Beyond Monocular Deraining: Stereo Image Deraining via Semantic Understanding

Kaihao Zhang[1], Wenhan Luo[2], Wenqi Ren[3]✉, Jingwen Wang[2]
Fang Zhao[4], Lin Ma[2], and Hongdong Li[1,5]

[1] Australian National University
[2] Tencent AI Lab
[3] Institute of Information Engineering, Chinese Academy of Sciences
[4] Inception Institute of Artificial Intelligence
[5] ACRV

Abstract. Rain is a common natural phenomenon. Taking images in the rain however often results in degraded quality of images, thus compromises the performance of many computer vision systems. Most existing de-rain algorithms use only one single input image and aim to recover a clean image. Few work has exploited stereo images. Moreover, even for single image based monocular deraining, many current methods fail to complete the task satisfactorily because they mostly rely on per pixel loss functions and ignore semantic information. In this paper, we present a Paired Rain Removal Network (PRRNet), which exploits both stereo images and semantic information. Specifically, we develop a Semantic-Aware Deraining Module (SADM) which solves both tasks of semantic segmentation and deraining of scenes, and a Semantic-Fusion Network (SFNet) and a View-Fusion Network (VFNet) which fuse semantic information and multi-view information respectively. We also propose new stereo based rainy datasets for benchmarking. Experiments on both monocular and the newly proposed stereo rainy datasets demonstrate that the proposed method achieves the state-of-the-art performance.

Keywords: Stereo deraining; semantic understanding; rethinking loop; view fusion; deep learning

1 Introduction

Stereo images processing has become an increasingly active research field in computer vision with the development of stereoscopic vision. Based on stereo images, many key technologies such as depth estimation [1–3], scene understanding [4–6] and stereo matching [7–9] have achieved a great success. As a common natural phenomenon, rain causes visual discomfort and degrades the quality of images, which can deteriorate the performance of many core models in outdoor vision-based systems. However, there are few studies for stereo deraining. In this paper, we address the problem of removing rain from stereo images.

In fact, stereo deraining has an intrinsic advantage over monocular deraining because the effects of identical rain streaks in corresponding pixels from stereo

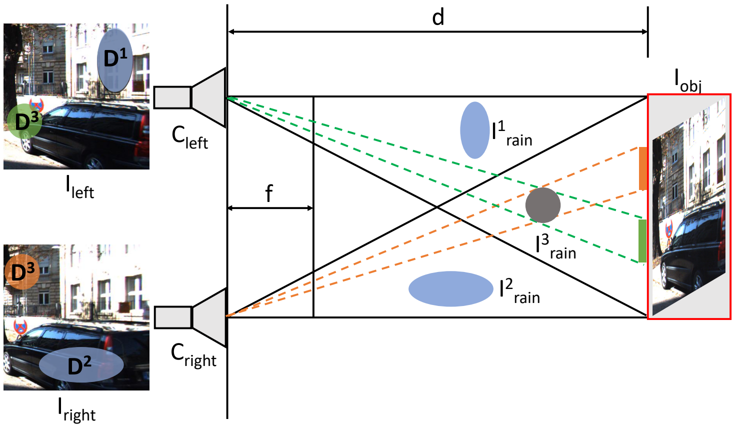


Fig. 1. The illustration of stereo cameras. One pair of images captured by stereo cameras. Same rain can cause different effects on images from two views.

images are different. As Fig. 1 shows, the mapping of object I_{obj} on stereo images can be represented as

$$I_{left} = I_{obj} * \frac{d}{f}, \qquad I_{right} = I_{obj}^{ref} * \frac{d}{f}, \tag{1}$$

where d and f are the distance between object and camera and the camera focal length, respectively. I_{obj}^{ref} is the reflection of I_{obj}. Assuming that the object I_{obj} is in the middle of two cameras, the lengths of identical objects, I_{obj}^{ref} and I_{obj}, on stereo views are the same. However, the effects of rain across stereo images are different. For example, the degraded regions by rain I_{rain}^1 on the two images can be denoted as

$$D_{left}^1 = I_{rain}^1 * \frac{d_{rain}}{f}, \qquad D_{right}^1 = 0. \tag{2}$$

I_{rain}^1 degrades the quality of the object on the left image but does not affect the visual comfort of the right view. d_{rain} is the distance between the camera and the raindrop. There is also rain influencing different regions on both stereo images like I_{rain}^3. The image in Fig. 1 shows the different effects of identical rain streaks on stereo views.

Moreover, the geometric cue and semantics provide important prior information, serving as a latent advantage for removing rain. Recently, most deep monocular deraining methods achieve a great success by reconstructing objects based on pixel-level objective functions like MSE. However, these methods ignore modeling the geometric structure of objects and understanding the semantic information of scenes, which in fact benefit deraining. Hu *et al.* [10] try to remove rain via depth estimation, but they also fail to understand the rainy scenes.

In this paper, we first propose a semantic-aware deraining module, *SADM*, which removes rain by leveraging scene understanding. Fig. 2 illustrates the concept of *SADM*. It contains two parts. The first part is an encoder which takes a rainy image as input and encodes it as semantic-aware features. Then the representations are fed into the second part, a conditional generator, to transform them into the deraining image and scene segmentation. Based on a multi-task shared learning mechanism and different input conditions, the single *SADM* is capable of jointly removing rain and understanding scenes. To further enhance the understanding of input images, a *Semantic-Rethinking Loop* is proposed to utilize the difference between the outputs of the conditional generators in different stages.

