Close the Loop: A Unified Bottom-up and Top-down Paradigm for Joint Image Deraining and Segmentation

Yi Li[1], **Yi Chang**[1,2*], **Changfeng Yu**[1], **Luxin Yan**[1]

[1] National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology

[2] Pengcheng Laboratory

{li_yi, yichang, ycf, yanluxin}@hust.edu.cn

Abstract

In this work, we focus on a very practical problem: image segmentation under rain conditions. Image deraining is a classic low-level restoration task, while image segmentation is a typical high-level understanding task. Most of the existing methods intuitively employ the bottom-up paradigm by taking deraining as a preprocessing step for subsequent segmentation. However, our statistical analysis indicates that not only deraining would benefit segmentation (bottom-up), but also segmentation would further improve deraining performance (top-down) in turn. This motivates us to solve the rainy image segmentation task within a novel top-down and bottom-up unified paradigm, in which two sub-tasks are alternatively performed and collaborated with each other. Specifically, the bottom-up procedure yields both clearer images and rain-robust features from both image and feature domains, so as to ease the segmentation ambiguity caused by rain streaks. The top-down procedure adopts semantics to adaptively guide the restoration for different contents via a novel multi-path semantic attentive module (SAM). Thus the deraining and segmentation could boost the performance of each other cooperatively and progressively. Extensive experiments and ablations demonstrate that the proposed method outperforms the state-of-the-art on rainy image segmentation.

Introduction

The image deraining (Fu et al. 2020; Deng et al. 2020) and segmentation (Zhang et al. 2018; Yang et al. 2018) have made great progress during the past few years. The former is a classic low-level restoration task, while the latter is a typical high-level understanding task. Most of the previous methods focus on one task and consider the two tasks separately. However, there are fewer works considering the practical problem: image segmentation under rain conditions.

To solve this problem, the relationship between degradation and segmentation has been preliminarily studied. On one hand, a number of works analyze the influence of various degradations and their removal to high-level segmentation (Sakaridis, Dai, and Van Gool 2018; Pei et al. 2021). For example, Kamann *et al.* (Kamann and Rother 2020) reached a conclusion that segmentation models generalize well for image noise/blur, however, not with respect to weather corruptions. On the other hand, pioneer works introduce high-level tasks to evaluate the low-level restoration. For example, detection (Li et al. 2019) and recognition (Scheirer et al. 2020) are employed to evaluate deraining performance.

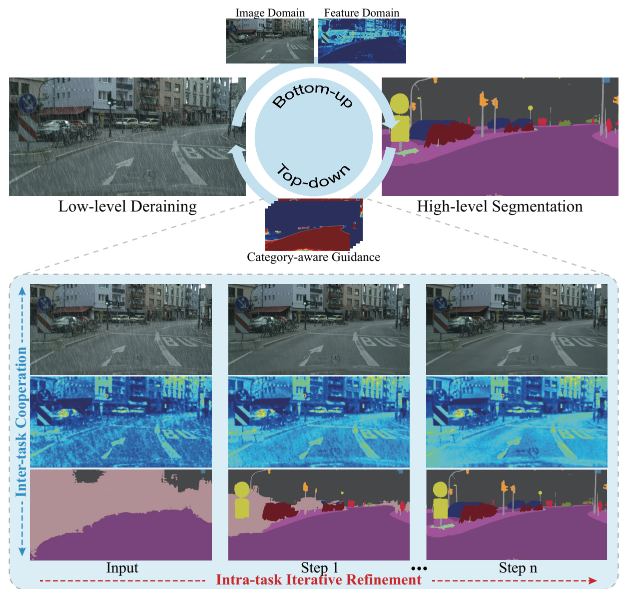


Figure 1: Illustration of the proposed unified bottom-up and top-down paradigm. The low-level deraining (including both the image and feature domains) and high-level segmentation benefit from each other progressively, in which a better deraining result facilitates better segmentation, meanwhile a better segmentation offers better guidance for restoration. Below is the visualization results of the image and feature domain deraining, and segmentation in each iterative step.

