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

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

1. Dual-tree Complex Wavelet Transformation

The DTCWT [1] is an improved form of the discrete wavelet transformation (DWT) [2]. Due to lack of the direction selective property, the DWT may not represent a 2-D signal effectively. Compared to DWT, the DTCWT achieve good direction selectivity and reduces the translation sensitivity of the DWT. The DTCWT overcomes the limitations of the DWT by introducing complex wavelets and the tree structure. The input signal is decomposed and reconstructed by two independent real number filter banks. The one-dimensional complex wavelet function can be expressed as:

$$\psi(x) = \psi_h(x) + j\psi_g(x), \tag{1}$$

where $j = \sqrt{-1}$, $\psi_h(x)$ and $\psi_g(x)$ are wavelet basis functions, and they are the real part and imaginary part of the mother wavelet, respectively. For the two-dimensional complex wavelet function, it can be defined as:

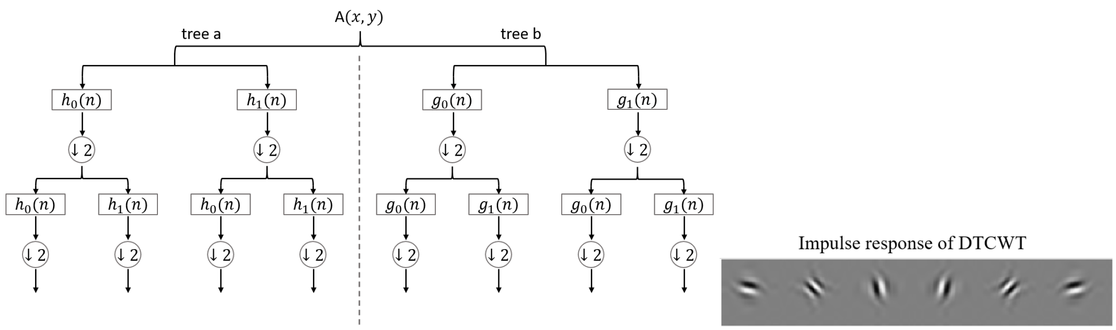
$$\psi(x, y) = \psi(x)\psi(y) = [\psi_h(x) + j\psi_g(x)][\psi_h(y) + j\psi_g(y)], \tag{2}$$

The detailed procedure of the 2-D DTCWT is demonstrated in Figure 1a. It consists of two branches of filters (i.e., tree a and tree b) with the same frequency responses and the filters in tree a should be reverse to tree b. For perfect reconstruction, the high-pass filter (HPF) and the low-pass filter (LPF) should form a Hilbert transform pair and the phase difference should be 90°. Moreover, the LPFs in two trees must differ by half a sample period. With this transform, dynamic texture feature extraction can be achieved. In the 2-D DTCWT, six high-pass subbands and one or two low-pass subbands at each level are produced. Therefore, more directions in edges (i.e., ±15°, ±45°, ±75°) and more image information can be acquired. Its illustration is shown in Figure 1b. Furthermore, the better ability of shift invariance can provide the high robustness to noise in high-frequency components.

2. Residue Estimation Block

The architecture of the residue estimation block (REB) is presented in Figure 2. In the REB, initially, the input is concatenated with the proposed aggregate wavelet component (AWC). Then, the input image is projected to the higher dimension space by several feature extraction modules. Each module consists of multi-pooling operations which have been defined in the regular paper and the Res2Net [3] architecture. Then, the short-cut operation is applied and several feature extraction

*Indicates equal contribution.



(a) (b)

Figure 1: **The illustration of the dual-tree complex wavelet transform (DTCWT).** (a) The pipeline of the DTCWT: Tree a and Tree b are real filters reverse to each other. The two trees represent real and imaginary parts of complex coefficients, respectively. Note that, A(x,y) is the input image, and $h_0(n)$ and $h_1(n)$ are low-frequency filter and high-frequency filter for real part while $g_0(n)$ and $g_1(n)$ are low-frequency filter and high-frequency filter for imaginary part. (b) The impulse response of the DTCWT with six orientations (i.e., ±15°, ±45°, ±75°).

