

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

Wei-Ting Chen^{1,3*}

jimmy3505090@gmail.com

Cheng-Che Tsai^{1*}

r08943148@ntu.edu.tw

Hao-Yu Fang²

danielfang60609@gmail.com

I-Hsiang Chen²

f09921058@ntu.edu.tw

Jian-Jiun Ding²

jjding@ntu.edu.tw

Sy-Yen Kuo²

sykuo@ntu.edu.tw

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

² Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan

³ ASUS Intelligent Cloud Services
Asustek Computer Inc.
Taipei, Taiwan

Abstract

Images acquired from rainy scenes usually suffer from bad visibility which may damage the performance of computer vision applications. The rainy scenarios can be categorized into two classes: moderate rain and heavy rain scenes. Moderate rain scene mainly consists of rain streaks while heavy rain scene contains both rain streaks and the veiling effect (similar to haze). Although existing methods have achieved excellent performance on these two cases individually, it still lacks a general architecture to address both heavy rain and moderate rain scenarios effectively. In this paper, we construct a hierarchical multi-direction representation network by using the contourlet transform (CT) to address both moderate rain and heavy rain scenarios. The CT divides the image into the multi-direction subbands (MS) and the semantic subband (SS). First, the rain streak information is retrieved to the MS based on the multi-orientation property of the CT. Second, a hierarchical architecture is proposed to reconstruct the background information including damaged semantic information and the veiling effect in the SS. Last, the multi-level subband discriminator with the feedback error map is proposed. By this module, all subbands can be well optimized. This is the first architecture that can address both of the two scenarios effectively. The code is available in <https://github.com/cctakaet/ContourletNet-BMVC2021>.

* Equal Contribution

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

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

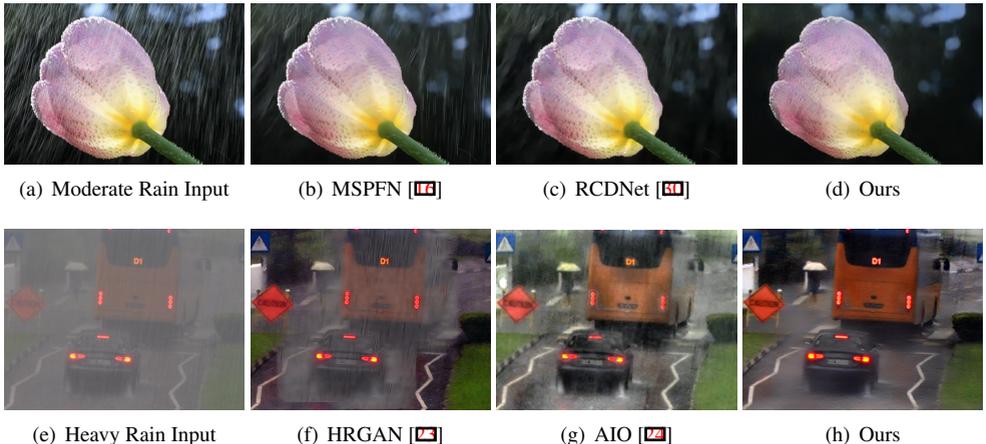


Figure 1: Visual comparison of the proposed method and state-of-the-art rain removal methods under moderate rain and heavy rain scenarios. It shows that the proposed method has generalized ability for addressing various rain scenarios.

1 Introduction

Rain streaks is an atmospheric phenomenon which usually leads to poor visibility. Similar to haze [9] and snow [5], it may degrade the performance of high-level vision applications, such as object detection and semantic segmentation. Rainy scenarios can be categorized into: (i) the moderate rain scene and (ii) the heavy rain scene (please refer Figure 1). The moderate rain scene mainly consists of rain streaks with irregular distribution and can be modeled as:

$$\mathbf{I} = \mathbf{J} + \sum_i^n \mathbf{S}_i, \quad (1)$$

where \mathbf{S}_i is the rain streak in the i^{th} layer, \mathbf{I} is the rain image, and \mathbf{J} is the clean image. Several algorithms have been proposed to address the moderate rain scenario [6, 9, 10, 16, 17, 19, 25, 30, 31, 32, 33, 35, 36]. Kang *et al.* [19] applied the image decomposition technique based on dictionary learning and morphology. Fu *et al.* [10] adopted discriminatively intrinsic characteristics, Li *et al.* [25] applied layer separation, and Wang *et al.* [30] proposed the SPANet to remove rain streaks. Jiang *et al.* [16] proposed the multi-scale progressive fusion strategy to achieve excellent results in moderate rain removal.