Consequently, the existing rainy image segmentation methods can be mainly classified into three categories: segmentation-oriented bottom-up methods, restoration-oriented top-down methods, and multi-task parallel methods. The key idea of the bottom-up methods is to first get rid of negative effects of the degradation, so as to improve the feature discrimination for subsequent segmentation. The degradation removal procedure can be explicit in image do-

*Corresponding Author

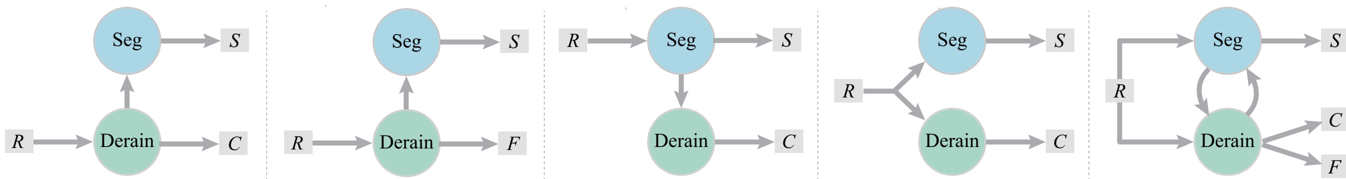


Figure 2: Comparison of existing rainy image segmentation techniques. R, S, C, and F represent rainy image, segmentation prediction, clear image, and clean feature, respectively. Rather than focusing on unidirectional promotion for deraining or segmentation, the proposed method considers unified bidirectional facilitation for both tasks via an iterative manner.

main (Sharma et al. 2018; Porav, Bruls, and Newman 2019; Guo et al. 2020; Liu et al. 2020) [Fig. 2(a)] or implicit in feature domain (Valada et al. 2017; Sakaridis, Dai, and Van Gool 2020; Dai et al. 2020) [Fig. 2(b)]. On the contrary, the main goal of the top-town methods is to improve the restoration performance with high-level semantic guidance providing sufficient structural priors (Hu et al. 2019; Ren et al. 2018; Zhang et al. 2021) [Fig. 2(c)]. The last one of the multi-task parallel methods is capable of jointly removing rain and understanding scenes with shared learning mechanism (Zhang et al. 2020) [Fig. 2(d)].

Most of the previous methods either follow the bottom-up paradigm or the top-down paradigm, which is only a unidirectional promotion for deraining or segmentation. Few of them (Valada et al. 2017; Zhang et al. 2020; Guo et al. 2020) have noticed a simple yet important problem that low-level restoration and high-level segmentation could be mutually promoted and tightly coupled. In this work, we first analyze how rain influences segmentation, and discover that less degradation brings less semantic ambiguity, so as to better segmentation results. Then, we analyze how semantic information affects deraining, and prove that explicit semantic prior plays an important role in promoting restoration. These motivate us to bridge the gap between low-level deraining and high-level segmentation within a unified bidirectional bottom-up and top-down paradigm and propose the unified bidirectional cooperation network (UBCN) [Fig. 2(e)].

Specifically, for the bottom-up stream, we propose to eliminate the negative influence of rainy effect from both image and feature-level instead of single perspective, where both an image-level deraining module and a feature-level domain adaptation module are designed for rain robust segmentation. As for the top-down stream, we propose to explicitly embed the semantic information into the restoration network via a novel multi-path semantic attentive module (SAM) for learning spatial content-aware features to facilitate deraining. The two unidirectional streams are iteratively fed into the other stream collaboratively. Consequently, a better deraining result can be better represented by the network, thus facilitating segmentation; meanwhile, better semantic information can offer more structured semantic constraints in the solution space for better deraining, as shown in Fig. 1. Overall, we summarize main contributions:

- Through statistical analysis, we prove that not only deraining would facilitate segmentation, but also segmentation would further improve deraining. Thus we propose a unified bottom-up and top-down paradigm along with the unified bidirectional cooperation network, considering bidirectional promotion between deraining and segmentation in an iterative feedback manner.
- Compared with previous methods where only the image or feature domain is utilized to eliminate rain effect for segmentation, we propose a joint image- and feature-level domain adaptation scheme, so as to better get rid of the negative influence of rain in segmentation.
- We present a novel semantic attentive module by embedding semantic prior into deraining networks to explicitly learn content-aware features for adaptive restoration. Extensive experiments demonstrate the superiority of our method both quantitatively and qualitatively.