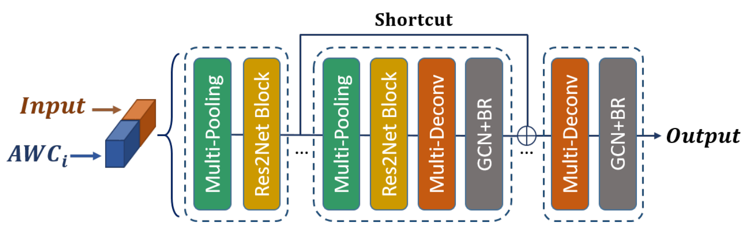
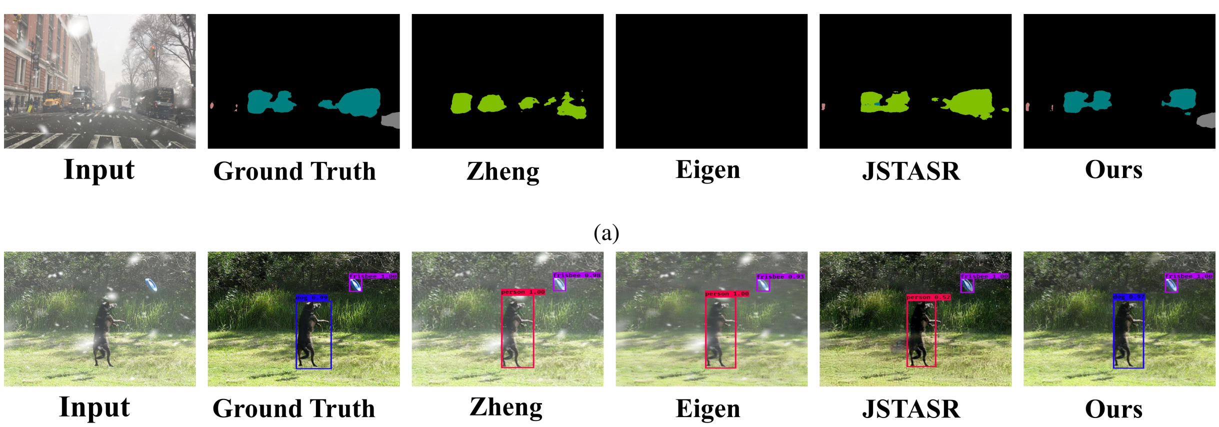


Figure 2: **The architecture of the residue estimation block (REB).**



(b)

Figure 3: **Visual comparison of high-level vision applications after desnowing** (a) Results of semantic segmentation; (b) Results of the object detection.

modules are used. For the decoder part, we apply several techniques including multi-deconvolution [4, 5], global convolution [6], and boundary refinement [6] to improve the performance of the network. The architectures of both HR and LR sub-networks are based on the proposed REB. However, more parallel kernels (i.e., kernel sizes with 2, 3, 5, 7, and 9) are used in LR and the filter depth of LR is wider than that of HR because low-frequency component recovery involves more complex semantic information.

Table 1: **Performances of object detection and semantic segmentation using the proposed and state-of-the-art methods for desnowing.** The datasets of COCO and BDD are adopted, respectively. Note that the baseline denotes the results of these applications without applying any desnowing method.

Metrics		Methods							
		Baseline	**Zheng** [7]	**Eigen** [8]	**DAD** [9]	**CycleGAN** [10]	**JSTASR** [11]	**All in One** [12]	**Ours**
Object Detection- Algorithm: YOLOv3 [13]; Dataset: COCO [14]									
mAP(%)	wDH	68.5	65.7	62.7	64.1	62.4	67.2	66.6	**79.3**
	woDH	64.9	57.2	61.6					
Semantic Segmentation- Algorithm: DeepLab [15]; Dataset: BDD [16]									
mIoU/mPA(%)	wDH	80.1/86.8	80.0/85.6	79.5/86.0	74.8/83.6	72.1/81.5	79.9/86.4	78.7/86.4	**84.6/90.4**
	woHD	80.6/87.6	75.2/81.7	77.2/84.8					

3. High-Level Vision Applications

To prove that the proposed method can benefit high-level vision applications, in Table 1 and Figure 3, we conduct several experiments for i) object detection and ii) semantic segmentation in snow scenarios. In these experiments, the proposed method or state-of-the-art methods are adopted as the pre-process technique for desnowing. First, we randomly select 650 images from COCO [14] and BDD [16] datasets, respectively. Then, we synthesize snowflakes, snow streaks, and the veiling effect based on the procedure in Section 4.1 of the regular paper. We adopt the snow removal algorithm and then apply YOLO-v3 [13] and DeepLab [15] for object detection and semantic segmentation on these images, respectively. From the results in Table 1 and Figure 3, one can see that the performance of these applications can be much improved by the proposed methods. Moreover, compared to state-of-the-art methods, using the proposed desnowing algorithm is even more helpful for improving the performance of high-level vision applications in snow scenarios.

References

[1] 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. 1

[2] B. L. Sturm, "Stéphane Mallat: A Wavelet Tour of Signal Processing," 2007. 1

[3] 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. 1

[4] 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. 2

[5] 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. 2

[6] 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. 2

[7] 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. 3

[8] 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. 3

[9] 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. 3

[10] 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. 3

[11] 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. 3

[12] 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. 3

[13] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. 3

[14] H. Caesar, J. Uijlings, and V. Ferrari, "Coco-stuff: Thing and stuff classes in context," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1209–1218. 3

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818. 3

[16] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving video database with scalable annotation tooling," *arXiv preprint arXiv:1805.04687*, vol. 2, no. 5, p. 6, 2018. 3