Unlike moderate rain scenarios, heavy rain scenes usually contain rain streaks and strong veiling effect which is similar to the haze. Its formulation can be expressed as [23]:

$$\mathbf{I} = \mathbf{T} \odot (\mathbf{J} + \sum_i^n \mathbf{S}_i) + (\mathbf{1} - \mathbf{T}) \odot \mathbf{A}, \quad (2)$$

where \mathbf{A} is the atmospheric light, \mathbf{T} is the transmission map, and \odot is pixel-wise multiplication. Due to multi-flux scattering and veiling effect, the background information under the heavy rain scenario is blurrier than that in the moderate rain case. Li *et al.* [23] proposed a two-stage optimization strategy termed the HRGAN which extracts \mathbf{T} , \mathbf{A} , and \mathbf{S} for physical model-based recovery, and applied adversarial learning to optimize the results.

Although existing rain removal methods can reconstruct desirable results for the moderate rain or the heavy rain scenario individually, there exists a limitation. These methods cannot address both types of rain scenes with the same architecture. In Figure 2, we present

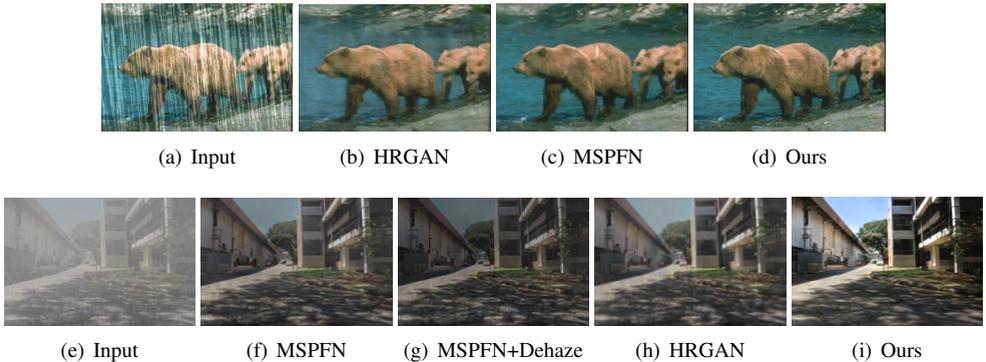


Figure 2: **Illustration of the limitation in existing methods.** (a) and (e): the inputs of moderate rain and heavy rain. (b)-(d): derained results by the heavy rain removal method (i.e., HRGAN [23]), MSPFN [16] and the proposed method. (f): derained results by moderate rain removal (i.e., MSPFN). (g): MSPFN + dehazed method [8]. (h): HRGAN and (i): the proposed method. Note that, for the fair comparison, both (b) and (f) are retrained with moderate rain and heavy rain (HR) datasets, respectively.

an example that applies a heavy rain removal method (the HRGAN [23]) to address moderate rain scenes and uses a moderate rain removal method (the MSPFN [16]) to handle the heavy rain scenes. We can observe that, for the former case, the recovered results tend to have color distortion problem because the veiling effect removal in the HRGAN may over-dehaze an moderate rain image. Moreover, the result tends to have residual rain streaks. For the latter case, the residual veiling effect may exist in the recovered results because the veiling effect removal is not considered by the moderate rain removal. Moreover, even if we combine the state-of-the-art veiling effect removal strategy (i.e., dehazing strategy) [8] with moderate rain removal, the recovered results are still limited because the heavy rain scenario is not a simple combination of rain streaks and the veiling effect. It suffers from the multi-flux scattering [23], resulting in blurry problem in the heavy rain scene.