Related Work

Image Deraining. Single image deraining has been widely studied during past few years, including the optimization-based methods (Kang, Lin, and Fu 2011; Chen and Hsu 2013; Luo, Xu, and Ji 2015; Li et al. 2016; Zhu et al. 2017; Chang, Yan, and Zhong 2017) and the deep learning-based methods (Fu et al. 2017; Zhang and Patel 2018; Li et al. 2018; Yasarla and Patel 2019; Yang et al. 2019; Zhu et al. 2019; Wang et al. 2019, 2020a; Yang et al. 2020). Most of existing deraining methods mainly aim at visual appearance and the quantitative PSNR/SSIM metrics. Recently, high-level information has been taken into consideration for practical application. On one hand, researchers begin to take high-level tasks, such as classification (Qian et al. 2018; Li, Cheong, and Tan 2019), detection (Li et al. 2019), and segmentation (Jiang et al. 2020) as the derain evaluation indexes. On the other hand, high-level semantic knowledge has been employed to guide deraining process in a top-down manner (Xu et al. 2021). For example, Zhang *et al.* (Zhang et al. 2020) generated segmentation map through a multi-task framework to improve stereo deraining. Compared with previous methods which utilize semantics with only concatenation operation, we make the utmost of spatial cues provided by segmentation and propose a novel semantic attentive module (SAM), explicitly adopting semantic priors for adaptive restoration of different contents.

Semantic Segmentation. Although considerable progress has been made in semantic segmentation, most of the existing methods mainly focus on degradation-free scenes (Chen et al. 2018; Yu et al. 2020) and may encounter significant

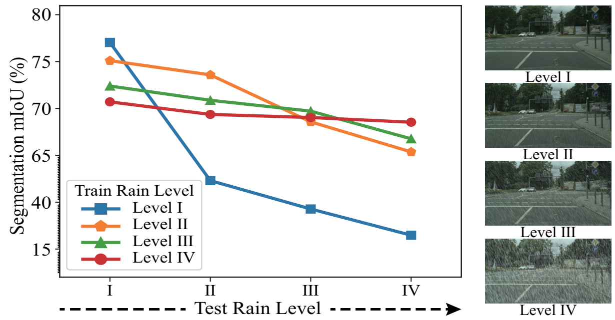


Figure 3: Influence of different rain levels on high-level segmentation. We train the segmentation models on four rain levels and show corresponding segmentation results.

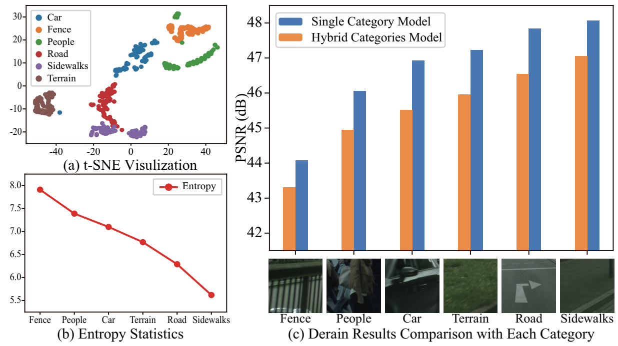


Figure 4: The intrinsic discrepancy between different image contents and their influences on deraining performance. (a) and (b) visualize the t-SNE and entropy of typical patches, respectively. (c) show deraining results for different patches.

performance drop when facing adverse weather. In recent years, the influence of various degradations on segmentation has been widely studied, such as rain (Porav et al. 2019), haze (Sakaridis, Dai, and Van Gool 2018; Dai et al. 2020) and low-illumination (Sakaridis, Dai, and Van Gool 2020). To solve the degraded image segmentation, the researchers mainly start from the bottom-up paradigm. The key idea is first to get rid of the adverse effect of degradation in image domain (Porav et al. 2019) or feature domain (Dai et al. 2020; Sakaridis, Dai, and Van Gool 2020; Wang and Zhang 2021), and then acquire better segmentation prediction in clearer domain. For example, Porav *et al.* (Porav et al. 2019) suppressed rain effects in image domain by explicitly restoring a clear image from the rainy one to improve segmentation performance. Halder *et al.* (Halder, Lalonde, and Charette 2019) eliminated rain in feature domain and directly learned a robust feature representation for rain to facilitate segmentation. In this work, we propose to solve rainy image segmentation from both image- and feature-domain, so as to better get rid of the negative influence of the rain.

