

# ContourletNet: A Generalized Rain Removal Architecture Using Multi-Direction Hierarchical Representation

## Supplementary Material

Wei-Ting Chen<sup>1,3\*</sup>

jimmy3505090@gmail.com

Cheng-Che Tsai<sup>1\*</sup>

r08943148@ntu.edu.tw

Hao-Yu Fang<sup>2</sup>

danielfang60609@gmail.com

I-Hsiang Chen<sup>2</sup>

f09921058@ntu.edu.tw

Jian-Jiun Ding<sup>2</sup>

jjding@ntu.edu.tw

Sy-Yen Kuo<sup>2</sup>

sykuo@ntu.edu.tw

<sup>1</sup> Graduate Institute of Electronics

Engineering

National Taiwan University

Taipei, Taiwan

<sup>2</sup> Department of Electrical Engineering

National Taiwan University

Taipei, Taiwan

<sup>3</sup> ASUS Intelligent Cloud Services

Asustek Computer Inc.

Taipei, Taiwan

## 1 Introduction of Contourlet Transform

The contourlet transform (CT) [1] is an efficient and effective technique for geometric information analysis which can achieve better representation in both locality and directionality. Compared to the wavelet transform (WT), the CT can represent the features in a sparser way because it provides better spatial locality, anisotropy, and directionality properties while the WT only extracts vertical and horizontal details. The comparison of the WT and the CT is shown in Figure 1. The overview of the CT is presented in Figure 2(a). It mainly consists of two operations: the Laplacian pyramid (LP) [2] and directional filter banks (DFB) [3]. Given an image, first, the LP filter decomposes it into the semantic subband (SS) and a high-frequency subband. Then, the high-frequency subband is further decomposed into several subspaces with  $2^k$  direction via directional filters where  $k$  is a positive integer. We term these subspaces multi-direction subbands (MSs). The illustration of multi-direction subbands is shown in Figure 2(b). Note that MSs mainly consist of residual rain streaks and textural information. For these components, the direction is very important. Therefore, the contourlet decompose the high-frequency subband into several subspaces according to directions to capture the edge and textural features with a variety of orientations. By the CT, multiresolution and multidirection decomposition can be achieved efficiently and abundant directional features can be extracted.

\* Equal Contribution

©2021. The copyright of this document resides with its authors.

It may be distributed unchanged freely in print or electronic forms.

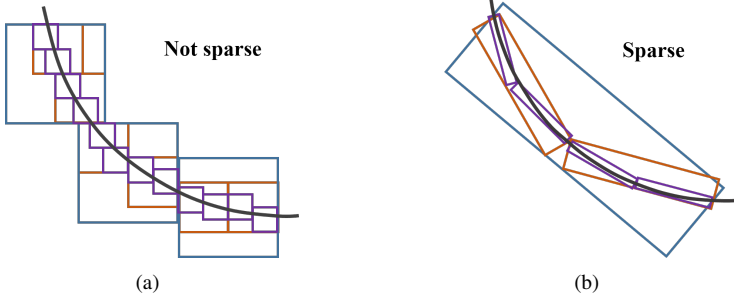


Figure 1: Using (a) the wavelet and (b) the contourlet to represent a bent curve. Note that a sparser representation can be achieved by the contourlet.

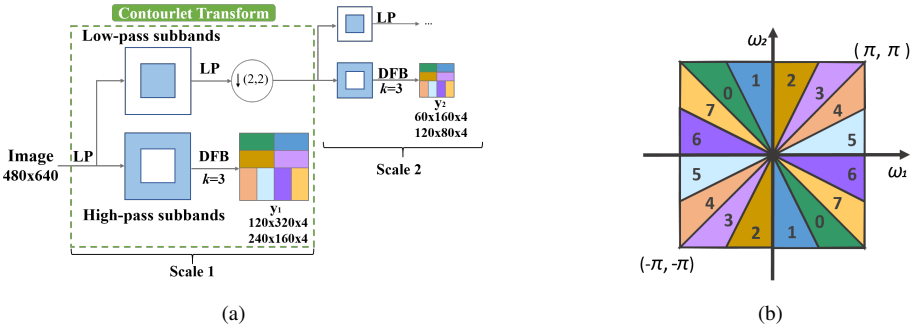


Figure 2: **Illustration of the Contourlet Transformation (CT).** (a) The CT with the hierarchical architecture. Note that the green dash box denotes the CT operation. (b) The multi-direction property of the CT. The eight-direction case is illustrated (i.e.,  $k = 3$ ).