Thus, in this paper, to adequately address both moderate rain and heavy rain scenarios with one architecture, we propose an effective recovered framework called the ContourletNet. First, the contourlet transform (CT) [7] is embedded in our network to generate several multi-direction subbands (MS) and one semantic subband (SS). The MS contains the information of rain streaks and detail such as edges and textures while the SS mainly consists of the contextual information and the veiling effect. With the CT, the information of rain streaks can be well retrieved because rain streaks are usually in different directions. We construct two sub-networks termed multi-direction subband recovery (MSR) and semantic subband recovery (SSR) to recover all subbands adequately. Second, to achieve better recovery of the SS, we investigate both heavy rain and moderate rain images. We find an interesting phenomenon: with the increase of the CT decomposition level, the difference between heavy rain/haze images and the corresponding ground truths in the SS is reduced. Based on this observation, we propose to integrate the hierarchical architecture and the CT in our network. Last, to improve the performance of the ContourletNet, the multi-level semantic subbands discriminator with the feedback error map is proposed. All semantic subbands and the recovered image are distinguished by the multi-level discriminator to enhance the contextual quality.

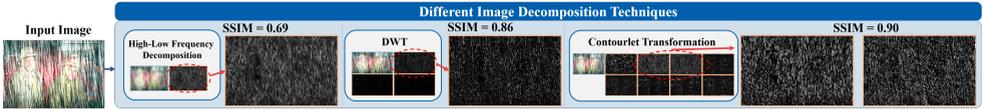


Figure 3: **Comparison of using different rain streak extraction methods.** One can see that the CT can achieve better rain streaks retrieval, especially for the bevelled directions. The rain streaks of input image mainly contain vertical and bevelled direction. Moreover, the value of SSIM indicates the structural similarity between the rain mask and each subband. Note that we add two subbands to calculate the SSIM for the CT.

Experiments show that our proposed method achieves superior performance than state-of-the-art approaches for both moderate rain and heavy rain scenarios. As far as we know, our method is the first work which can deal with moderate rain and heavy rain with one architecture effectively.

2 Proposed Method

2.1 Effective rain streak extraction

Rain streak extraction is essential for moderate rain and heavy rain scenarios. We first introduce the contourlet transform (CT) and illustrate why it is helpful for rain streak extraction.

Contourlet Transform. The CT [10] is an effective technique for geometric information analysis. It can achieve better representation in both locality and directionality. Its detail is presented in Supplementary Material. It mainly consists of two operations: the Laplacian pyramid (LP) [11] and directional filter banks (DFB) [10]. Given an image, first, the LP filter decomposes it into a 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. We term these subspaces multi-direction subbands (MSs). With the CT, the multi-resolution and multi-direction features can be extracted effectively.

Motivation. For both moderate rain and heavy rain removal, rain streak extraction plays a crucial role. Rain streaks are not uniform but have different directions and intensities. There are some existing strategies which can extract rain streaks, including high-low frequency decomposition (HL) [12], the discrete wavelet transform (DWT) [13], and learnable convolution kernels with a guidance [14]. For the DWT and HL, as shown in Figure 3, although they can extract some rain streaks information, the performance is still limited compared with the CT. The reason is as follows. The DWT only contains filter with two directions (i.e., vertical and horizontal), which leads to that the interpretation of bevelled rain streaks is limited. For HL, the information of rain streaks cannot be retrieved appropriately and stably because all high-frequency information (e.g., edges, details, and rain streaks) in various directions are included in a single subband. Moreover, it has no lossless inverse transformation, which limits the performance of reconstruction. For learnable convolution kernels, though rain information in different directions can be extracted by the well-trained filter from the CNN, achieving global optimization is challenging since the patterns of rain streak in real-world scenarios are too complicated to be covered and modeled comprehensively. Therefore, to address these limitations, we embed the CT into our network to extract rain streaks with better spatial locality and directionality. With this property, the effective representation of

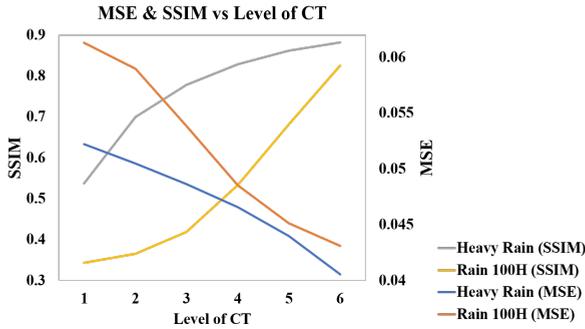


Figure 4: **Investigating the response of the CT level:** The MSE and the SSIM of the rain images (Rain 100H and heavy rain datasets) and the corresponding ground truths in the semantic subband (i.e., SS).

rain can be achieved by leveraging the empirical decomposition process based on the prior knowledge of rain streaks.

