International Journal of Computer Vision (2022) 130:1754–1769
https://doi.org/10.1007/s11263-022-01620-w

Beyond Monocular Deraining: Parallel Stereo Deraining Network Via Semantic Prior

Kaihao Zhang[1] · Wenhan Luo[2] · Yanjiang Yu[3] · Wenqi Ren[2] · Fang Zhao[4] · Changsheng Li[3] · Lin Ma[5] · Wei Liu[6] · Hongdong Li[1]

Received: 8 May 2021 / Accepted: 15 April 2022 / Published online: 17 May 2022

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. In addition, we also introduce an Enhanced Paired Rain Removal Network (EPRRNet) which exploits semantic prior to remove rain streaks from stereo images. We first use a coarse deraining network to reduce the rain streaks on the input images, and then adopt a pre-trained semantic segmentation network to extract semantic features from the coarse derained image. Finally, a parallel stereo deraining network fuses semantic and multi-view information to restore finer results. 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. https://github.com/HDCVLab/Stereo-Image-Deraining.

Keywords Stereo image deraining · Parallel stereo network · View fusion · Deep learning

1 Introduction

Stereo image 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 (Godard et al., 2017; Liu et al., 2015; Riegler et al., 2019) , scene understanding (Eslami et al., 2016; Shao et al., 2015; Zhao et al., 2017) and

Communicated by Andreas Geiger.

✉ Wenhan Luo
whluo.china@gmail.com

Kaihao Zhang
kaihao.zhang@anu.edu.au

Yanjiang Yu
yuyanjiang87@gmail.com

Wenqi Ren
rwq.renwenqi@gmail.com

Fang Zhao
fang.zhao@inceptioniai.org

Changsheng Li
lcs@bit.edu.cn

Lin Ma
forest.linma@gmail.com

Wei Liu
wl2223@columbia.edu

Hongdong Li
hongdong.li@anu.edu.au

[1] Australian National University, Canberra, Australia

[2] Sun Yat-sen University, Guangzhou, China

[3] Beijing Institute of Technology, Beijing, China

[4] Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

[5] Meituan Group, Beijing, China

[6] Tencent, Shenzhen, China

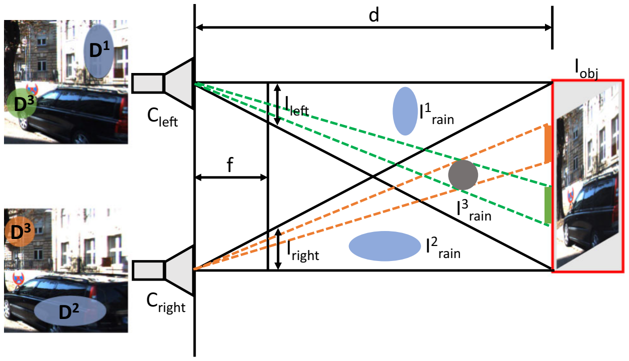


Fig. 1 The illustration of stereo cameras. One pair of images I_{left} and I_{right} is captured by stereo cameras C_{left} and C_{right}. I_{obj} is an object and I_{obj}^{ref} is the reflection of I_{obj}. d is distance between the object and the camera. f is the camera focal length. The same rain I_{rain}^{3} can cause different effects on images from two views

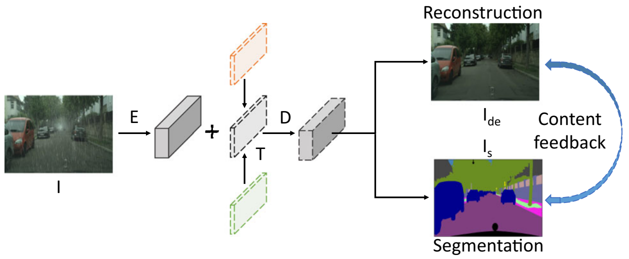


Fig. 2 The architecture of the proposed semantic-aware deraining module. Rainy images are fed into the encoder E to extract features. T represents the task labels, which can be deraining labels or scene labels. Then the decoders D generate deraining and segmentation results for different tasks

stereo matching (Luo et al., 2016; Chang & Chen, 2018; Pang et al., 2017) 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 images are different. As Fig. 1 shows, the degraded regions by rain I_{rain}^{1} on the two images are different. I_{rain}^{1} degrades the quality of the object on the left image but does not affect the visual comfort of the right view. There is also rain influencing different regions on both stereo images like I_{rain}^{3}. However, the degraded regions are still different in images from different views. This will provide additional information compared with the case of a single image in monocular deraining.

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. (2019) try to remove rain via depth estimation, but they 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. Figure 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.

Considering that there exist semantic segmentation models which are trained on clean images. We also propose an *Enhanced Paired Rain Removal Network (EPRRNet)* to remove rain via utilizing the semantic prior extracted from pre-trained semantic segmentation networks. The *EPRRNet* consists of three sub-networks: a coarse deraining network, a pre-trained semantic segmentation network and a parallel stereo deraining network. The coarse deraining network first reduces rain streaks from input rainy images and generates derained results. The pre-trained semantic segmentation network then estimates the semantic labels from the derained images. Finally, the derained images and semantic labels are fed into a parallel multi-scale deraining network to generate the final derained images via extracting semantic-attentional features.

Our proposed *EPRRNet* is a variant of the *PRRNet* with a number of critical extensions: (1) We directly use a pre-trained semantic network to extract semantic labels as prior. In this way, our *EPRRNet* is flexible to use up-to-date semantic segmentation networks to help remove rain streak. Meanwhile, the coarse deraining network can alleviate the negative effect of rain for semantic segmentation. (2) A multi-scale derained network is proposed to extract multi-scale features to obtain better details. (3) We use a new parallel cooperation way to fuse stereo images which can make better use of stereo information from low-level to high-level layers.

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 four-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 extend the *PRRNet* to an enhanced version, *EPRRNet*. It is a multi-scale network which can flexibly use up-to-date semantic segmentation model to obtain semantic prior. Meanwhile, it also use a parallel cooperation way to fuse stereo information from low-level to high-level layers.
- Finally, we synthesize two stereo rainy datasets for stereo deraining, which may be the largest datasets for stereo image deraining. Experiments on the monocular and stereo rainy datasets show that the proposed *PRRNet* and *EPRRNet* achieve 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

$$O = B + R\,, \tag{1}$$

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. (2011) 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. (2013) and Zhu et al. (2017) use sparse coding based methods to remove rain from a single image. Some works aim to remove rain based on low-rank representation (Chen & Hsu, 2013; Zhang et al., 2017) . Chen and Hsu (2013) 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. (2016) 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 (Ledig et al., 2017; Niu et al., 2020; Zhang et al., 2021) , deblurring (Zhang et al., 2018, 2020; Li et al., 2021; Zhang et al., 2022) , dehazing (Ren et al., 2016; Zheng et al., 2021) , which also include deraining (Li et al., 2019; Zhang et al., 2019; Fu et al., 2017, ?; Yang et al., 2017; Zhang & Patel, 2018; Li et al., 2018; Eigen et al., 2013; Qian et al., 2018; Zheng et al., 2019; Zhang et al., 2021, ?, ?; Yasarla et al., 2019, 2020; Yasarla & Patel, 2020; Yang et al., 2020, 2021; Li et al., 2020, 2019; Yang et al., 2019; Zamir et al., 2021; Deng et al., 2020) . 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 (Hu et al., 2019) , heavy rain regions (Li et al., 2019) or density (Zhang & Patel, 2018) . However, to the best of our knowledge, there are few deep deraining works which try to remove rain via scene understanding (Long et al., 2015) .