Low- and High-level Vision Interaction. Recently, exploring the interaction between low- and high-level vision tasks is drawing attention, e.g., joint image denoising and segmentation (Liu et al. 2020), face image restoration and landmark detection (Sun et al. 2019; Ma et al. 2020). One intuitive way is to establish a pipeline framework to allow one task to facilitate the other (Wang et al. 2016; Ren et al. 2018; Shen et al. 2020; Liu et al. 2020; Sharma et al. 2018). Another research line is to construct a parallel multi-task framework, which treats the two tasks equally with two parallel branches (Huang, Le, and Jaw 2021; Wang et al. 2020b). Unfortunately, these unidirectional approaches have not fully taken the interaction within different tasks. In fact, the collaborative relationship has been explored in face restoration and recognition. Zhang *et al.* (Zhang et al. 2011) presented a joint blind face image restoration and recognition method based on the sparse representation. Ma *et al.* (Ma et al. 2020) proposed a joint face super-resolution and landmark detection network with iterative collaboration. In this work, we concentrate on the task of joint image deraining and segmentation within a unified top-down and bottom-up paradigm.

Unified Bidirectional Cooperation Network

Relation between Deraining and Segmentation

In this section, we briefly analyze the mutual influence between low-level deraining and high-level segmentation.

Bottom-up: how rain degradation affects segmentation. Intuitively, heavy rain may cause great impacts on segmentation due to severe damage of image structures and content, while light rain may cause slight effects. To quantitatively analyze how rain affects segmentation, we evaluate the performance of the segmentation model when applying different rain levels images for training and testing. Specifically, we synthesize four levels of rain (approximate 0, 50, 150, 250 mm/hr) on *Cityscapes* dataset and train four individual segmentation models (Zhao et al. 2017), as shown in Fig. 3. Then, we test the four trained models on different testing levels, corresponding to different color lines.

We have three key observations. First, the segmentation results of all models decrease monotonically when the test rain level increases. *We can conclude that the lighter the rain is, the better the segmentation result is.* Second, the best segmentation result for all cases is clean image training and clean image testing. *That is to say, the clean image is most discriminative for segmentation without any ambiguity caused by the artifacts.* Third, the best segmentation result for each rain level is always obtained only when the training dataset and the testing dataset match. This suggests that the segmentation model should be adaptively associated with the rainy dataset. The first two phenomenons indicate that proper rain removal could indeed facilitate better segmentation. The third motivates us that the segmentation should be tightly coupled with the deraining, as so to well accommodate different rain levels. In the next section, we will introduce the details about how we construct the overall network and how we get rid of rain effects for segmentation.

Top-down: how semantic facilitates deraining. Each image contains abundant content, such as the textured tree, smooth road, and sharp edges of the artificial buildings. That is exactly the semantics that offers pixel-level category information about the content. Thus, we raise a question: *how does different content affect the deraining?*