2.2 Hierarchical Decomposition for Rain Removal

The CT decomposes an image into multi-direction subbands (MSs) and the semantic subband (SS). Although appropriate recovery of the MS components can be achieved effectively by the residual strategy due to their sparsity [14, 20], reconstructing the SS component may be challenging since it generally contains complicated content such as residual rain streaks, the degraded background, and the veiling effect. To alleviate this issue, we propose a hierarchical architecture which only decomposes the SS component in each CT level. Specifically, we only recover the SS component at the bottom level and reconstruct several MS components in other levels. This idea is inspired by the investigation on the Rain 100H [34] and heavy rain datasets [23] which is shown in Figure 4. We observed that the difference between the rain image and the corresponding ground truth in the SS may decrease with the number of levels, which means that hierarchical decomposition can benefit the reconstruction of the SS.

For the moderate rain scene, with the hierarchical CT, most rain streaks are retrieved in MS components, which means that the SS component is closer to the ground truth. For the heavy rain scenario, rain streaks can be extracted by MS components, too. However, heavy rain images also contain the veiling effect, which mainly consists of the scene radiance \mathbf{J} , the atmospheric light \mathbf{A} , and the transmission \mathbf{T} [23]. \mathbf{A} is usually assumed as a constant. Although \mathbf{T} is a depth-related feature, it may become a constant-like matrix after an adequate iteration of decomposition because high-frequency components such as edge or texture are extracted. Thus, according to the formulation of veiling effect, the relation between the scene radiance \mathbf{J} and the heavy rain image \mathbf{I} can be approximated to a constant matrix mapping. Based on the above analysis, the difficulty of recovering the SS component can be reduced effectively with the proposed hierarchical CT. Therefore, the combination of the CT and the hierarchical architecture can both improve the ability of feature representation and achieve a better local optima.

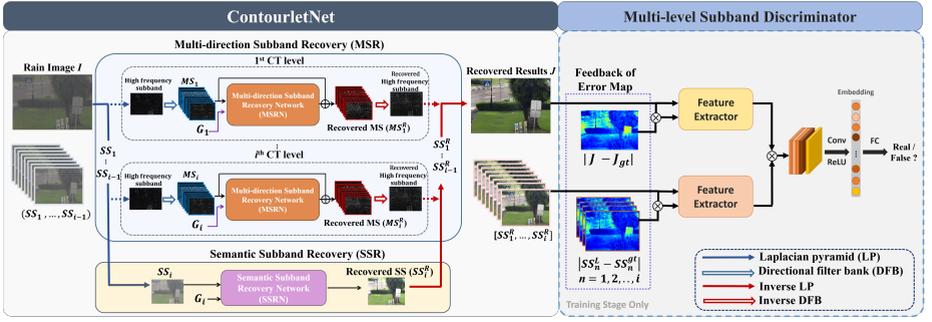


Figure 5: **The overview of the proposed rain removal network.** It consists of the ContourletNet and the multi-level subband discriminator. The ContourletNet contains two main modules: multi-direction subband recovery (MSR) and semantic subband recovery (SSR). The feature extractor in multi-level subband discriminator is based on Res2Net [13].

2.3 Network Architecture

The flowchart of the proposed network is shown in Figure 5. It consists of two subnetworks: (i) the ContourletNet and (ii) the multi-level subband discriminator. We leverage the GAN [13] to optimize the performance of the ContourletNet. First, the input image is fed into the ContourletNet to produce the clean image. Then, different from other learning-based architectures [15, 67], the recovered image and its recovered MS at different levels are distinguished by the discriminator and the feedback of the error map mechanism. The architecture is illustrated as follows.