Based on *SADM*, we then present a stereo deraining model, *Paired Rain Removal Network (PRRNet)*, which consists of *SADM*, *Semantic-Fusion Network (SFNet)* and *View-Fusion Network (VFNet)*. *SADM* is utilized to learn the semantic information and reconstruct deraining images, while *SFNet* and *VFNet* are to fuse the semantic information with coarse deraining images, and obtain the final deraining images by fusing stereo views, respectively. Currently, there is no public large-scale stereo rainy datasets. In order to evaluate the performance of the proposed method and compare against the state-of-the-art methods, two large stereo rainy image datasets are thus constructed.

In summary, the contributions of this paper are three-fold:

- Firstly, a multi-task shared learning deraining model, *SADM*, is proposed to remove rain via scene understanding. This model not only considers pixel-level objective functions like previous methods, but also models the geometric structure and semantic information of input rainy images. Inside *SADM*, a novel *Semantic-Rethinking Loop* is employed to further strengthen the connection between scene understanding and image deraining.
- Secondly, we propose *PRRNet*, the first semantic-aware stereo deraining network. *PRRNet* fuses the semantic information and multi-view information via *SFNet* and *VFNet*, respectively, to obtain the final stereo deraining images.
- Thirdly, we synthesize two stereo rainy datasets for stereo deraining, which may be the largest datasets for stereo image deraining. Experiments on monocular and stereo rainy datasets show that the proposed *PRRNet* achieves the state-of-the-art performance on both monocular and stereo deraining.

2 Related Work

2.1 Single Image Deraining

Deraining from a single rainy image is a highly ill-posed task, whose mathematical formulation is expressed as

$$O = B + R\,, \tag{3}$$

where O, B and R are the observed rainy image, the latent clean image and the rain-streak component, respectively.

For traditional methods of recovering the clean deraining image B from the rainy version O, Kang *et al.* [11] first detect rain from the high/low frequency part of input images based on morphological component analysis and remove rain streaks in the high frequency layer via dictionary learning. Similarly, Huang *et al.* [12] and Zhu *et al.* [13] use sparse coding based methods to remove rain from a single image. Some works aim to remove rain based on low-rank representation [14, 15]. Chen *et al.* [14] generalize a low-rank model from matrix to tensor structure, which does not need the rain detection and dictionary learning stage. In addition, Li *et al.* [16] use a GMM trained on patches from natural images to model the background patch priors.

Recently, deep learning achieves significant success in low-level vision tasks such as image super-resolution [17, 18], deblurring [19, 20], dehazing [21, 22], which also include deraining [23–32]. These methods learn a mapping between input rainy images and their corresponding clean version using CNN/RNN based models. Some of them use an attention mechanism to pay attention to depth [10], heavy rain regions [33] or density [28]. However, to the best of our knowledge, there are few deep deraining works which try to remove rain via scene understanding [34].

2.2 Video Deraining

Video deraining is to obtain a clean video from an input rainy video. Compared with single image deraining, methods for video deraining can not only learn the spatial information, but also leverage temporal information in removing rain.

Traditional methods try to use prior-based methods to use the temporal context and motion information [35, 36]. Researches formulate rain streaks based on their intrinsic characteristics [37–41] or propose some learning-based methods to improve the performance of deraining models [42–46]. For example, Santhaseelan *et al.* [39] and Barnum *et al.* [47] extract phase congruence features and Fourier domain features, respectively, to remove rain streaks. Chen *et al.* [42] apply photo-metric and chromatic constraints to detect rain and utilize filters to remove rain in the pixel level.

Deep learning methods are also proposed for video deraining [48–51]. Chen *et al.* [50] propose a robust deep deraining model via applying super-pixel segmentation to decompose the scene into depth consistent unites. Liu *et al.* [48] depict rain streaks via a hybrid rain model, and then present a dynamic routing residue recurrent network via integrating the hybrid model and using motion information. Yang *et al.* [51] consider the additional degradation factors in real world and propose a two-stage recurrent network for video deraining. Their model is able to capture more reliable motion information at the first stage and keep the motion consistency between frames at the second stage. Although these methods use the information of multiple rainy images, all of them extract features from a sequence of monocular frames and ignore the stereo views.

2.3 Stereo Deraining

Stereo images provide more information from cross views and have thus been utilized to improve the performance of various computer vision tasks, including traditional problems [1, 4, 7] and novel tasks [52–55]. However, there are few methods that leverage the stereo images to remove rain so far. Yamashit *et al.* [56] remove the rain via utilizing disparities between stereo images to detect positions of noises and estimate true disparities of images regions hidden into rain. In order to obtain the deraining left-view images, Kim *et al.* [57] warp the spatially adjacent right-view frames and subtract warped frames from the original frames. However, these traditional methods do not consider the importance of semantic information. Meanwhile, the strong capability of learning features implied in deep neural networks is also ignored by them.

3 The Semantic-aware Deraining Module

The ultimate goal of our work is to recover the deraining images from their corresponding rainy versions. In order to improve the capability of our model, a semantic-aware deraining module is proposed to learn semantic features based on clean images, rainy images and semantic labels. In this section, we will first introduce the consolidation of different tasks in Sec. 3.1 and how to train the proposed module based on images and semantic-annotated images in Sec. 3.2. Then, a semantic-rethinking loop is discussed in Sec. 3.3 to further enhance our module and extract powerful features.