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 such as the temporal context and motion information (Garg & Nayar, 2004, 2006) . Researchers formulate rain streaks based on their intrinsic characteristics (Zhang et al., 2006; Liu et al., 2009; Santhaseelan & Asari, 2015; Brewer & Liu, 2008; Jiang et al., 2017) or propose some learning-based methods to improve the performance of deraining models (Chen & Chau, 2013; Tripathi & Mukhopadhyay, 2012; Kim et al., 2015; Wei et al., 2017; Ren et al., 2017) . For example, Santhaseelan and Asari (2015) and Barnum et al. (2010) extract phase congruence features and Fourier domain features, respectively, to remove rain streaks. Chen and Chau (2013) apply photometric 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 (Liu et al., 2018, ?; Chen et al., 2018; Yang et al., 2019) . Chen et al. (2018) propose a robust deep deraining model via applying super-pixel segmentation to decompose the scene into depth consistent unites. Liu et al. (2018) 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. (2019) consider the additional degradation factors in the 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 (Godard et al., 2017; Eslami et al., 2016; Luo et al., 2016) and novel tasks (Jeon et al., 2018; Li et al., 2018; Chen et al., 2018; Zhou et al., 2019) . However, there are few methods that leverage the stereo images to remove rain so far. Tanaka et al. (2006) 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 derained left-view images, Kim et al. (2014) warp the spatially adjacent right-view frames and subtract the 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 Sect. 3.1 and how to train the proposed module based on images and semantic-annotated images in Sect. 3.2. Then, a semantic-rethinking loop is discussed in Sect. 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 the derained ones (Li et al., 2019) . Inspired by Hu et al. (2019), 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. Figure 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

$$P = D(E(I), T)\,, \tag{2}$$

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

$$I_{de} = \sigma_{de}(P \mid T_{de})\,, \tag{3}$$

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

The branch of understanding scenes can be formulated as

$$I_{seg} = \sigma_{seg}(P \mid T_{seg})\,, \tag{4}$$

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

Based on the conditional architecture (Zhao et al., 2018) , 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:

$$\mathcal{L}_{de} = ||I_c - \sigma_{de}(D(E(I_{rainy}), T_{de}))||^2\,, \tag{5}$$

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

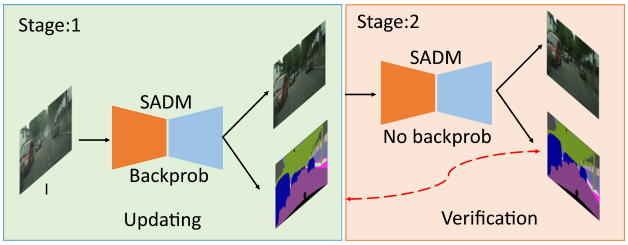


Fig. 3 The Semantic-rethinking Loop. During training, rainy images are fed into *SADM* to generate deraining and segmentation results in stage I. Then the deraining images are fed into the *SADM* to generate segmentation results again in stage II. Through comparing the two segmentation results from rainy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

remove rain from rainy images by leveraging semantic information. The learning process of scene understanding can be denoted as

$$\mathcal{L}_{seg} = \sigma_h(I_{seg}^{gt}, I_{seg}) , \tag{6}$$

where I_{seg} and I_{seg}^{gt} indicate the scene understanding of the model and ground truth labels from auxiliary training sets. In practice, we set the task of image deraining as $T_{de} = 1$ and task of semantic segmentation as $T_{seg} = 0$. Among the network, we use 2-dimensional maps to indicate different task. σ_h is the cross-entropy loss function.

3.3 Semantic-Rethinking Loop

Semantic information plays an important role in various tasks of computer vision (Shen et al., 2018, 2020; Zhang et al., 2021, 2019, 2020, ?) . 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. Figure 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 derain. 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

$$\mathcal{L}_{con} = ||I_{seg}^{ver} - I_{seg}^{up}|| , \tag{7}$$

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

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 Sect. 4.1, and then two core sub-networks will be discussed in Sects. 4.2 and 4.3. Finally, the objective functions to train the proposed model will be presented in Sect. 4.4. Section 4.5 provides implementation details.

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 Sect. 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.

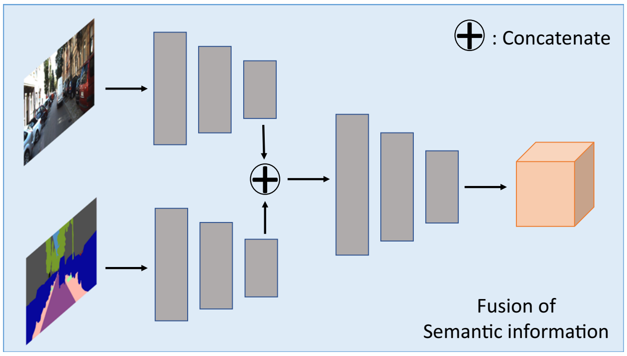


Fig. 4 The architecture of *SFNet*. The coarse deraining images and semantic segmentation results from *SADM* are fed into *SFNet* to generate features volume with semantic information

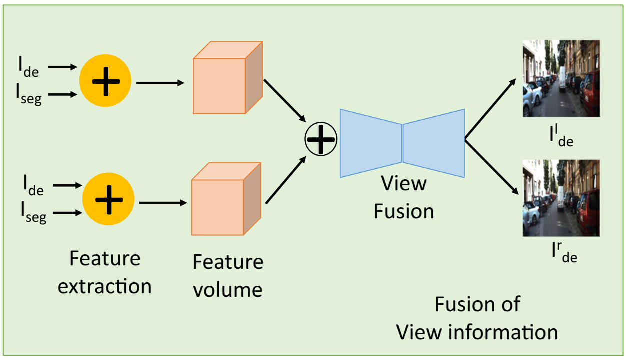


Fig. 5 The architecture of *VFNet*. Features volumes from stereo images are fused to generate final stereo deraining images

4.3 VFNet

Figure 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 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

$$\mathcal{L}_{view} = ||I_{de}^{left} - I_{gt}^{left}|| + ||I_{de}^{right} - I_{gt}^{right}||, \tag{8}$$

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

$$\mathcal{L}_f = \mathcal{L}_{de} + \lambda_1 \mathcal{L}_{seg} + \lambda_2 \mathcal{L}_{con} + \lambda_3 \mathcal{L}_{view}, \tag{9}$$

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.

4.5 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 × 3 × 3) which are utilized to fuse the semantic information. *VFNet* contains five ResBlocks (He et al., 2016) to generate final deraining results. Each ResBlock consists of three CNN layers of 64 × 3 × 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.

5 The Enhanced Paired Rain Removal Network

Currently, there exist several well pre-trained networks for semantic segmentation which can extract satisfactory semantic information. However, the unwanted rain streaks make the labels extracted from the rainy images incorrect. In this section, we introduce an Enhanced Paired Rain Removal Network (EPRRNet) under the "coarse-to-fine" scheme to reduce the effect of rain streaks and make better use of the semantic prior. We first give an overview of the proposed *EPRRNet* model. Then, we introduce four core modules, coarse deraining network, semantic segmentation network, Enhanced SFNet (*ESFNet*) and Enhanced VFNet (*EVFNet*). Sections 3, 4 and 5 are in an evolving sequence. Sections 3 and 4 are the same as our previous conference paper (Zhang et al., 2020) . While the proposed EPRRN network is our best model in this paper.