ContourletNet. The ContourletNet can be divided into: i) multi-direction subband recovery (MSR) and ii) semantic subband recovery (SSR). First, the CT decomposes the input image into the MS and the SS. Then, the MS is reconstructed by the multi-direction subband recovery network (MSRN). The SS component at the bottom level is recovered by the semantic subband recovery network (SSRN) with pixel-to-pixel reconstruction. Two sub-networks are based on the proposed contourlet predictor (CP) architecture to recover the components. More details about CP, SSRN, and MSRN are presented in the Supplementary Material. Initially, the input is concatenated with the proposed aggregate contourlet component G_i :

$$G_i = \sigma(\mathbf{I}) \oplus \sigma(\mathbf{SS}_1) \oplus \cdots \oplus \sigma(\mathbf{SS}_{i-1}), \quad (3)$$

where G_i denotes the aggregate contourlet component at the i^{th} level, \oplus is the concatenate operation, \mathbf{I} is the rainy image, \mathbf{SS}_i is the SS component at level i , and σ is the multi-pooling architecture [4]. The idea in (3) is that, instead of down-sampling the input directly, the aggregate contourlet component enables the reconstruction network to acquire more information from the SS components in previous layers. The SS part contains semantic and contextual information which can benefit the reconstruction process. The use of multi-pooling is to prevent the information loss which usually happens in down-sampling. Then, the MSRN and SSRN will reconstruct the information in each subband based on the input and the aggregate contourlet component.

Multi-Level subbands Discriminator. To better optimize the recovered results, the multi-level subbands discriminator is proposed, as shown in the right side of Figure 5. In previous works, the final recovered result is generally fed into the discriminator directly. This may limit the performance because it is hard to optimize SS components at all levels based on the

final recovered result. To address this, we propose to distinguish both the recovered result and all recovered SS components jointly. Our idea is that the recovered SS is an important information to evaluate the performance of each level because it contains both multi-direction information and semantic information in the previous layers of the inverse CT. With the SS, the information in the previous layer can be optimized simultaneously. It can prevent the error from amplifying massively if the error is generated at the bottom layer. Moreover, to improve the performance of the discriminator, inspired by the attention mechanism, the feedback of the error map is proposed. That is, the input is multiplied by the error map, which is the L1-norm difference between the input and its ground truth. Then, the inputs are concatenated with their corresponding attentive component to the discriminator. It can make the affected regions get more attention from the discriminator.

2.4 Loss Function

In the proposed hierarchical contourlet-based rain removal network, three losses are applied: (i) contourlet loss, (ii) the perceptual loss, and (iii) the adversarial loss.

Contourlet Loss. The contourlet loss is defined as follows:

$$\mathcal{L}_C = \|(\mathbf{SS}_i^R - \mathbf{SS}_i^{gt})\|_2 + \sum_{m=1}^i \sqrt{\|(\mathbf{MS}_m^R - \mathbf{MS}_m^{gt})\|^2 + \varepsilon^2}, \quad (4)$$

where \mathbf{SS}_i^R and \mathbf{SS}_i^{gt} are the predicted semantic subband component and its corresponding ground truth at the i^{th} level. \mathbf{MS}_i^R and \mathbf{MS}_i^{gt} are the predicted multi-direction subbands and their corresponding ground truths, and ε is the slack value which can keep the values in multi-direction subbands non-zero to prevent the texture details from vanishing [22].

Adversarial Loss. It is defined as

$$\mathcal{L}_{Adv} = \min_G \max_D [E_{(\mathbf{J}_{SS} + \mathbf{J}) \sim p(\mathbf{J}_{SS} + \mathbf{J})} [\log D(\mathbf{J}_{SS} + \mathbf{J})] + E_{(\mathbf{I}_{SS} + \mathbf{I}) \sim p(\mathbf{I}_{SS} + \mathbf{I})} [\log(1 - D(G(\mathbf{I}_{SS} + \mathbf{I})))]], \quad (5)$$

where $\mathbf{J}_{SS} + \mathbf{J}$ and $\mathbf{I}_{SS} + \mathbf{I}$ are the sets of the clean images and rain images with their corresponding SS components, respectively. D is the discriminator, and G is the generator. The overall loss of the proposed network is

$$\mathcal{L}_{Overall} = \mathcal{L}_C + \lambda_1 \mathcal{L}_{Perceptual} + \lambda_2 \mathcal{L}_{Adv}, \quad (6)$$

where $\mathcal{L}_{Perceptual}$ is the perceptual loss [18] and we set the $\lambda_1 = 10^{-3}$ and $\lambda_2 = 10^{-4}$.