3.1 The Consolidation of Different Tasks

Currently, most deep deraining methods directly learn the transformation from rainy images to derained ones [23]. Inspired by [10], which proposes a depth-aware network to jointly learn depth estimation and image deraining via two different sub-networks. In this paper, an autoencoder architecture is employed to merge different tasks in the learning stage. Fig. 2 illustrates the architecture of the proposed module. Images are input into the encoder of the proposed module to extract semantic features F. Then the semantic features F combined with a task label T are fed into the following decoder architecture to obtain a prediction P corresponding to label T. Based on different task labels like *deraining* or *scene understanding*, different outputs will be obtained. The learning stage can be formulated as

$$P = D(E(I), T)\,, \tag{4}$$

where E and D are the encoder and decoder of *SADM*, respectively. I is the input image. T represents the label of different tasks. Based on the output of the encoder and T, different predictions will be derived.

The branch of image deraining can be denoted as

$$I_{de} = \sigma_{de}(P \mid T_{de})\,, \tag{5}$$

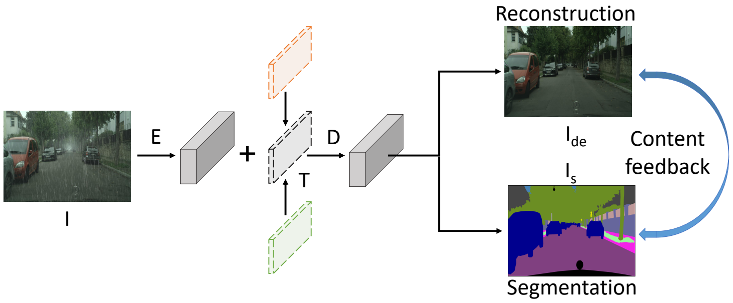


Fig. 2. The architecture of the proposed semantic-aware deraining module. Rainy images are fed into the encoder to extract features. Then the decoders generate deraining and segmentation results for different tasks.

where T_{de} corresponds to the label of deraining image. σ_{de} is the mapping function.

The branch of understanding scenes can be denoted as

$$I_{seg} = \sigma_{seg}(P \mid T_{seg}) \,, \tag{6}$$

where T_{seg} corresponds to the semantic segmentation label. σ_{seg} is a softmax function.

Based on the conditional architecture [58], the proposed *SADM* can jointly learn scene understanding and image deraining, which can extract more powerful semantic-aware features via sharing the information learned from different tasks, therefore being beneficial to multiple tasks.

3.2 Image Deraining and Scene Segmentation

Image Deraining. When T is set to T_{de}, the output of the proposed module is the deraining image. To learn the image deraining model, we compute the image reconstruction loss based on the MSE loss function:

$$\mathcal{L}_{de} = ||I_c - \sigma_{de}(D(E(I_{rainy}), T_{de}))||^2 \,, \tag{7}$$

where I_c is the clean image.

Scene Segmentation. Most existing deraining methods focus on pixel-level loss function and thus fail to model the geometric and semantic information. This makes it difficult for models to understand the input image and generate deraining results with favorable details. To address this problem, we remove rain from rainy images by leveraging semantic information. The learning process of scene understanding can be denoted as

$$\mathcal{L}_{seg} = \sigma_h(I_{seg}^{gt}, I_{seg}) \,, \tag{8}$$

where I_{seg} and I_{seg}^{gt} indicate the scene understanding of the model and ground truth labels from auxiliary training sets. σ_h is the cross-entropy loss function.

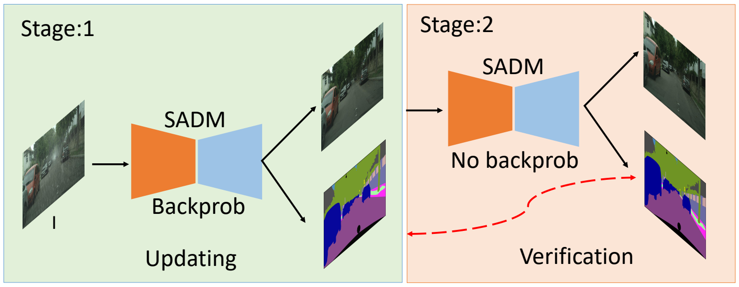


Fig. 3. The Semantic-rethinking Loop. During training, rainy images are fed into *SADM* to generate deraining and segmentation results in stage I. Then the deraining images are utilized to generate segmentation results again in stage II. Through comparing the two segmentation results from rainy and deraining images, *SADM* can better understand scenes and remove the undesired rain. *SADM*s in the two stages share the weights.

3.3 Semantic-rethinking Loop

Semantic information plays an important role in various tasks of computer vision [59–64]. In order to further enhance the semantic understanding of our model and help remove rain, a semantic-rethinking loop is proposed to refine the error-prone semantic understanding. Fig. 3 illustrates its scheme. It consists of an "updating" part and a "verification" part, whose core architecture is the semantic-aware deraining module, which has been illustrated in Fig. 2.