5.1 Overall

In our preliminary work (Zhang et al., 2020) , we build a *SADM* to extract semantic labels from the input rainy images and then apply an *SFNet* and a *VFNet* to fuse semantic prior and information from stereo views, respectively. However, the unwanted rain streaks make the labels extracted from the

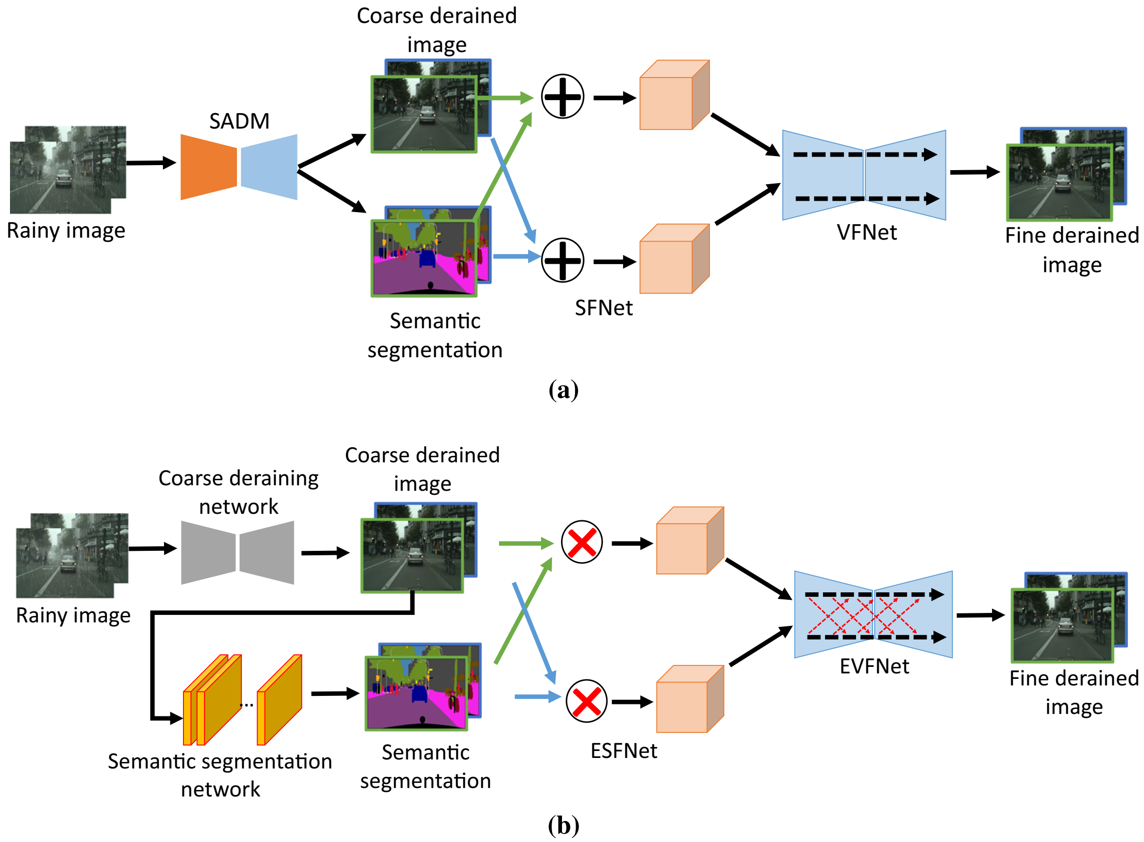


Fig. 6 Overview of the proposed model. The top figure is the framework of the state-of-the-art stereo deraining method (Zhang et al., 2020) , which directly predict the semantic labels from rainy images to help their following process. The bottom figure is our proposed enhanced version, which first reduces the rain streaks via a coarse deraining network. In addition, the proposed network replaces the *SFNet* and *VFNet* with *ESFNet* and *EVFNet* in a semantic-attentional and parallel cooperation ways, respectively

rainy images incorrect. In this section, we improve from the following aspects:

- To reduce the effect of rain streaks and extract better semantic labels, we first use a coarse deraining network to obtain coarse derained results, and then apply a pre-trained network to predict the semantic labels.
- The *SFNet* is replaced by an *ESFNet* to fuse the semantic prior. Specially, we apply a semantic-guided attention mechanism to learn semantic-attentional features, and then combine with non-semantic-attention features to remove rain streaks.
- The *VFNet* is replaced by an *EVFNet* to fuse the information from stereo views. The proposed *EVFNet* is a parallel stereo network. Stereo images are fed into networks to extract two different types of information. Different from the *VFNet* which fuses the high-level features from two views to obtain the final derained images, the cooperation between stereo images in our proposed *EVFNet* is carried out from the low-level layers to high-level layers in a parallel way.

The differences between the *PRRNet* and *EPRRNet* are shown in the Fig. 6.

5.2 Coarse Deraining Network

To reduce the negative effect of the rain streaks, we build a coarse deraining network to obtain the coarse deraining results:

$$I_{de}^{c} = G(I_i)\,, \tag{10}$$

where I_i and G are the input rainy images and the coarse deraining network, respectively.

The coarse deraining network is similar to the *SADM* with several variations. Firstly, as the output is only coarse deraining images, we use a one-branch output network to replace the two-branch model. Secondly, we use DenseBlocks to build the coarse deraining network instead of AutoEncoder. Specially, it consists of three modules, *i.e.*, the pre-processing module, the backbone module and the post-processing module. The overall architecture of the proposed network is shown in the Fig. 7. The pre-processing modules includes

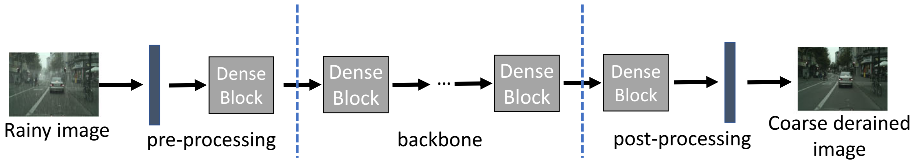


Fig. 7 The architecture of the coarse deraining network. It consists of three modules, *i.e.*, pre-processing module, backbone module and post-processing module. Its input and output are the rainy images and coarse derained results, respectively

a CNN and a DenseBlock to generate 16 feature maps. The backbone module consists of five DenseBlocks, whose input is the output of the pre-processing model. The details of DenseBlock is similar to (Zhang et al., 2020) . The feature maps in the backbone module is also set to 16. We use ReLU as the activation function for all convolutional layers. In order to restore derained images with better details, we build a post-processing module to remove the artifacts and improve the quality of the restored images. It consists of a DenseBlock and a convolutional layer.

5.3 Semantic Segmentation Network

We use a hierarchical multi-scale model (Tao et al., 2020) to predict the semantic information. The output of the coarse deraining network is fed into the semantic segmentation network S to generate the predication of semantic labels I_{seg}:

$$I_{seg} = S(I_{de}^{c}) , \tag{11}$$

where $I_{seg} \in \mathbb{R}^{H \times W \times K}$. H, W and K represent the height, weight of input images, and the number of semantic classes. The prediction of semantic information encode different objects of input images like face, cars and buildings and can be applied as priors for recovering derained images. In this paper, the K is set as 30 to obtain 30 semantic labels for fine deraining process.