3 Experimental Result

3.1 Dataset and Implementation Detail

Several rain datasets are leveraged to validate the derained ability of the proposed method. For the moderate rain, we adopt the Rain100H and Rain100L datasets proposed in [34], and Rain800 [37] for training and evaluation. For the heavy rain, we adopt the heavy rain dataset proposed in [23]. The details of the datasets can be found in the Supplementary Material.

For the implementation detail, the initial learning rate is 10^{-4} and is multiplied by 0.75 after 10 epochs until the 250th epoch. The Adam optimizer [21] is applied, 16 directions and 4 levels are adopted in the CT. The proposed network is implemented on Nvidia RTX Titan GPU and trained with 1000 epochs. It takes thirty hours for training. Our network is trained based on PyTorch.

Table 1: **Quantitative evaluation for comparison with other existing methods on the heavy rain dataset.** The **DHF** and **DRF** denote dehaze first and derain first, respectively.

Metrics	RCDNet [40]		MSPFN [46]		BRN [47]		RCDNet*	MSPFN*	BRN*	CycleGAN [53]	Pix2Pix [45]	HRGAN [23]	AIO [24]	Ours
	DHF	DRF	DHF	DRF	DHF	DRF								
PSNR	16.42	15.81	16.69	16.03	16.76	16.68	23.15	23.92	22.38	21.62	22.43	24.78	24.71	25.45
SSIM	0.741	0.725	0.776	0.734	0.748	0.726	0.897	0.872	0.874	0.826	0.854	0.882	0.898	0.912
CIEDE 2000	13.32	14.01	12.77	13.69	12.63	13.36	5.14	4.52	6.38	7.23	6.26	7.40	4.89	4.07

Table 2: **Quantitative evaluation for comparison with other state-of-the-art methods on the existing moderate rain datasets.** (PSNR/SSIM/CIEDE2000)

Dataset	DDN [40]	JORDER [54]	SPANet [51]	PreNet [26]	BRN [47]	DRDNet [6]	RCDNet [30]	MSPFN [46]	Ours
Rain100L	32.170/0.927/3.73	36.830/0.972/2.08	35.280/0.966/2.21	37.100/0.977/1.52	38.160/0.982/1.11	36.950/0.978/1.18	39.990/0.986/0.48	36.400/0.973/1.27	40.040/0.988/0.43
Rain900	21.970/0.827/6.23	23.620/0.835/5.79	22.410/0.838/5.73	26.610/0.831/4.88	26.720/0.865/4.61	26.320/0.874/4.47	27.330/0.869/4.54	27.500/0.876/4.22	27.890/0.881/4.06
Rain100H	21.870/0.782/6.02	24.520/0.805/4.21	25.020/0.843/4.45	28.980/0.892/3.21	30.470/0.903/2.44	28.330/0.876/3.67	31.280/0.909/2.19	28.660/0.860/3.73	31.440/0.911/2.07

3.2 Comparison with State-of-the-art Methods

¹ We evaluate our method on two rain scenarios: moderate rain scenario and heavy rain scenario. For the moderate rain scenario, we apply several existing methods including DDN [40], JORDER [54], SPANet [51], PreNet [26], BRN [47], DRDNet [6], RCDNet [30], and MSPFN [46] for comparison. For fair evaluation, we ensure all models are trained with the same training set provided by the corresponding dataset.

For the heavy rain scenario, we adopt the HRGAN [23] and AIO [24] for comparison. Moreover, referring to the experiments in [23], we also compare our method with other heavy rain removal strategies: (i) moderate rain removal strategies combined with the dehazing strategy [8]; (ii) the moderate rain removal methods retrain with the heavy rain dataset; (iii) the image translation methods (the CycleGAN [53] and Pix2Pix [45]). For better identification, '*' denotes that the models retrain with the heavy rain dataset. Three metrics are applied for quantitative evaluation: the structural similarity (SSIM), the peak signal to noise ratio (PSNR), and the CIEDE 2000 color difference. A lower value of CIEDE2000 means less color distortion.

