNN network. After applying some post-processing techniques, the data will be sent to CARLA for the final simulation. The bottom block of the Figure 4 shows the block diagram of this approach.

4.3 Evaluation

In order to quantitatively evaluate how well the snow removal technique performs, we plan to apply two metrics: the structural similarity (SSIM) and the peak signal to noise ratio (PSNR).

In order to evaluate the performance of semantic segmentation, we plan to use IoU. IoU, also known as the Jaccard Index, stands for Intersection over Union. It is the most popular evaluation metric used in the object detection benchmarks. As Figure 5 shows[20], IoU is the area of the overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. The IoU metric ranges from 0–1 with 0 signifying no overlap while 1 signifying perfectly overlapping segmentation. The mean of IoU for multi-class segmentation is calculated by averaging the IoU of each class. We will use Keras for our implementation of evaluation metric.

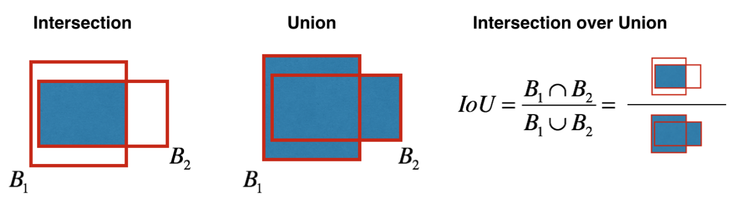

Figure 5: Intersection over Union calculation

5 Demonstration

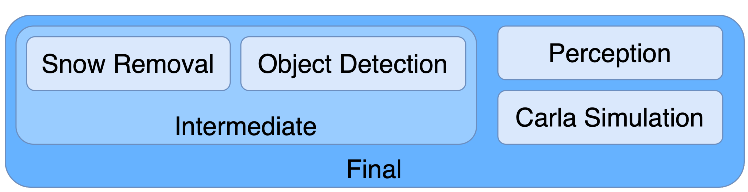

Figure 6: demonstration structure

Our demonstrations can be divided into four parts: snow removal, object detection, perception, and CARLA simulation. The detail structure is shown in Figure 6.

5.1 Intermediate Demonstrations

The intermediate demonstration mainly focuses on the snow removal and object detection. It mainly focuses on demonstrating the feasibility of the scheme.

5.1.1 Snow Removal Demo

The snow removal demonstration aims to visualize and evaluate the validity of the snow removal techniques. In this demonstration, snowing images are picked randomly from the test dataset. Then, snow removal techniques will be applied on them and output the snow-removed images. For the quantitative evaluation, we will apply SSIM and PSNR metrics as mentioned earlier. Additionally, we will put the original and de-snowed images side-by-side for the qualitative evaluation.

5.1.2 Object Detection Demo

The object detection demonstration aims to visualize and evaluate the performance of the object detection on the snow-removed dataset. In this demonstration, snowing images are picked randomly from the snow-removed dataset. Semantic segmentation will be performed on them and output the labeled images. The labeled result will include class labels, bounding boxes, and object masks. Sample output is shown in Figure 7. For the quantitative evaluation, we will apply IoU [15]. Additionally, we will put the de-snowed and labeled images side-by-side for the illustration.

Figure 7: Expected object detection demo output [7]

5.2 Final Demonstrations

The final demonstrations will utilize the previously mentioned contents and built upon them. Final demonstrations will focus on the perception and CARLA simulation. It will mainly focus on demonstrating the results and efficiency of the scheme.

5.2.1 Perception Demo

The perception demonstration aims to visualize and evaluate the performance of the perception module we will be developing in the CARLA simulation environment. In this demonstration, dataset will be created by running two identical simulations with snow as the only variable. The experiment (snowing) dataset will be collected in snowing simulation environment and the controlled (not snowing) dataset will be collected in not snowing simulation. Perception module with snow-removal will be performed on the experiment dataset, while another perception module without snow-removal is being performed on the controlled dataset. Two perception modules will both have labeled results, which contain class labels, bounding boxes, and object masks. Evaluation metrics for this demonstration will be IoU [15] and precision-recall curve (PR Curve). We will put the de-snowed and labeled images side-by-side for the illustration.