## 2 Detail of the proposed network architecture

As we mentioned in the manuscript, the proposed ContourletNet consists of multi-direction subband recovery network (MSRN) and the semantic subband recovery network (SSRN). In each sub-network, to improve the performance of feature extraction and reconstruction, we propose a novel network architecture called the contourlet predictor (CP) shown in Figure 3. Initially, the input is concatenated with the proposed aggregate contourlet component  $G_i$  which is defined as

$$G_i = \sigma(\mathbf{I}) \oplus \sigma(\mathbf{SS}_1) \oplus \cdots \oplus \sigma(\mathbf{SS}_{i-1}), \quad (1)$$

where  $G_i$  denotes the aggregate contourlet component at the  $i^{th}$  level,  $\oplus$  presents the concatenate operation,  $\mathbf{I}$  is the rain image, and  $\mathbf{SS}_i$  is the SS component at level  $i$ , and  $\sigma$  is the multi-pooling architecture [24]. The idea of (1) has been illustrated in the manuscript. For the encoder part of the CP, the Res2Net [25] is adopted to perform feature extraction. For the decoder part, the techniques of multi-deconvolution [26], shortcut, global convolution (GCN) and boundary refinement (BR) [27] are applied. The recovery of texture and structure is crucial since rain may deteriorate these contents. According to [26] multi-deconvolution is helpful for reconstructing the structural information while the GCN and BR can enhance the edge detail. Both the MSRN and the SSRN adopt the architecture of the CP as the backbone. However, the target output of MSRN is the residual difference while that of SSRN is the

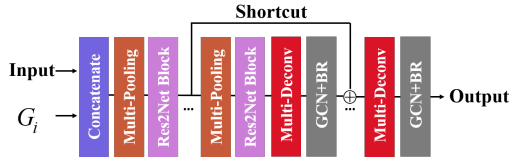


Figure 3: The architecture of the contourlet predictor (CP).

ground truth of SS, respectively. Besides, the filter depth is wider and more parallel kernels are connected in the SSRN because SS components require more complicated features and semantic information than MS components in the recovered process.

About the multi-level discriminator, we adopt the Res2Net [24] as the feature extractor for two input components (i.e., recovered image and multi-level SS components). Then, we concatenate these feature maps and leverage a fully connected layer to evaluate the quality of recovered result.

### 3 More Experimental Results

We present more results on real-world and synthetic datasets under the moderate rain and heavy rain scenarios. For the moderate rain, we adopt the Rain100H and Rain100L datasets proposed in [15], and Rain800 [16] for training and evaluation. Rain100H contains 1800 synthetic images for training and 100 synthetic images for testing. Rain100L contains 200 synthetic images for training and 100 synthetic images for testing. Rain800 contains 700 synthetic images for training and 100 synthetic images for testing. For the heavy rain, we adopt the heavy rain dataset proposed in [9], which contains 9000 synthetic images with rain streaks and the veiling effect. 7000 images are applied for training and 2000 images are applied for testing.

#### 3.1 Effectiveness of Contourlet Transformation.

To prove the effectiveness of the CT in rain scenarios, we apply several existing feature extraction techniques for comparison: vanilla convolution operation with  $3 \times 3$  kernel size (**VConv**); multi-scale convolution kernel (**Mconv**) [14]; Laplacian pyramid (**LP**) [8]; high-low frequency decomposition [9]; DWT [13]. We apply these rain streaks extraction techniques on the proposed network and evaluate the performance on real-world scenario. As shown in Figure 4, one can see that, the proposed method can achieve better performance on rain streak removal compared with other extraction techniques, and generate more decent result.

#### 3.2 Effectiveness of the Multi-level Discriminator and the Feedback Error Map.

To further prove the effectiveness of the proposed multi-level discriminator (**MD**) and the feedback error map (**FEM**), we shown a visual comparison in Figure 5. One can see that, with the proposed **MD**, the reconstructed subbands have better resolution and recovered quality. Moreover, the **MD** can suppress the residual rain streaks and achieve better rain removal results.