In the training stage, the "updating" part takes a rainy image as input, and then generates the deraining image and semantic segmentation. Loss functions introduced in above sections are calculated and then update the weights of layers in the semantic-aware deraining module. Then the deraining image obtained in the "updating" part is fed into the "verification" part to obtain new semantic segmentation. The semantic understanding can improve the performance of deraining, which will be demonstrated in the next section. However, rain increases the difficulty of scene understanding. Via comparing segmentation results in different parts and pushing them to be close, *SADM* can better understand scenes and thus better deraining. Both "updating" and "verification" parts employ the semantic-aware deraining module. The main difference between the "updating" and "verification" parts is that the weights in semantic-aware deraining module are updated in the "updating" part but fixed in the "verification" part. The semantic-rethinking loop provides the content feedback from the coarse-deraining image and improves the semantic understanding of *SADM*. In the testing stage, only the core semantic-aware deraining model is utilized to remove rain from images. The loss function can be noted as

$$\mathcal{L}_{con} = ||I_{seg}^{ver} - I_{seg}^{up}|| \,, \tag{9}$$

where I_{seg}^{ver} and I_{seg}^{up} are the semantic segmentation results from the "verification" and "updating" parts, respectively.

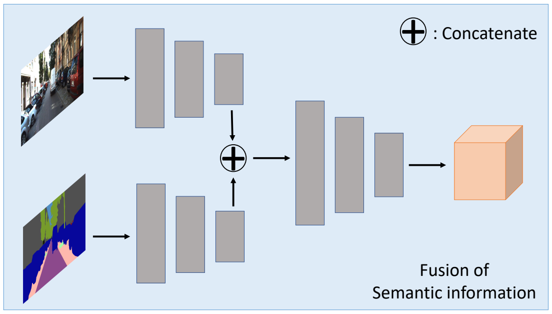


Fig. 4. The architecture of *SFNet*. The coarse deraining images and semantic segmentation results from *SADM* are fed into *SFNet* to generate features volume with semantic information.

4 The Paired Rain Removal Network

In order to remove rain from stereo images, we further present a *PRRNet* based on *SADM*. The overall of the proposed network will firstly be introduced in Sec. 4.1, and then two core sub-networks will be discussed in Sec. 4.2 and 4.3. Finally, the objective functions to train the proposed model will be presented in Sec. 4.4.

4.1 Network Architecture

PRRNet consists of three sub-networks, *i.e.*, *SADM*, Semantic-Fusion Net (*SFNet*) and View-Fusion Net (*VFNet*). *SADM* is introduced in Sec. 3 to jointly remove rain and understand semantic information. Semantic-Fusion Net is utilized to combine the semantic information with coarse deraining images, while View-Fusion Net is to combine information from different views to obtain final deraining images. Due to the above-mentioned stereo semantic-aware deraining module, the proposed *PRRNet* simultaneously considers cross views and semantic information to help remove rain from images.

4.2 SFNet

The architecture of *SFNet* is shown in Fig. 4. The input is semantic segmentation and coarse deraining images from *SADM*. Given that the semantic information can help remove rain, we first process them individually and concatenate them, and then forward them into the following layers, to generate feature volume, which is utilized for generating final deraining results.

4.3 VFNet

Fig. 5 illustrates the architecture of *VFNet*. The input is extracted fusion features from *SFNet*. The features extracted from the right view are helpful to remove

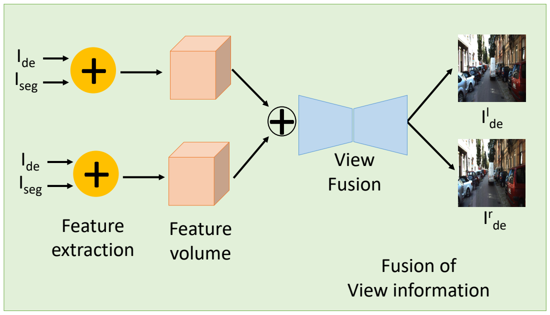


Fig. 5. The architecture of *VFNet*. Features volumes from stereo images are fused to generate final stereo deraining images.

the rain in the left-view image. Similarly, removing the rain from the right-view image also takes advantage of features captured from the left-view image. Through the *VFNet*, the final finer deraining stereo images are obtained. The loss function in this part can be denoted as

$$\mathcal{L}_{view} = ||I_{de}^{left} - I_{gt}^{left}|| + ||I_{de}^{right} - I_{gt}^{right}||, \tag{10}$$

where I_{de}^{left} and I_{de}^{right} are stereo deraining images from *VFNet*, respectively. I_{gt}^{left} and I_{gt}^{right} are the clean version of the stereo images.

4.4 Objective Functions

The loss function consists of two kinds of data terms, which are calculated based on semantic understanding and deraining reconstruction images. The final loss function can be written as

$$\mathcal{L}_f = \mathcal{L}_{de} + \lambda_1 \mathcal{L}_{seg} + \lambda_2 \mathcal{L}_{con} + \lambda_3 \mathcal{L}_{view}, \tag{11}$$

where $\mathcal{L}_{de}$ and $\mathcal{L}_{view}$ are utilized to remove the rain from rainy images, and $\mathcal{L}_{seg}$ and $\mathcal{L}_{con}$ push the model to understand scenes better, which are helpful for stereo deraining. λ_1, λ_2 and λ_3 are three parameters to balance different loss functions, which are set as 1.0, 0.2 and 1.0, respectively.

5 Experiments

5.1 Datasets

RainKITTI2012 dataset. To the best of our knowledge, there are no benchmark datasets that provide stereo rainy images and their corresponding ground-truth clean version. In this paper, we first use Photoshop to create a synthetic RainKITTI2012 dataset based on the public KITTI stereo 2012 dataset [65]. The training set contains $4,062$ image pairs from various scenarios, and the testing set contains $4,085$ image pairs. The size of images is 1242×375.