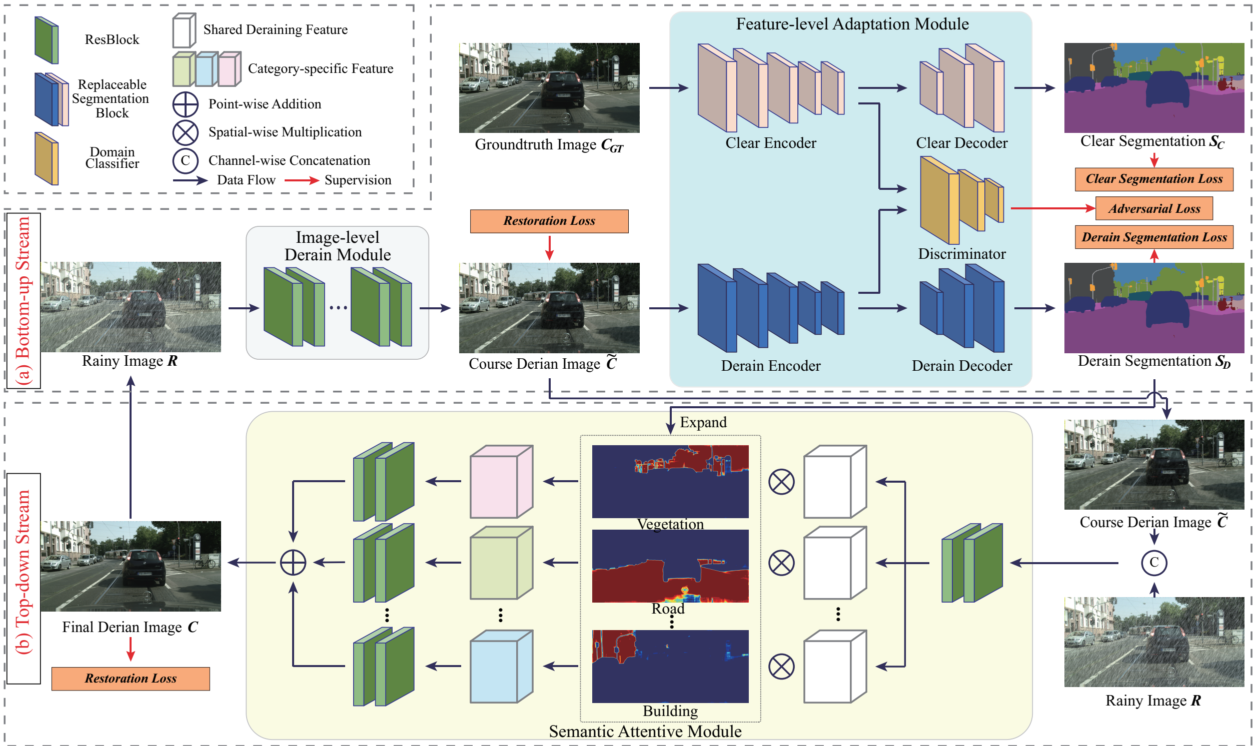


Figure 5: Overview of the proposed UBCN, which is mainly composed of two streams: (a) bottom-up stream for rain-robust semantic representation and (b) top-down stream for adaptive image restoration. Given a rainy image, both image-level deraining and feature-level adaptation are performed to get segmentation rid of rain. Then the obtained segmentation prediction is utilized as explicit instructions for adaptive deraining on different semantic contents via semantic attentive module. Finally, the top-down restoration and bottom-up segmentation are iteratively performed and collaborate with each other.

To answer this question, we first demonstrate that different contents have intrinsic low-dimensional manifolds. We crop a number of $256 * 256$ patches from the original $2048 * 1024$ image in *Cityscapes* dataset. Each patch should contain 95% of a certain category, such as fence, people, or car. Here, we choose six categories as representatives. Then, we perform t-SNE (Van der Maaten and Hinton 2008) on these patches to visualize the two-dimensional distribution of these patches. In Fig. 4(a), we can observe that all categories of the patches have been clustered distinctly. *That is to say, different contents have different feature representation*. Moreover, we calculate the mean entropy of each category to measure the degree of randomness (complexity) in the patch. In Fig. 4(b), we can observe that the entropy gradually decreases in terms of fence, people, car, terrain, road, and sidewalks. The t-SNE and entropy results verify that there exist large discrepancies between different category patches.

Here, we analyze how these different category patches affect deraining performance. We train six deraining models where each one is trained specifically on same category patches, and also a hybrid model for mixed category patches. The results are shown in Fig. 4(c). The result of category-aware single model is consistently better than the hybrid model, indicating that high-level category prior information would definitely benefit low-level deraining. Moreover, it is very interesting that the deraining results are inverse to the entropy to each category: the more complex the semantic patch is (higher entropy), the worse the deraining result is.

Overall, different category patches possess significant variance on both visual appearance and intrinsic dimension, which lead to different difficulties of deraining. We argue that learning category-specific features can be beneficial to the deraining. This motivates us to exploit high-level semantic information to adaptively guide the low-level deraining.