Analysis on Heavy Rain Scene. The quantitative comparisons are shown in Table 1. From Table 1, one can see that the proposed method outperforms other state-of-the-art heavy rain removal strategies in all metrics. Moreover, based on the visual comparison in Figures 6(a) and 6(b), one can see that, for existing heavy rain removal strategies, the recovered results tend to have residual rain streaks and color distortion. However, our method can solve these problems effectively and provide better visual quality compared to other methods.

Analysis on Moderate Rain Scene. Table 2 presents the results of quantitative evaluation on conventional moderate rain datasets. One can see that the proposed method can achieve the best performance for moderate rain removal compared with other existing methods in all datasets and all metrics. Figures 6(c) and 6(d) present the visual comparison on synthesized and real-world datasets. One can see that, compared with the proposed method, the results of existing methods may have more residual rain streaks (see the results presented in Figure 6(d)). These results show that the proposed method can achieve better rain removal and reconstruct the background effectively.

Based on the analysis above, the proposed method can achieve excellent performance on **both moderate rain and heavy rain scenarios**, which proves that the ContourletNet has the generalized ability on rain removal task.

¹Due to the limited space, more experimental results are reported in the supplementary material.

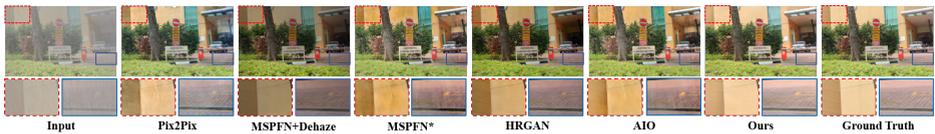
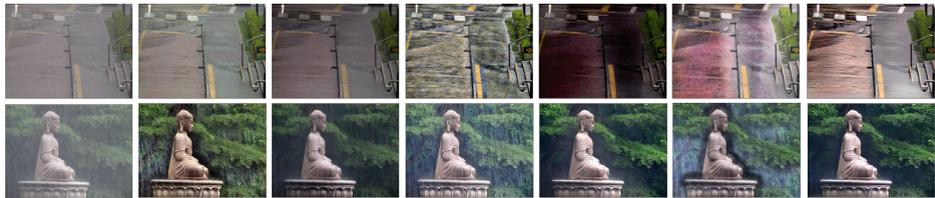
(a) Synthesized **heavy rain** images(b) Real-world **heavy rain** images(c) Synthesized **moderate rain** images.(d) Real-world **moderate rain** images

Figure 6: Visual comparison with other state-of-the-arts algorithms for some examples in heavy rain scenario and moderate rain scenarios.

3.3 Ablation Study

To verify the effectiveness of each of the proposed modules in this paper, five combinations are performed on the heavy rain and Rain 100H datasets: (1) the proposed ContourletNet with multi-level discriminator (**C**) (2) the ContourletNet with single-level discriminator (**C w/o MD**); (2) **C** without the hierarchical architecture (**C w/o H**); (3) **C** without the aggregate contourlet component (**C w/o G**); (4) **C** without the contourlet predictor (**C w/o CP**); (5) **C** without the feedback error map mechanism (**C w/o FEM**). Table 3(a) shows that the best performance can be achieved if all proposed techniques (the hierarchical architecture, the aggregate contourlet component, the contourlet predictor, the feedback of error map, and the multi-level contourlet components discriminator) are adopted.

To prove the effectiveness of the CT, we apply several existing feature extraction techniques instead of the CT in the proposed network for comparison: the vanilla convolution



(a) Input

(b) MSPFN



(c) RCDNet

(d) Ours

Figure 7: Failure case of the proposed method and state-of-the-art rain removal methods under night-time rain scene.

Table 3: Quantitative evaluation for ablation study.