5.4 Enhanced SFNet

As the semantic information is able to provide strong global prior for images restoration, we take the estimated semantic labels as a guidance to learn semantic-attentional features, which are concatenated with features without semantic attention, to help remove rain streaks.

Specially, we first use a pre-processing module to encode the input coarse derained images, which are then fed into a CNN network to obtain attention weights via modeling the semantic-guided attention mechanism. Similar to Hu et al. (2019), the CNN network contains 7 convolutional layers. Its output is a set of attention weights $\{A_i, A_2, ..., A_n\}$, which are fed into a Softmax layer to normalize and obtain the attention weights $\{W_i, W_2, ..., W_n\}$, which is formulated as:

$$W_i = \frac{e^{A_c}}{\sum_{c=1}^{n} e^{A_c}} , \tag{12}$$

where c is the channel of features. Figure 8 shows the process of fusing semantic-attentional features.

5.5 Enhanced VFNet

In order to make better use of the information from stereo views, we introduce a parallel stereo network to convolve features from stereo images. Different from the above VFNet which combines stereo information in the last layers, the *EVFNet* makes use of stereo information from low-level to high-level layers.

Figure 9 illustrates the overall architecture of EVFNet, the backbone is similar to the coarse deraining network. The EVFNet contains two share-weights streams to process input features in a parallel way. It divides the input two stereo images into two sub-nets and use a regular convolution over each sub-net to extract features. In order to exchange the information across stereo images, a stereo-image fusion module is applied between two streams as Fig. 9. Specially, the output of each DenseBlock from the top sub-net is combined with the output of each DenseBlock from the bottom sub-net. Finally, the output of backbone is fed into the post-processing modules to obtain final derained results.

6 Experiments

6.1 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 versions. In this paper, we first use Photoshop to create a synthetic RainKITTI2012 dataset based on the public KITTI stereo 2012 dataset (Geiger et al., 2013) . Specifically, we synthesize rainy images by referring to the official guidance of Photoshop in https://www.photoshopessentials.com/photo-effects/photoshop-weather-effects-rain/. We use this method to synthesize rainy images because some previous methods

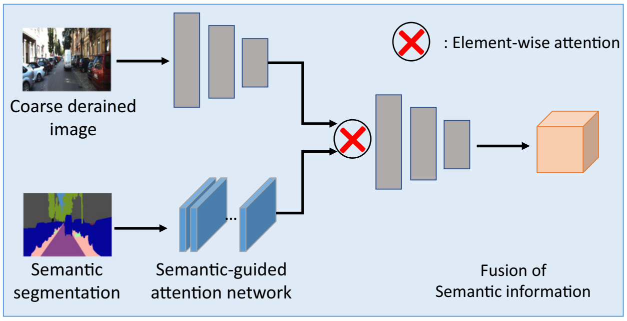


Fig. 8 The architecture of the EVFNet. The input is the feature volume extracted from stereo images. Then both of them are fed into a parallel stereo backbone. The output of each DenseBlock from the left image is concatenated with the output from the right image. In this way, the communication between stereo images are conducted from low-level to high-level layers. The dense block in EVFNet backbone is the same as the Fusion part. The blocks in different parts share weights

use the same tool to synthesize rainy images and achieve satisfied deraining performance (Zhang & Patel, 2018; Fu et al., 2017) . In addition, Photoshop has a good reputation in the field of image processing. 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 (Geiger et al., 2013) . 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. (2019) based on the Cityscapes dataset (Cordts et al., 2016) . 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.

6.2 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 the one 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 reports the quantitative comparison results on the dataset of RainKITTI2012. Figure 10 shows the exemplar qualitative comparison results on the same dataset. *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. Moreover, in order to make better use of semantic information and stereo views, we also build two enhanced versions of *PRRNet*, *i.e.*, *EPPRNet(monocular)* and *EPPRNet(stereo)*, whose performance results are also shown in Table 1 and Fig. 10.

Table 1 Ablation study results on the RainKITTI2012 dataset. Both PSNR and SSIM values are reported

Methods	PSNR	SSIM
PRRNet(D)	30.71	0.923
PRRNet(D+S)	31.56	0.928
PRRNet(D+S+L)	31.89	0.930
EPRRNet(monocular)	32.38	0.935
PRRNet(stereo)	33.01	0.936
EPRRNet(stereo)	34.13	0.947

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. 10. Additional components incrementally improve the visibility of the input image, and the image generated by *PRRNet(stereo)* is the much closer to the ground truth. The proposed *EPRRNet* achieves better performance. Specially, in the monocular case, *EPRRNet(monocular)* achieves the best performance. But it is still inferior to the *PRRNet(stereo)*. We suspect that the cross-view information indeed provides more cues for deraining. *EPRRNet(stereo)* is better than *PRRNet(stereo)*, and achieves the best performance among all the variants. This demonstrates the advantage of using semantic prior and parallel stereo network.

6.3 Stereo Deraining

We quantitatively and qualitatively compare our *PRRNet* with current state-of-the-art methods, which include DDN (Yang et al., 2017) , DID-MDN (Zhang & Patel, 2018) , DAF-Net (Hu et al., 2019) and DeHRain (Li et al., 2019) . Tables 2 and 3 show the quantitative results on our synthesized RainKITTI2012 and RainKITTI2015 datasets, respectively. In both tables, our monocular versions, *PRR-*

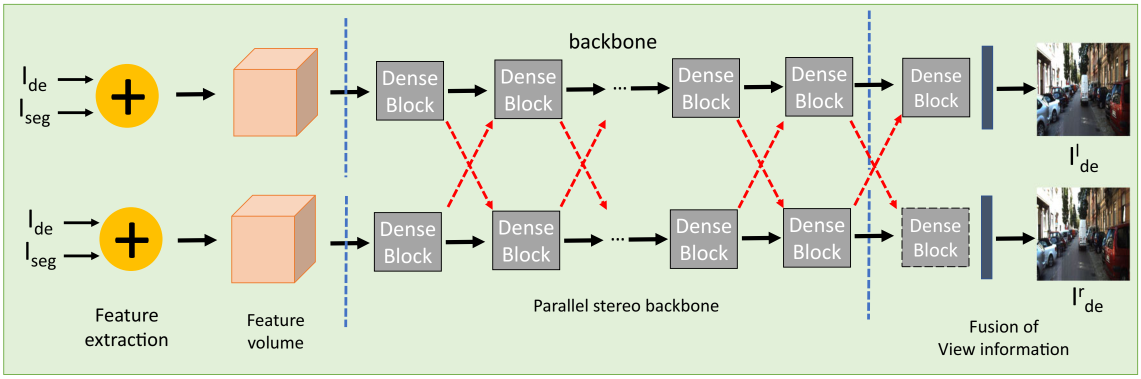


Fig. 9 The architecture of the ESFNet. The input is coarse derained images and semantic labels. The semantic-attentional features are calculated and fed into the following layers to help remove rain streaks