5.2.2 CARLA Simulation Demo

The CARLA simulation demonstration aims to visualize and evaluate the performance of the system we will be developing in the CARLA simulation environment. In this demonstration, a snowing simulation environment will be set up. Three simulations will be performed separately. The first one will use the snow removal techniques and Mask R-CNN. The second one will only use Mask R-CNN. The last one will run with the baseline perception module. The heaviness of snow will be the only variable. The behavior of the simulated autonomous vehicle will be monitored and analyzed. Evaluation methods for this demonstration will be mean time between failure (MTBF). We will put their behavior side-by-side for the comparison.

5.3 Timeline and Work Partition

#Week	Date	Progress	#Week	Date	Progress
5	2/14	Snow Removal & Object Detection	11	3/28	Perception & Carla & Refine
6	2/21		12	4/4	
7	2/28		13	4/11	Tune, Test and Refine
8	3/7	Tune, Test and Refine	14	4/18	Summarize Project
9	3/14	Intermediate Summary	15	4/25	**Final Presentation**
10	3/21	**Intermediate Project Demo**			

Figure 8: project timeline

Before the intermediate demonstration, snow removal and object detection will be our focal point. After the intermediate demonstration, perception and CARLA simulation will become our focal point. The overall timeline is shown in Figure 8 and our work partition is shown in Figure 9.

	Towards Intermediate Demo	Towards Final Demo
Seo Young Kim	Snow Removal	Refine Snow Removal & Object Detection
Chen Zhao	Carla Simulation	Perception & Carla Simulation
Chang Liu	Object Detection	Perception & Carla Simulation

Figure 9: work partition

6 Conclusion

In conclusion, snow weather do challenge the robustness of Autonomous Vehicles. With this opportunity, we want to use computer vision techniques to improve the perception performance of AVs in snowy weather conditions.

References

[1] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, 2016.

[2] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *CoRR*, abs/1601.07661, 2016.

[3] Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen Kuo. Pms-net: Robust haze removal based on patch map for single images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11673–11681, 2019.

[4] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Chen-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *European Conference on Computer Vision*, 2020.

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

[6] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. 2017.

[7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.

[8] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, 2009.

[9] Zhi Li, Juan Zhang, Zhijun Fang, Bo Huang, Xiaoyan Jiang, Yongbin Gao, and Jenq-Neng Hwang. Single image snow removal via composition generative adversarial networks. *IEEE Access*, 7:25016–25025, 2019.

[10] 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.

[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 *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6, 2014.

[12] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[14] Wenqi Ren, Sibo Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, 2016.

[15] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[16] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, 2013.

[17] Viacheslav Voronin, Evgenii Semenishchev, Marina Zhdanova, Roman Sizyakin, and Alexander Zelenskii. Rain and snow removal using multi-guided filter and anisotropic gradient in the quaternion framework. page 26, 09 2019.

[18] 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 : a publication of the IEEE Signal Processing Society*, 26 8:3936–3950, 2017.

[19] Jing Xu, Wei Zhao, Peng Liu, and Xianglong Tang. An improved guidance image based method to remove rain and snow in a single image. *Comput. Inf. Sci.*, 5:49–55, 2012.

[20] Iffat Zafar and Giounona Tzanidou. *Hands-on convolutional neural networks with tensorflow: Solve Computer Vision problems with modeling in tensorflow and python*. Packt, 2018.

[21] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30:7419–7431, 2021.

[22] Xianhui Zheng, Yinghao Liao, Wei Guo, Xueyang Fu, and Xinghao Ding. Single-image-based rain and snow removal using multi-guided filter. pages 258–265, 11 2013.