(a) Effectiveness of the proposed modules

Dataset	C	C w/o MD	C w/o H	C w/o G	C w/o CP	C w/o FEM
Heavy	25.45/0.91	24.98/0.89	25.01/0.89	24.71/0.89	24.88/0.88	25.13/0.90
Rain 100H	31.44/0.91	30.01/0.89	29.88/0.89	30.23/0.90	29.97/0.89	30.41/0.90

(b) Effectiveness of Contourlet Transform

Dataset	VConv	MConv	HL	LP	DWT	Ours
Heavy	21.57/0.85	22.13/0.86	23.38/0.87	24.01/0.88	24.11/0.89	25.45/0.91
Rain 100H	26.21/0.84	27.14/0.85	29.31/0.87	29.51/0.88	29.63/0.88	31.44/0.91

(c) Level of CT decomposition versus the performance of recovery.

Dataset	Lv 1	Lv 2	Lv 3	Lv 4	Lv 5	Lv 6
Heavy	21.95/0.86	23.88/0.88	24.79/0.90	25.45/0.91	25.48/0.91	25.50/0.91
Rain100H	29.69/0.87	30.18/0.88	30.72/0.90	31.44/0.91	31.47/0.91	31.48/0.91

operation with a 3×3 kernel (**VConv**); multi-scale convolution kernel (**Mconv**) [28]; Laplacian pyramid (**LP**) [11]; HL [23]; DWT [29]. The results in Table 3(b) indicate that using the CT as the rain streak extractions can achieve the best result on deraining. Moreover, we present a visual comparison on the real-world scene in the Supplementary Material.

In Table 3(c), we discuss the effect of decomposition levels for the CT and the recovered performance. One can see that increasing the levels of decomposition can benefit the reconstruction quality generally. The more levels of decomposition, the better performance the proposed method may have.

4 Conclusion

In this paper, to mitigate the limitation in existing rain removal methods, a novel ContourletNet is proposed. First, to address the residual rain streaks in the recovered image, we leverage the CT to decompose the input image to several multi-direction subbands and a semantic subband at each level. Second, based on observing the semantic subbands of rainy images and their corresponding ground truths, the hierarchical recovery process is proposed. Last, the multi-level subband discriminator is proposed to suppress the error propagation in each subband. Extensive experiments prove the effectiveness of the proposed ContourletNet for both heavy and moderate rain scenarios. In the future work, we will address the limitation of the proposed method as shown in Figure 7. Specifically, the proposed method may fail for night-time rainy scene. There still exists residual rain streaks in the recovered images.

5 Acknowledgement

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

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/CVF 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. Jstar: 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] Sen Deng, Mingqiang Wei, Jun Wang, Yidan Feng, Luming Liang, Haoran Xie, Fu Lee Wang, and Meng Wang. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14560–14569, 2020.
- [7] 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.
- [8] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [9] Yingjun Du, Jun Xu, Qiang Qiu, Xiantong Zhen, and Lei Zhang. Variational image deraining. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2406–2415, 2020.
- [10] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3855–3863, 2017.
- [11] 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.

- [12] 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.
- [13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [14] Tiantong Guo, Hojjat Seyed Mousavi, Tiep Huu Vu, and Vishal Monga. Deep wavelet prediction for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 104–113, 2017.
- [15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [16] 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 de-raining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020.
- [17] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4057–4066, 2017.
- [18] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711, 2016.
- [19] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing*, 21(4):1742–1755, 2011.
- [20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.
- [21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [22] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(11):2599–2613, 2018.
- [23] 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.
- [24] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3175–3185, 2020.

- [25] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2736–2744, 2016.
- [26] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3937–3946, 2019.
- [27] 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.
- [28] 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.
- [29] Bob L Sturm. Stéphane Mallat: A Wavelet Tour of Signal Processing, 2007.
- [30] 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.
- [31] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12270–12279, 2019.
- [32] Wei Wei, Lixuan Yi, Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Should we encode rain streaks in video as deterministic or stochastic? In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2516–2525, 2017.
- [33] Hao-Hsiang Yang, Chao-Han Huck Yang, and Yu-Chiang Frank Wang. Wavelet channel attention module with a fusion network for single image deraining. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 883–887. IEEE, 2020.
- [34] 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.
- [35] Wenhan Yang, Jiaying Liu, Shuai Yang, and Zongming Guo. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning. *IEEE Transactions on Image Processing*, 28(6):2948–2961, 2019.
- [36] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2726–2736, 2020.
- [37] 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.

- [38] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.