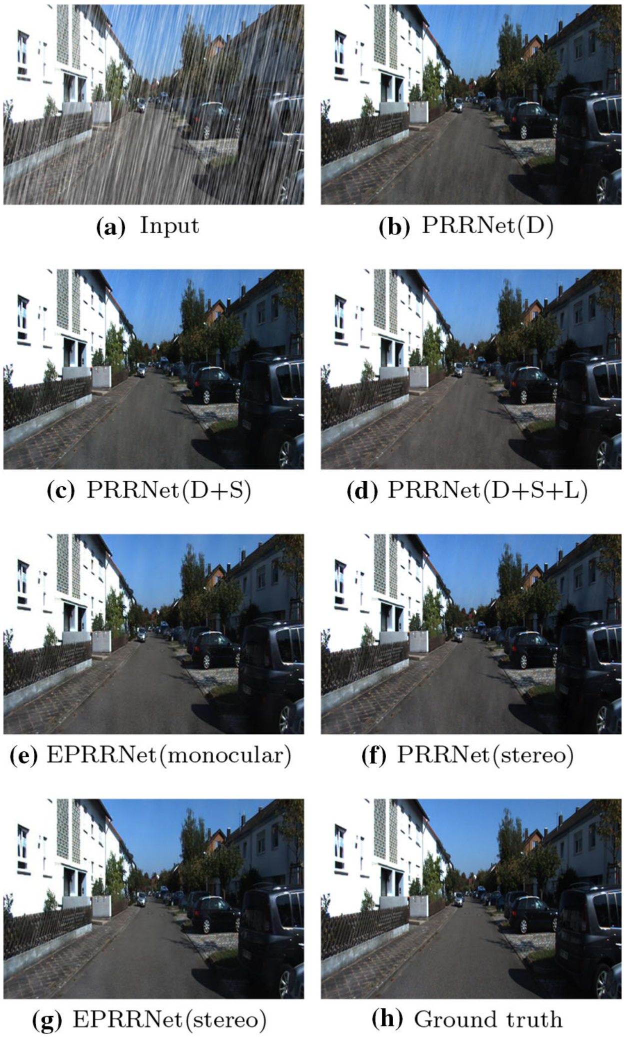


Fig. 10 Exemplar results of deraining evaluation of different variant models on the dataset of RainKITTI2012

Table 2 Quantitative evaluation on the RainKITTI2012 dataset

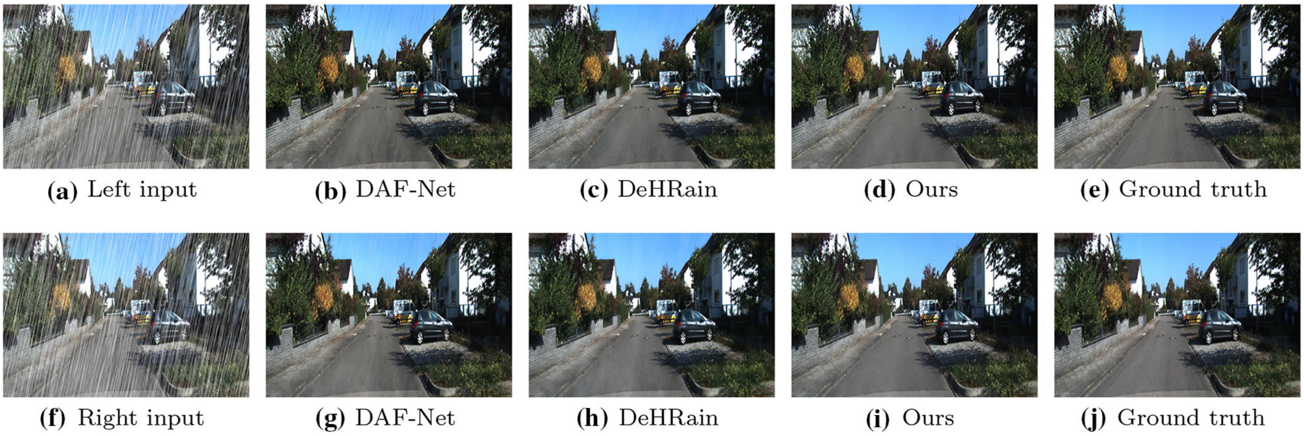
Methods	PSNR	SSIM
DDN (Fu et al., 2017)	29.43	0.904
DID-MDN (Zhang & Patel, 2018)	29.14	0.901
DAF-Net (Hu et al., 2019)	30.44	0.914
DeHRain (Li et al., 2019)	31.02	0.923
PRRNet(monocular)	31.89	0.930
EPRRNet (monocular)	**32.38**	**0.935**
PRRNet(stereo)	33.01	0.936
EPRRNet (stereo)	**34.13**	**0.947**

Table 3 Quantitative evaluation on the RainKITTI2015 dataset

Methods	PSNR	SSIM
DDN (Fu et al., 2017)	29.23	0.906
DID-MDN (Zhang & Patel, 2018)	28.97	0.899
DAF-Net (Hu et al., 2019)	30.17	0.915
DeHRain (Li et al., 2019)	30.84	0.921
EPRRNet(monocular)	31.64	0.932
EPRRNet(monocular)	**32.71**	**0.936**
PRRNet(stereo)	32.58	0.937
EPRRNet(stereo)	**33.83**	**0.943**

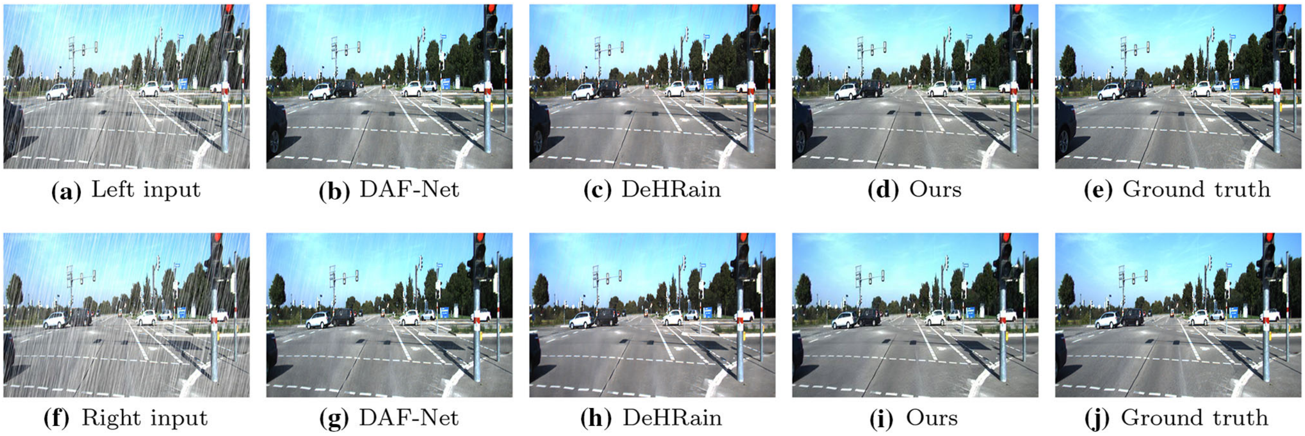
Net(monocular) and *EPRRNet(monocular)*, outperform the existing state-of-the-art methods, with remarkable gain. The model *PRRNet(stereo)* and *EPRRNet(stereo)* achieve the best performance with additional improvement. This demonstrates the superiority of stereo deraining over monocular deraining.

Figures 11 and 12 compare the qualitative performance of our method *EPRRNet(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 truth.


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

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

Fig. 11 Exemplar results of qualitative evaluation of current SOTA models on RainKITTI2012


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

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

Fig. 12 Exemplar results of qualitative evaluation of current SOTA models on RainKITTI2015

6.4 Monocular Deraining

The proposed *EPRRNet* 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 monocular version of our models, *PRRNet(monocular)* and *EPRRNet(monocular)*, with the state-of-the-art methods, including DID-MDN (Zhang & Patel, 2018) , RESCAN (Li et al., 2018) , JOB (Zhu et al., 2017) , GMMLP (Li et al., 2016) , DSC (Luo et al., 2015) , DCPDN (Zhang & Patel, 2018) , and DAF-Net (Hu et al., 2019) , from both quantitative and qualitative aspects.