RainKITTI2015 dataset. The KITTI2015 dataset is another set from the KITTI stereo 2015 dataset [65]. Therefore, we also synthesize a RainKITTI2015 dataset, whose training set and testing set contain $4,200$ and $4,189$ pairs of images, respectively.

Cityscapes dataset. Cityscapes dataset is utilized as the semantic segmentation data to train *PRRNet*. This dataset contains various urban street scenes and provides images with pixel-wise segmentation labels. It includes $2,975$ images and their corresponding ground truth semantic labels.

RainCityscapes dataset. This dataset is built by Hu *et al.* [10] based on the Cityscapes dataset [66]. The training set contains $9,432$ rainy images and the corresponding clean images and depth labels. For evaluation, the testing set contains $1,188$ images. We use this dataset to evaluate the performance of monocular deraining.

5.2 Implementation Details

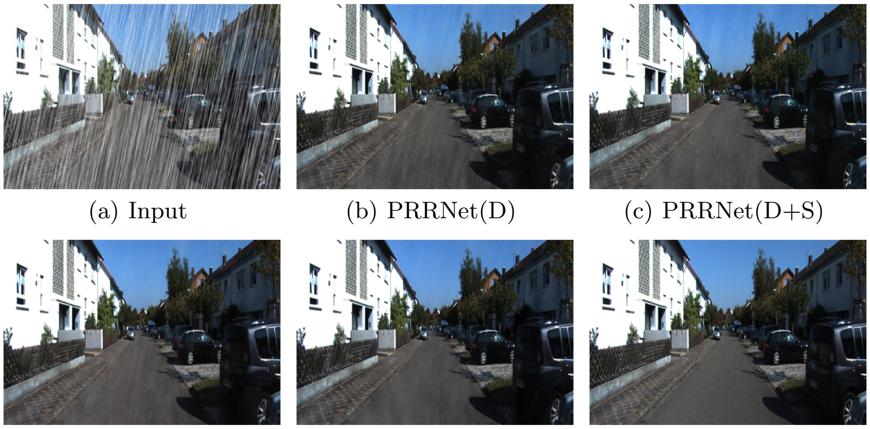
SADM is an encoder-decoder architecture. The encoder network consists of 13 CNN layers, which is initialized by a VGG16 network pre-trained for object classification. The decoder also has 13 CNN layers. *SFNet* contains three CNN layers ($32 \times 3 \times 3$) which are utilized to fuse the semantic information. *VFNet* contains five ResBlocks [67] to generate final deraining results. Each ResBlock consists of three CNN layers of $64 \times 3 \times 3$ kernels and two ReLU activation layers. The proposed *PRRNet* is trained with Pytorch library. The base learning rate is set to 10^{-4} and then declined to 10^{-5}. The model is updated with the batch size of 2 during the training stage. The branches of deraining and segmentation in *SADM* are optimized based on the data from RainKITTI2012/2015 and Cityscapes, respectively.

5.3 Ablation Study

The proposed *PRRNet* takes advantage of semantic information to remove rain from images. In order to show the effectiveness of semantic information, we compare the performance of our model with that which is trained without semantic information. Another advantage of *PRRNet* is that it fuses the varying information in corresponding pixels across two stereo views to remove rain. Therefore, we also compare models trained on monocular and stereo images. Table 1 and Fig. 6 show the quantitative and qualitative comparison results. *PRRNet(D)* is the model trained on monocular images with the single deraining task. *PRRNet(D+S)* is the one trained on monocular images with both deraining and segmentation tasks. *PRRNet(D+S+L)* is the model trained on monocular images with the above two tasks plus the semantic-rethinking loop. *PRRNet(stereo)* is our full model trained based on stereo images.

Table 1. *Ablation study on the RainKITTI2012 dataset.*

Methods	PSNR	SSIM
PRRNet (D)	30.71	0.923
PRRNet (D+S)	31.56	0.928
PRRNet (D+S+L)	31.89	0.930
PRRNet (stereo)	33.01	0.936


(a) Input (b) PRRNet(D) (c) PRRNet(D+S)

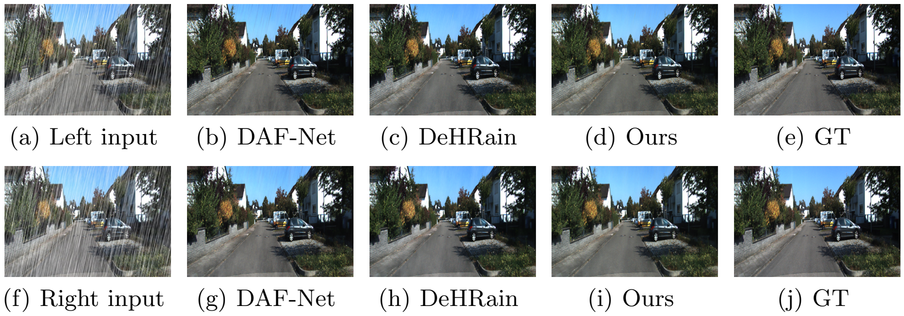
(d) PRRNet(D+S+L) (e) PRRNet(stereo) (f) Ground truth

Fig. 6. Deraining evaluation of different baseline models on RainKITTI2012.