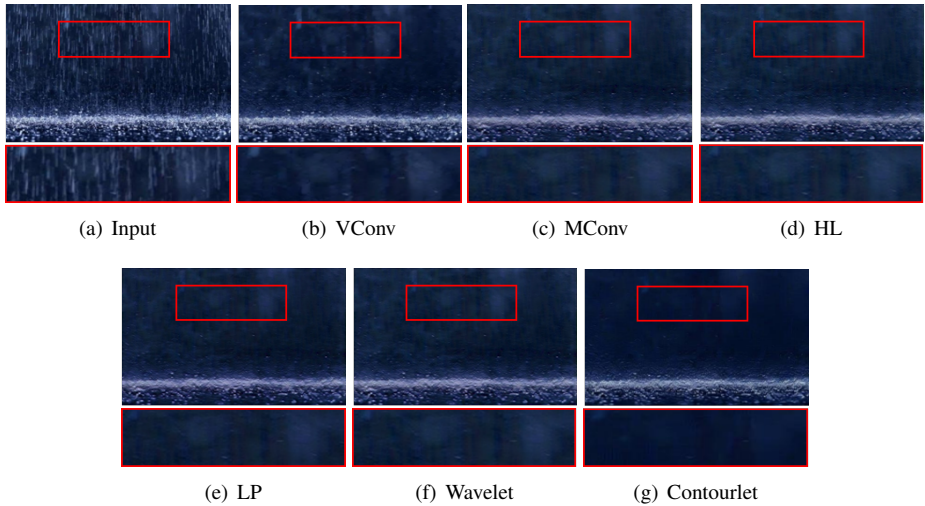


Figure 4: **Visual comparison of different decomposition methods on the real-world scene.**

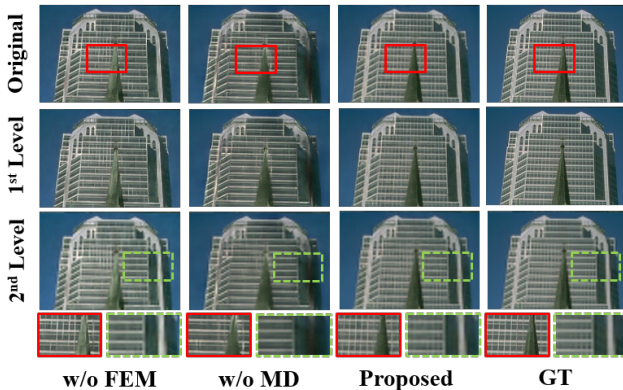


Figure 5: **Effectiveness of the multi-level subband discriminator and the feedback error map.** Note that 'Original' indicates the image without CT decomposition.

Table 1: **Computational resource evaluation for the ContourletNet and existing algorithms.** (Image size:  $640 \times 480$ )

Method	MSPFN [8]	RCDNet [10]	BRN [10]	HRGAN [9]	Ours
Time	0.58	1.50	0.41	0.30	0.11
Parameters	15,823,424	3,166,355	412,518	40,627,038	9,911,603
FLOPs	$2.84 \times 10^{12}$	$9.09 \times 10^{11}$	$9.23 \times 10^{11}$	$8.68 \times 10^{11}$	$8.94 \times 10^{10}$

### 3.3 More Recovered Results

**Heavy Rain Scenario.** In Figure 6 and Figure 7, we present more rain removal results on the synthetic heavy rain and real-world datasets for the heavy rain scenario. The results indicate



that the proposed method can achieve better reconstruction performance compared to other existing methods.

**Moderate Rain Scenario.** In this subsection, we present more rain removal results on the synthetic dataset to prove the effectiveness of the proposed method. The comparison is shown in Figure 8. From this comparison, one can notice that, the proposed method can remove both rain streaks and the veiling effect and retain more detail information qualitatively. Furthermore, the proposed method can achieve the best quantitative evaluation in terms of the PSNR, the SSIM, and the CIEDE 2000 compared to other methods. Moreover, we show more recovered results in the real-world dataset in Figure 9. Compared to other methods, the proposed method can remove rain streaks even more clearly and retain more detail information.