The quantitative results on the RainCityscapes dataset are shown in Table 4. DID-MDN (Zhang & Patel, 2018) and DCPDN (Zhang & Patel, 2018) perform well and DAF-Net (Hu et al., 2019) outperforms these two methods. Our monocular version *EPRRNet(monocular)* achieves the best performance compared with all the compared methods on this task, revealing the effectiveness of taking semantic segmentation into consideration and the semantic-rethinking loop. Figure 13 compares its qualitative performance with different methods. The results show that the monocular version of our *EPRRNet* also achieves the best performance in terms of monocular image deraining.

6.5 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. Figure 14 shows the qualitative results on two exemplar images from the Internet. Compared to other competing methods, the proposed method *EPRRNet(monocular)* 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. In addition, we provides some more deraining results in Fig. 15. And by the way, the result of the proposed method still has many rain streaks left on the images. Though we have achieved the state-of-the-art performance, there is

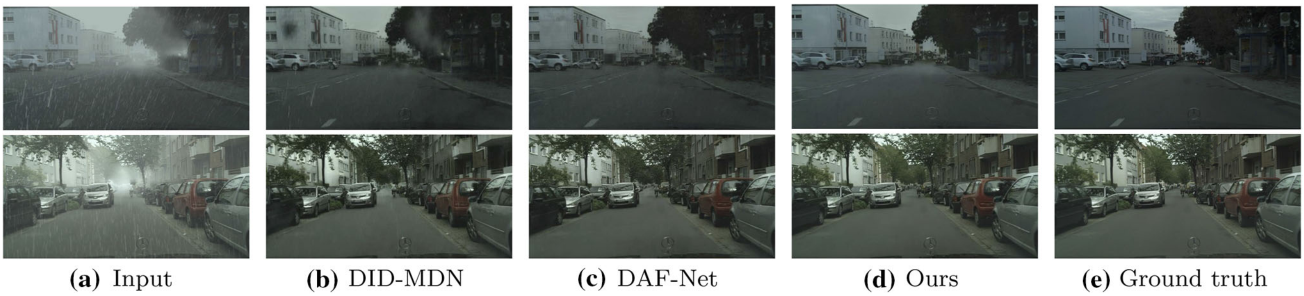


Fig. 13 Exemplar results of qualitative evaluation of current state-of-the-art models on the RainCityscapes dataset

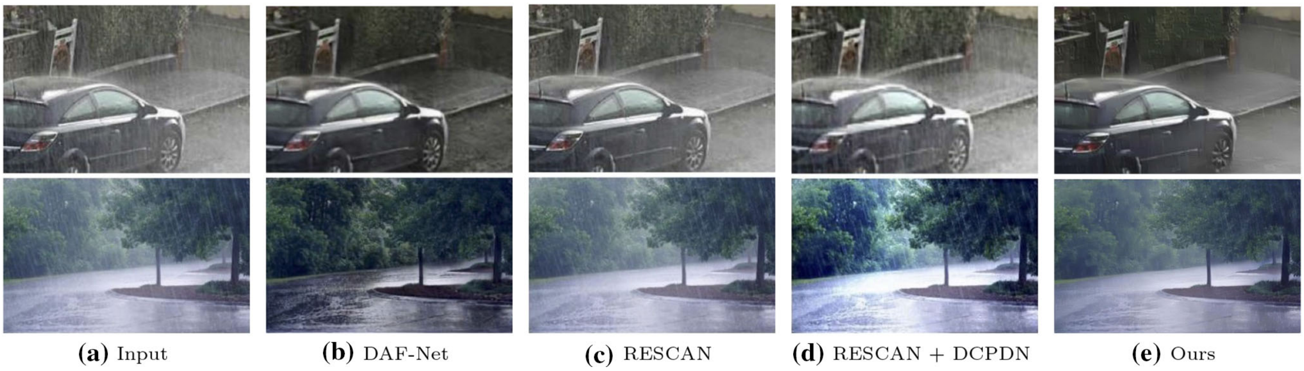


Fig. 14 Qualitative evaluation on real-world rainy images. From left to right are the input images, results of DAF-Net (Hu et al., 2019) , RESCAN (Li et al., 2018) , RESCAN + DCPDN (Zhang & Patel, 2018) and ours, respectively

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

Methods	PSNR	SSIM
DID-MDN (Zhang & Patel, 2018)	28.43	0.9349
RESCAN (Li et al., 2018)	24.49	0.8852
JOB (Zhu et al., 2017)	15.10	0.7592
GMMLP (Li et al., 2016)	17.80	0.8169
DSC (Luo et al., 2015)	16.25	0.7746
DCPDN (Zhang & Patel, 2018)	28.52	0.9277
DAF-Net (Hu et al., 2019)	30.06	0.9530
DRD-Net (Deng et al., 2020)	30.13	0.9535
MPRNet (Zamir et al., 2021)	30.96	0.9721
***PRRNet*(monocular)**	**30.44**	**0.9688**
***EPRRNet*(monocular)**	**31.11**	**0.9741**

still improvement space, especially in the most challenging scenery. We will continue to develop more advanced methods to improve the performance of image deraining.

6.6 Discussion

Autonomous driving has become an increasingly active research field in computer vision with the development of stereoscopic vision (Chen et al., 2015) . Based on stereo images, many key technologies such as depth estimation (Godard et al., 2017; Liu et al., 2015; Riegler et al., 2019) , scene understanding (Eslami et al., 2016; Shao et al., 2015; Zhao et al., 2017) and stereo matching (Luo et al., 2016; Chang & Chen, 2018; Pang et al., 2017) have achieved great success. As an inevitable natural phenomenon in the wild, rain causes visual discomfort and degrades the quality of images, which can deteriorate the performance of many core models, thus increasing the latent danger of autonomous driving (Li et al., 2019) . The stereo deraining methods have potential to improve the quality of stereo images.

In addition, the parameters in this paper are set by our experience. The proposed model with our parameters can achieve the reported results. Though parameters with other settings may achieve better performance, it is not the focus of this paper.

Finally, our experimental results verify that stereo image deraining has an advantage over the monocular deraining. This is corresponding to our analysis in the introduction section. In future, we will consider to explore more relationship between stereo images and rain streaks to design network for stereo image deraining.