The results in Table 1 suggest that, the plain *PRRNet(D)* accomplishes the task fairly well. Additionally considering the semantic segmentation task, *PRRNet(D+S)* improves the performance. With the semantic-rethinking loop, the results are further improved by *PRRNet(D+S+L)*. However, the improvement is not as significant as that from *PRRNet(D+S+L)* to *PRRNet(stereo)* in the stereo case. This is also verified by the qualitative results in Fig. 6. Additional components incrementally improve the visibility of the input image, and the image generated by *PRRNet(stereo)* is the closest to the ground truth.

5.4 Stereo Deraining

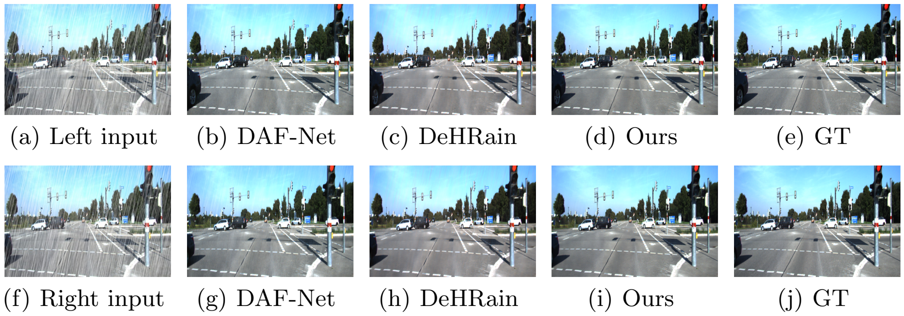
We quantitatively and qualitatively compare our *PRRNet* with current state-of-the-art methods, which include DDN [27], DID-MDN [28], DAF-Net [10] and DeHRain [33]. Table 2 and Table 3 show the quantitative results on our synthesized RainKITTI2012 and RainKITTI2015 datasets, respectively. In both tables, our monocular version, *PRRNet(monocular)*, outperforms the existing state-of-the-art methods, with remarkable gain. The model *PRRNet(stereo)* achieves the best performance with additional improvement. This demonstrates the superiority of stereo deraining over monocular deraining.



(a) Left input (b) DAF-Net (c) DeHRain (d) Ours (e) GT

(f) Right input (g) DAF-Net (h) DeHRain (i) Ours (j) GT

Fig. 7. Qualitative evaluation of current SOTA models on RainKITTI2012.



(a) Left input (b) DAF-Net (c) DeHRain (d) Ours (e) GT

(f) Right input (g) DAF-Net (h) DeHRain (i) Ours (j) GT

Fig. 8. Qualitative evaluation of current SOTA models on RainKITTI2015.

Figs. 7 and 8 compare the qualitative performances between our method *PRRNet(stereo)* and various state-of-the-art methods. The results produced by our method exhibit the smallest portion of artifacts, by referring to the ground truths.

5.5 Monocular Deraining

The proposed *PRRNet* is not only able to remove rain from stereo images, but also has the advantage of removing rain from a single image with its monocular version. In this section, we also evaluate it on the monocular dataset RainCityscapes. We compare the *PRRNet*'s monocular version, *PRRNet(monocular)*, with the state-of-the-art methods, including DID-MDN [28], RESCAN [29], JOB [13], GMMLP [16], DSC [69], DCPDN [68], and DAF-Net [10], from both quantitative and qualitative aspects.

The quantitative results on the RainCityscapes dataset are shown in Table 4. DID-MDN [28] and DCPDN [68] perform well and DAF-Net [10] outperforms these two methods. Our monocular version *PRRNet(monocular)* achieves the best performance compared with all the compared methods on this task, revealing the effectiveness of taking semantic segmentation into consideration and

Table 2. *Quantitative evaluation on the RainKITTI2012 dataset.*

Methods	PSNR	SSIM
DDN [26]	29.43	0.904
DID-MDN [28]	29.14	0.901
DAF-Net [10]	30.44	0.914
DeHRain [33]	31.02	0.923
PRRNet(monocular)	**31.89**	**0.930**
PRRNet(stereo)	**33.01**	**0.936**

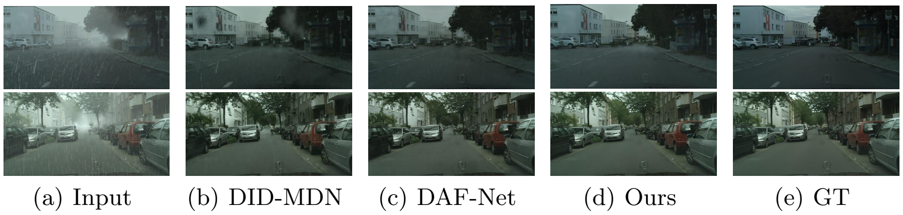
Table 3. *Quantitative evaluation on the RainKITTI2015 dataset.*

Methods	PSNR	SSIM
DDN [26]	29.23	0.906
DID-MDN [28]	28.97	0.899
DAF-Net [10]	30.17	0.915
DeHRain [33]	30.84	0.921
PRRNet(monocular)	**31.64**	**0.932**
PRRNet(stereo)	**32.58**	**0.937**

the semantic-rethinking loop. Fig. 9 compares its qualitative performance with different methods. The results show that the monocular version of our *PRRNet* also achieves the best performance in terms of monocular image deraining.