### 3.4 Computational Resource Evaluation

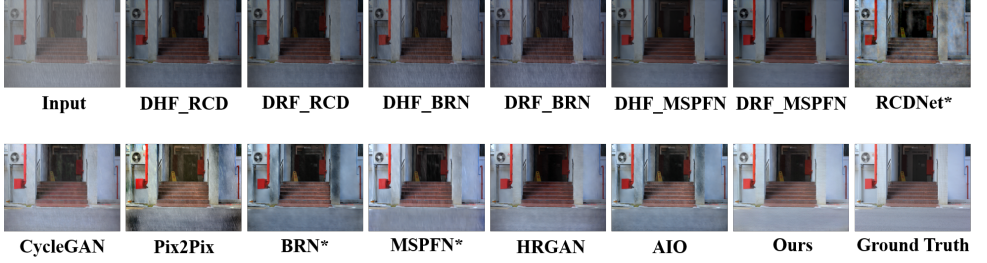
Table 1 presents the computational resource evaluation. It shows that the proposed method can achieve the favorable performance but with less FLOPs and time consumption.

### 3.5 Applications for Object Detection and Vehicle Plate Identification

Moreover, we show that the proposed method can benefit the performance of object detection and vehicle plate identification on the Google API service in Figure 10 and Figure 11, respectively. Note that the license plate number can be validly identified **ONLY** in the image recovered by the proposed method. Furthermore, the results recovered by our method can detect the object accurately compared to other methods. It is due to that the proposed method can effectively remove rain streaks and well reconstruct the detail.



(a)



(b)



(c)

Figure 6: Visual comparison on the heavy rain dataset.

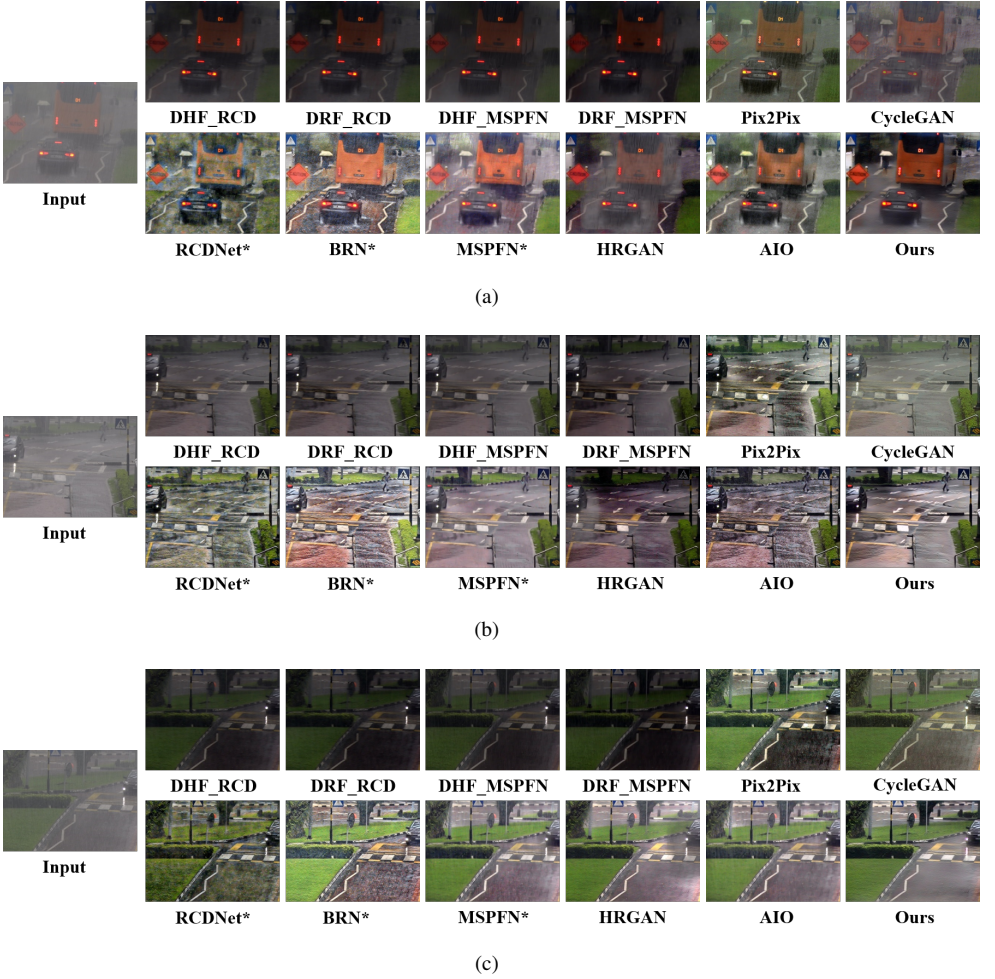


Figure 7: Comparison of heavy rain removal results on real-world scenarios with three existing strategies.