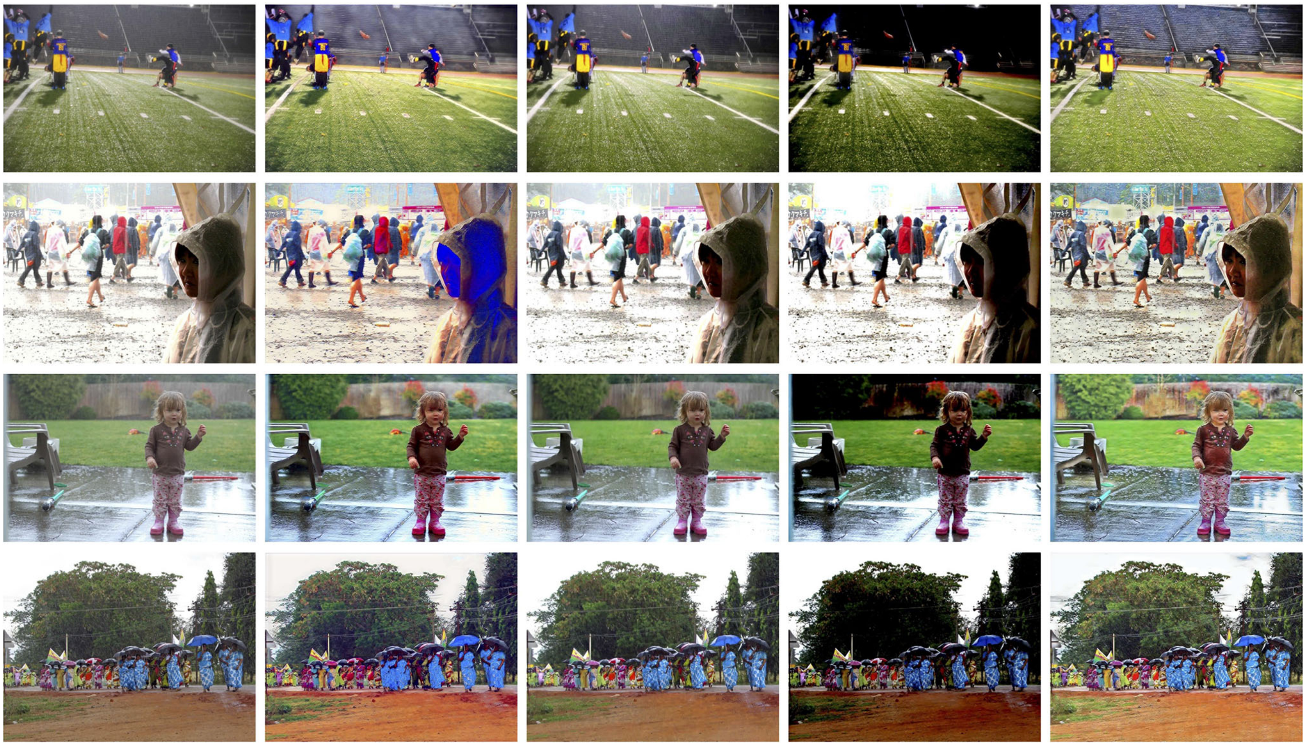


Fig. 15 Qualitative evaluation on real-world rainy images. From left to right are the input images, results of DRD-Net (Deng et al., 2020), DeHRain (Li et al., 2019), MPRNet (Zamir et al., 2021) and ours, respectively

7 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. An enhanced version, *i.e.*, *EPRRNet* is developed with a parallel stereo fusion module. The experimental results show that our proposed *PRRNet* and *EPRRNet* outperform the state-of-the-art methods on both monocular and stereo image deraining.

Acknowledgements This research was funded in part by ARC-Discovery grant projects (DP 190102261 and DP220100800), a Ford Alliance URP grant, and National Natural Science Foundation of China (NSFC) under Grant 62122013, U2001211.

References

Barnum, P. C., Narasimhan, S., & Kanade, T. (2010). Analysis of rain and snow in frequency space. *International Journal of Computer Vision (IJCV), 86,* 256–274.

Brewer, N., & Liu, N. (2008). 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).*

Chang, J.R., & Chen, Y.S. (2018). Pyramid stereo matching network. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).*

Chen, Y.L., & Hsu, C.T. (2013). A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Proceedings of the IEEE international conference on computer vision (ICCV).*

Chen, C., Seff, A., Kornhauser, A., & Xiao, J. (2015) Deepdriving: Learning affordance for direct perception in autonomous driving. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV).*

Chen, J., Tan, C.H., Hou, J., Chau, L.P., & Li, H. (2018). 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).*

Chen, D., Yuan, L., Liao, J., Yu, N., & Hua, G. (2018). Stereoscopic neural style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).*

Chen, J., & Chau, L. P. (2013). A rain pixel recovery algorithm for videos with highly dynamic scenes. *IEEE Transactions on Image Processing (TIP), 23,* 1097–1104.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016) The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).*

Deng, S., Wei, M., Wang, J., Feng, Y., Liang, L., Xie, H., Wang, F.L., & Wang, M. (2020). Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14560–14569.

Eigen, D., Krishnan, D., & Fergus, R. (2013). Restoring an image taken through a window covered with dirt or rain. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Eslami, S.A., Heess, N., Weber, T., Tassa, Y., Szepesvari, D., Hinton, G.E., et al. (2016). Attend, infer, repeat: Fast scene understanding with generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., & Paisley, J. (2017) Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Fu, X., Huang, J., Ding, X., Liao, Y., & Paisley, J. (2017). Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing (TIP), 26*(6), 2944–2956.

Garg, K., & Nayar, S.K. (2004). Detection and removal of rain from videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Garg, K., & Nayar, S.K. (2006). Photorealistic rendering of rain streaks. In: ACM Transactions on Graphics (TOG)

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research (IJRR), 32*(11), 1231–1237.

Godard, C., Mac Aodha, O., & Brostow, G.J. (2017). Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

https://www.photoshopessentials.com/photo-effects/photoshop-weather-effects-rain/.

Hu, X., Fu, C.W., Zhu, L., & Heng, P.A. (2019). Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Huang, D. A., Kang, L. W., Wang, Y. C. F., & Lin, C. W. (2013). Self-learning based image decomposition with applications to single image denoising. *IEEE Transactions on Multimedia (TMM), 16*(1), 83–93.

Jeon, D.S., Baek, S.H., Choi, I., & Kim, M.H. (2018). Enhancing the spatial resolution of stereo images using a parallax prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., & Wang, Y. (2017). 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)*

Kang, L. W., Lin, C. W., & Fu, Y. H. (2011). Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing (TIP), 21*(4), 1742–1755.

Kim, J.H., Sim, J.Y., & Kim, C.S. (2014). Stereo video deraining and desnowing based on spatiotemporal frame warping. In *The IEEE International Conference on Image Processing (ICIP)*.

Kim, J. H., Sim, J. Y., & Kim, C. S. (2015). Video deraining and desnowing using temporal correlation and low-rank matrix completion. *IEEE Transactions on Image Processing (TIP), 24*(9), 2658–2670.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, S., Araujo, I.B., Ren, W., Wang, Z., Tokuda, E.K., Junior, R.H., Cesar-Junior, R., Zhang, J., Guo, X., & Cao, X. (2019). Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, R., Cheong, L.F., & Tan, R.T. (2019). Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, B., Lin, C.W., Shi, B., Huang, T., Gao, W., & Jay Kuo, C.C. (2018). Depth-aware stereo video retargeting. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, R., Tan, R.T., & Cheong, L.F. (2020). All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3175–3185.

Li, R., Tan, R.T., Cheong, L.F., Aviles-Rivero, A.I., Fan, Q., & Schonlieb, C.B. (2019). Rainflow: Optical flow under rain streaks and rain veiling effect. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7304–7313.

Li, Y., Tan, R.T., Guo, X., Lu, J., & Brown, M.S. (2016). Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Li, X., Wu, J., Lin, Z., Liu, H., & Zha, H. (2018). Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *European conference on computer vision (ECCV)*.

Li, D., Xu, C., Zhang, K., Yu, X., Zhong, Y., Ren, W., Suominen, H., & Li, H. (2021). Arvo: Learning all-range volumetric correspondence for video deblurring. arXiv:2103.04260.