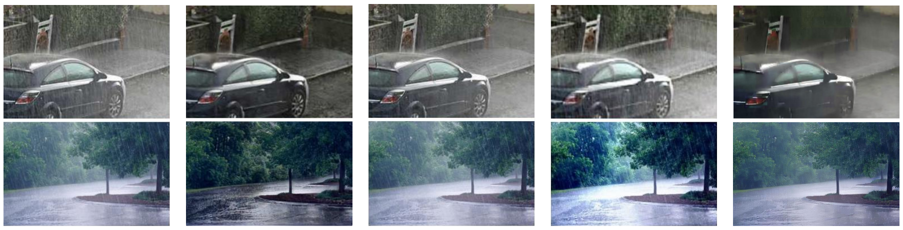
5.6 Evaluation on Real-world Images

To further verify the effectiveness of our method, we show its performance of deraining on the real world rainy images. Fig. 10 shows the qualitative results on two exemplar images from the Internet. Compared to other competing methods, the proposed method achieves better performance via understanding the scene structure. For example, DAF-Net seems to generate well-derained images, but the produced derained images suffer from color distortion (*e.g.*, the colors turn dark in the results). RESCAN and RESCAN+DCPDN perform worse than our method in removing rain.



(a) Input (b) DID-MDN (c) DAF-Net (d) Ours (e) GT

Fig. 9. Qualitative evaluation of current state-of-the-art models on the RainCityscapes dataset.



(a) Input (b) DAF-Net (c) RESCAN (d) RESCAN + DCPDN (e) Ours

Fig. 10. Qualitative evaluation on real rainy images. From left to right are the input images, DAF-Net [10], RESCAN [29], RESCAN + DCPDN [68] and ours, respectively.

Table 4. *Quantitative evaluation of current state-of-the-art models on the RainCityscapes dataset.*

Methods	PSNR	SSIM
DID-MDN [28]	28.43	0.9349
RESCAN [29]	24.49	0.8852
JOB [13]	15.10	0.7592
GMMLP [16]	17.80	0.8169
DSC [69]	16.25	0.7746
DCPDN [68]	28.52	0.9277
DAF-Net [10]	30.06	0.9530
***PRRNet*(monocular)**	**31.44**	**0.9688**

6 Conclusion

In this paper, we present *PRRNet*, the first stereo semantic-aware deraining network, for stereo image deraining. Different from previous methods which only learn from pixel-level loss functions or monocular information, the proposed model advances image deraining by leveraging semantic information extracted by a semantic-aware deraining model, as well as visual deviation between two views fused by two Fusion Nets, *i.e.*, *SFNet* and *VFNet*. We also synthesize two stereo deraining datasets to evaluate different deraining methods. The experimental results show that our proposed *PRRNet* outperforms the state-of-the-art methods on both monocular and stereo image deraining.

Acknowledgment

This work is funded in part by the ARC Centre of Excellence for Robotics Vision (CE140100016), ARC-Discovery (DP 190102261) and ARC-LIEF (190100080) grants. The authors gratefully acknowledge NVIDIA for GPU gift.

References

1. Godard, C., Mac Aodha, O., Brostow, G.J.: Unsupervised monocular depth estimation with left-right consistency. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
2. Liu, F., Shen, C., Lin, G.: Deep convolutional neural fields for depth estimation from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015)
3. Riegler, G., Liao, Y., Donne, S., Koltun, V., Geiger, A.: Connecting the dots: Learning representations for active monocular depth estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019)
4. Eslami, S.A., Heess, N., Weber, T., Tassa, Y., Szepesvari, D., Hinton, G.E., et al.: Attend, infer, repeat: Fast scene understanding with generative models. In: Advances in Neural Information Processing Systems (NeurIPS). (2016)
5. Shao, J., Kang, K., Change Loy, C., Wang, X.: Deeply learned attributes for crowded scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015)
6. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
7. Luo, W., Schwing, A.G., Urtasun, R.: Efficient deep learning for stereo matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016)
8. Chang, J.R., Chen, Y.S.: Pyramid stereo matching network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
9. Pang, J., Sun, W., Ren, J.S., Yang, C., Yan, Q.: Cascade residual learning: A two-stage convolutional neural network for stereo matching. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2017)
10. Hu, X., Fu, C.W., Zhu, L., Heng, P.A.: Depth-attentional features for single-image rain removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019)
11. Kang, L.W., Lin, C.W., Fu, Y.H.: Automatic single-image-based rain streaks removal via image decomposition. IEEE Transactions on Image Processing (TIP) (2011)
12. Huang, D.A., Kang, L.W., Wang, Y.C.F., Lin, C.W.: Self-learning based image decomposition with applications to single image denoising. IEEE Transactions on Multimedia (TMM) (2013)
13. Zhu, L., Fu, C.W., Lischinski, D., Heng, P.A.: Joint bi-layer optimization for single-image rain streak removal. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2017)
14. Chen, Y.L., Hsu, C.T.: A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2013)
15. Zhang, H., Patel, V.M.: Convolutional sparse and low-rank coding-based rain streak removal. In: IEEE Winter Conference on Applications of Computer Vision (WACV). (2017)
16. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016)

17. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
18. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision (ECCV). (2016)
19. Zhang, K., Luo, W., Zhong, Y., Ma, L., Liu, W., Li, H.: Adversarial spatio-temporal learning for video deblurring. IEEE Transactions on Image Processing (TIP) (2018)
20. Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., Li, H.: Deblurring by realistic blurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2020)
21. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision (ECCV). (2016)
22. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2017)
23. Li, S., Araujo, I.B., Ren, W., Wang, Z., Tokuda, E.K., Junior, R.H., Cesar-Junior, R., Zhang, J., Guo, X., Cao, X.: Single image deraining: A comprehensive benchmark analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019)
24. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2019)
25. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: A deep network architecture for single-image rain removal. IEEE Transactions on Image Processing (TIP) (2017)
26. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
27. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
28. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
29. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: European Conference on Computer Vision (ECCV). (2018)
30. Eigen, D., Krishnan, D., Fergus, R.: Restoring an image taken through a window covered with dirt or rain. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2013)
31. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
32. Zheng, Y., Yu, X., Liu, M., Zhang, S.: Residual multiscale based single image deraining. In: British Machine Vision Conference (BMVC). (2019)
33. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019)

34. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015)
35. Garg, K., Nayar, S.K.: Detection and removal of rain from videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2004)
36. Garg, K., Nayar, S.K.: Photorealistic rendering of rain streaks. In: ACM Transactions on Graphics (TOG). (2006)
37. Zhang, X., Li, H., Qi, Y., Leow, W.K., Ng, T.K.: Rain removal in video by combining temporal and chromatic properties. In: IEEE International Conference on Multimedia and Expo (ICME). (2006)
38. Liu, P., Xu, J., Liu, J., Tang, X.: Pixel based temporal analysis using chromatic property for removing rain from videos. Computer and Information Science (2009)
39. Santhaseelan, V., Asari, V.K.: Utilizing local phase information to remove rain from video. International Journal of Computer Vision (IJCV) (2015)
40. Brewer, N., Liu, N.: Using the shape characteristics of rain to identify and remove rain from video. In: Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR). (2008)
41. Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., Wang, Y.: 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 (CVPR). (2017)
42. Chen, J., Chau, L.P.: A rain pixel recovery algorithm for videos with highly dynamic scenes. IEEE Transactions on Image Processing (TIP) (2013)
43. Tripathi, A., Mukhopadhyay, S.: Video post processing: low-latency spatiotemporal approach for detection and removal of rain. IET Image Processing (2012)
44. Kim, J.H., Sim, J.Y., Kim, C.S.: Video deraining and desnowing using temporal correlation and low-rank matrix completion. IEEE Transactions on Image Processing (TIP) (2015)
45. Wei, W., Yi, L., Xie, Q., Zhao, Q., Meng, D., Xu, Z.: Should we encode rain streaks in video as deterministic or stochastic? In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2017)
46. Ren, W., Tian, J., Han, Z., Chan, A., Tang, Y.: Video desnowing and deraining based on matrix decomposition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
47. Barnum, P.C., Narasimhan, S., Kanade, T.: Analysis of rain and snow in frequency space. International Journal of Computer Vision (IJCV) (2010)
48. Liu, J., Yang, W., Yang, S., Guo, Z.: D3r-net: Dynamic routing residue recurrent network for video rain removal. IEEE Transactions on Image Processing (TIP) (2018)
49. Liu, J., Yang, W., Yang, S., Guo, Z.: Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
50. Chen, J., Tan, C.H., Hou, J., Chau, L.P., Li, H.: Robust video content alignment and compensation for rain removal in a cnn framework. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
51. Yang, W., Liu, J., Feng, J.: Frame-consistent recurrent video deraining with dual-level flow. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019)

52. Jeon, D.S., Baek, S.H., Choi, I., Kim, M.H.: Enhancing the spatial resolution of stereo images using a parallax prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
53. Li, B., Lin, C.W., Shi, B., Huang, T., Gao, W., Jay Kuo, C.C.: Depth-aware stereo video retargeting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
54. Chen, D., Yuan, L., Liao, J., Yu, N., Hua, G.: Stereoscopic neural style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
55. Zhou, S., Zhang, J., Zuo, W., Xie, H., Pan, J., Ren, J.S.: Davanet: Stereo deblurring with view aggregation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2019)
56. Tanaka, Y., Yamashita, A., Kaneko, T., Miura, K.T.: Removal of adherent waterdrops from images acquired with a stereo camera system. IEICE Transactions on Information and Systems (IEICE TIS) (2006)
57. Kim, J.H., Sim, J.Y., Kim, C.S.: Stereo video deraining and desnowing based on spatiotemporal frame warping. In: The IEEE International Conference on Image Processing (ICIP). (2014)
58. Zhao, F., Zhao, J., Yan, S., Feng, J.: Dynamic conditional networks for few-shot learning. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 19–35
59. Shen, Z., Lai, W.S., Xu, T., Kautz, J., Yang, M.H.: Deep semantic face deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 8260–8269
60. Shen, Z., Lai, W.S., Xu, T., Kautz, J., Yang, M.H.: Exploiting semantics for face image deblurring. International Journal of Computer Vision (2020)
61. Li, D., Rodriguez, C., Yu, X., Li, H.: Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In: The IEEE Winter Conference on Applications of Computer Vision. (2020) 1459–1469
62. Li, D., Yu, X., Xu, C., Petersson, L., Li, H.: Transferring cross-domain knowledge for video sign language recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 6205–6214
63. Zhang, J., Fan, D.P., Dai, Y., Anwar, S., Saleh, F.S., Zhang, T., Barnes, N.: Uc-net: uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 8582–8591
64. Zhang, J., Yu, X., Li, A., Song, P., Liu, B., Dai, Y.: Weakly-supervised salient object detection via scribble annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 12546–12555
65. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research (IJRR) (2013)
66. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016)
67. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). (2016)
68. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)

69. Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparse coding. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). (2015)