(a)



(b)



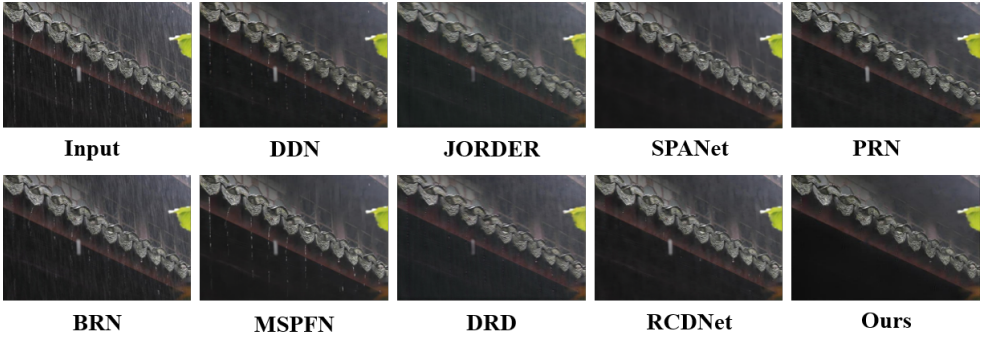
(c)

Figure 8: Comparison of moderate rain removal results on synthetic scenarios.





(a)



(b)



(c)

Figure 9: Comparison of moderate rain removal results on real-world scenarios.

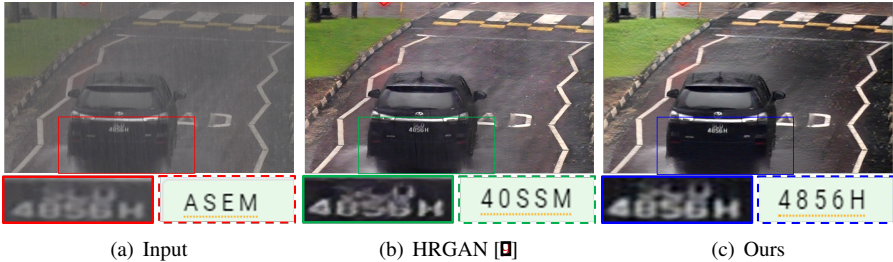


Figure 10: **Comparison of the license plate number identification results by the Google API after heavy rain removal.** One can see that, only the license plate in the image recovered by the proposed method can be identified correctly (i.e., 4856H), which indicates that the proposed method can potentially benefit the applications of high-level vision.



Figure 11: **Heavy Rain Removal** results for a realistic photograph by the HRGAN [9] and the proposed method with the corresponding **Object Detection** results by the Google Vision API.

## References

- [1] Roberto H Bamberger and Mark JT Smith. A filter bank for the directional decomposition of images: Theory and design. *IEEE transactions on signal processing*, 40(4): 882–893, 1992.
- [2] Peter Burt and Edward Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.
- [3] Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen 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*, pages 11681–11689, 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] Minh N Do and Martin Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 14(12): 2091–2106, 2005.
- [6] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight pyramid networks for image deraining. *IEEE transactions on neural networks and learning systems*, 31(6):1794–1807, 2019.
- [7] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE TPAMI*, 2020. doi: 10.1109/TPAMI.2019.2938758.
- [8] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020.
- [9] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1633–1642, 2019.
- [10] Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun. Large Kernel Matters-Improve Semantic Segmentation by Global Convolutional Network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361, 2017.
- [11] Dongwei Ren, Wei Shang, Pengfei Zhu, Qinghua Hu, Deyu Meng, and Wangmeng Zuo. Single image deraining using bilateral recurrent network. *IEEE Transactions on Image Processing*, 2020.
- [12] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016.



- [13] Bob L Sturm. Stéphane Mallat: A Wavelet Tour of Signal Processing, 2007.
- [14] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3103–3112, 2020.
- [15] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1357–1366, 2017.
- [16] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11):3943–3956, 2019.