Liu, F., Shen, C., & Lin, G. (2015). Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Liu, J., Yang, W., Yang, S., & Guo, Z. (2018). 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)*.

Liu, P., Xu, J., Liu, J., & Tang, X. (2009). Pixel based temporal analysis using chromatic property for removing rain from videos. *Computer and Information Science, 2*(1), 53–60.

Liu, J., Yang, W., Yang, S., & Guo, Z. (2018). D3r-net: Dynamic routing residue recurrent network for video rain removal. *IEEE Transactions on Image Processing (TIP), 28*(2), 699–712.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Luo, W., Schwing, A.G., & Urtasun, R. (2016). Efficient deep learning for stereo matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Luo, Y., Xu, Y., & Ji, H. (2015). Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., & Shen, H. (2020). Single image super-resolution via a holistic attention network. In *European conference on computer vision*, pp. 191–207. Springer.

Pang, J., Sun, W., Ren, J.S., Yang, C., & Yan, Q. (2017). Cascade residual learning: A two-stage convolutional neural network for stereo matching. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Qian, R., Tan, R.T., Yang, W., Su, J., & Liu, J. (2018). Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M.H. (2016). Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision (ECCV)*.

Ren, W., Tian, J., Han, Z., Chan, A., & Tang, Y. (2017). Video desnowing and deraining based on matrix decomposition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Riegler, G., Liao, Y., Donne, S., Koltun, V., Geiger, A. (2019). Connecting the dots: Learning representations for active monocular depth estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

Santhaseelan, V., & Asari, V.K. (2015). Utilizing local phase information to remove rain from video. International Journal of Computer Vision (IJCV)

Shao, J., Kang, K., Change Loy, C., & Wang, X. (2015). Deeply learned attributes for crowded scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Shen, Z., Lai, W.S., Xu, T., Kautz, J., & Yang, M.H. (2018). Deep semantic face deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8260–8269.

Shen, Z., Lai, W. S., Xu, T., Kautz, J., & Yang, M. H. (2020). Exploiting semantics for face image deblurring. *International Journal of Computer Vision, 128*(7), 1829–1846.

Tanaka, Y., Yamashita, A., Kaneko, T., & Miura, K.T. (2006). Removal of adherent waterdrops from images acquired with a stereo camera system. IEICE Transactions on Information and Systems (IEICE TIS).

Tao, A., Sapra, K., & Catanzaro, B. (2020). Hierarchical multi-scale attention for semantic segmentation. arXiv:2005.10821.

Tripathi, A., & Mukhopadhyay, S. (2012). Video post processing: Low-latency spatiotemporal approach for detection and removal of rain. *IET Image Processing, 6*(2), 181–196.

Wei, W., Yi, L., Xie, Q., Zhao, Q., Meng, D., & Xu, Z. (2017). Should we encode rain streaks in video as deterministic or stochastic? In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Yang, W., Liu, J., & Feng, J. (2019). Frame-consistent recurrent video deraining with dual-level flow. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., & Yan, S. (2017). Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Yang, W., Tan, R.T., Feng, J., Wang, S., Cheng, B., & Liu, J. (2021). Recurrent multi-frame deraining: Combining physics guidance and adversarial learning. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Yang, W., Tan, R. T., Feng, J., Guo, Z., Yan, S., & Liu, J. (2019). Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42*(6), 1377–1393.

Yang, W., Tan, R. T., Wang, S., Fang, Y., & Liu, J. (2020). Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(11), 4059–4077.

Yasarla, R., & Patel, V.M. (2019). Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8405–8414.

Yasarla, R., Sindagi, V.A., & Patel, V.M. (2020). Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2726–2736.

Yasarla, R., & Patel, V. M. (2020). Confidence measure guided single image de-raining. *IEEE Transactions on Image Processing, 29,* 4544–4555.

Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., & Shao, L. (2021). Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14821–14831.

Zhang, H., & Patel, V.M. (2017). Convolutional sparse and low-rank coding-based rain streak removal. In *IEEE Winter conference on applications of computer vision (WACV)*.

Zhang, H., & Patel, V.M. (2018). Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Zhang, H., & Patel, V.M. (2018). 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)*.

Zhang, J., Fan, D.P., Dai, Y., Anwar, S., Saleh, F.S., Zhang, T., & Barnes, N. (2020). 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*, pp. 8582–8591.

Zhang, K., Li, D., Luo, W., Lin, W.Y., Zhao, F., Ren, W., Liu, W., & Li, H. (2021) Enhanced spatio-temporal interaction learning for video deraining: A faster and better framework. arXiv:2103.12318.

Zhang, K., Li, D., Luo, W., Ren, W., Ma, L., & Li, H. (2021). Dual attention-in-attention model for joint rain streak and raindrop removal. arXiv:2103.07051.

Zhang, K., Li, D., Luo, W., Ren, W., Stenger, B., Liu, W., Li, H., & Yang, M.H. (2021). Benchmarking ultra-high-definition image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 14769–14778.

Zhang, X., Li, H., Qi, Y., Leow, W.K., & Ng, T.K. (2006) Rain removal in video by combining temporal and chromatic properties. In *IEEE international conference on multimedia and expo (ICME)*.

Zhang, K., Li, R., Yu, Y., Luo, W., Li, C., & Li, H. (2021). Deep dense multi-scale network for snow removal using semantic and geometric priors. arXiv:2103.11298 .

Zhang, K., Luo, W., Ma, L., & Li, H. (2019). Cousin network guided sketch recognition via latent attribute warehouse. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 9203–9210.

Zhang, K., Luo, W., Ren, W., Wang, J., Zhao, F., Ma, L., & Li, H. (2020). Beyond monocular deraining: Stereo image deraining via semantic understanding. In *European Conference on Computer Vision (ECCV)*.

Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., & Li, H. (2020). Deblurring by realistic blurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Zhang, K., Ren, W., Luo, W., Lai, W.S., Stenger, B., Yang, M.H., & Li, H. (2022). Deep image deblurring: A survey. arXiv:2201.10700.

Zhang, J., Yu, X., Li, A., Song, P., Liu, B., & Dai, Y. (2020). Weakly-supervised salient object detection via scribble annotations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12546–12555.

Zhang, K., Luo, W., Zhong, Y., Ma, L., Liu, W., & Li, H. (2018). Adversarial spatio-temporal learning for video deblurring. *IEEE Transactions on Image Processing (TIP), 28*(1), 291–301.

Zhang, H., Sindagi, V., & Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 30*(11), 3943–3956.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017) Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Zhao, F., Zhao, J., Yan, S., & Feng, J. (2018). Dynamic conditional networks for few-shot learning. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 19–35.

Zheng, L., Li, Y., Zhang, K., & Luo, W. (2021). T-net: Deep stacked scale-iteration network for image dehazing. arXiv:2106.02809.

Zheng, Y., Yu, X., Liu, M., & Zhang, S. (2019). Residual multiscale based single image deraining. In *British Machine Vision Conference (BMVC)*.

Zhou, S., Zhang, J., Zuo, W., Xie, H., Pan, J., & Ren, J.S. (2019). Davanet: stereo deblurring with view aggregation. In *Proceedings* *of the IEEE conference on computer vision and pattern recognition (CVPR)*.

Zhu, L., Fu, C.W., Lischinski, D., & Heng, P.A. (2017). Joint bi-layer optimization for single-image rain streak removal. In *Proceedings of the IEEE international conference on computer vision (ICCV)*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.