Overall Architecture

As we have analyzed above, the deraining could indeed facilitate better segmentation, and on the contrary, semantic information would be beneficial to deraining. To bridge the gap between low-level deraining and high-level segmentation, we propose a unified bidirectional cooperation network for joint image deraining and segmentation within a unified bottom-up and top-down paradigm, as shown in Fig. 5. The proposed UBCN is composed of two main streams within a cycle: one bottom-up stream for rain-robust semantic representation and one top-down stream for adaptive image deraining. In fact, we implement the cycle via an iterative manner to perform the two streams alternatively and collaboratively. On one hand, bottom-up process greatly eases rain degradation from both image- and feature-level to improve segmentation. On the other hand, top-down processes exploit semantic prior to explicit guide adaptive deraining on different semantic contents via semantic attentive module. At last, each iteration takes the output from last iteration as input, further improving results of both tasks.

Bottom-up: Joint Image & Feature Domain

The main goal of bottom-up methods is to get rid of rain effect in segmentation. Previous methods suppressed the rain artifacts either in the image domain or the feature domain so as to obtain rain-robust segmentation. Although the image-domain deraining could better visually remove the rain, the image details along with the corresponding discriminative feature may inevitably be damaged. On the contrary, the feature-domain adaptation method could well extract rainy-robust representation without losing original information.

In this work, we propose a joint image and feature adaptation network in bottom-up stream, as shown in Fig. 5(a). Specifically, explicit deraining is first performed via image-level deraining module (22 Resblocks) with restoration loss:

$$\mathcal{L}_{res_c} = ||\tilde{\boldsymbol{C}} - \boldsymbol{C}_{gt}||^2, \tag{1}$$

where $\tilde{\boldsymbol{C}}$ is the estimated derain image, $\boldsymbol{C}_{gt}$ is the clean image. Although it is nearly impossible to eliminate all rain effects directly from image domain, the derain image is much more similar to the clear one than the original rainy version.

Next, we design a feature-level adaptation module to enforce that the derain image $\tilde{\boldsymbol{C}}$ and clean image $\boldsymbol{C}_{gt}$ are also indistinguishable in the feature domain with adversarial loss:

$$\mathcal{L}_{adv} = log(1 - D(G_{de}(\tilde{\boldsymbol{C}}))) + log(D(G_{ce}(\boldsymbol{C}_{gt}))), \tag{2}$$

where G_{de} and G_{ce} are the encoder of the derain and clean image, respectively, and D is the discriminator. Note that, we employ the same encoder-decoder architecture for both clean and derain images, while we do not share the same weights for them. According to our experiment, the two patches with different weights, which means flexibility and representation, would have better performance.

Finally, after the decoder, we can obtain two segmentation results and utilize cross-entropy loss for optimization:

$$\mathcal{L}_{CSeg} = -\sum_c \boldsymbol{S}_{gt}^{(c)} log(\boldsymbol{S}_C^{(c)}), \tag{3}$$

$$\mathcal{L}_{DSeg} = -\sum_c \boldsymbol{S}_{gt}^{(c)} log(\boldsymbol{S}_D^{(c)}), \tag{4}$$

where c is the number of class, $\boldsymbol{S}_D$, $\boldsymbol{S}_C$, $\boldsymbol{S}_{gt}$ is segmentation results of derain, clean, and GT, respectively. In fact, each encoder-decoder is a segmentation network. Here we employ well-known backbones PSPNet50 (Zhao et al. 2017) and HRNet18 (Wang et al. 2021) as our replaceable segmentation network. Overall, joint image and feature adaptation would obtain satisfactory rain-robust segmentation results.

Top-down: Semantic Attentive Restoration

In the top-down stream, the high-level semantic information can be used to adaptively guide the low-level deraining. The key question is how to utilize the semantic properly. The most common way is a concatenate operation (Ren et al. 2018) between input and semantic information due to same spatial dimension. However, such straightforward operator may not fully explore spatial cues provided by semantics.

In this work, we propose a novel semantic attentive module to learn category-specific features in Fig. 5(b). The key idea of SAM is divide and conquer. First, the input images go through a set of residual blocks to extract shared feature representation. Then the segmentation map is divided into several paths with physical meaning attention to a certain category. Each segmentation map is expanded to the same size as the shared feature. These multi-patch shared features are point-wise multiplied with the expanded attention maps. Final, we fuse the divided feature to obtain derain image:

$$\hat{\boldsymbol{C}} = \sum_{c=1}^{C} F_c(\boldsymbol{S}_{D_c} \otimes F_d(\tilde{\boldsymbol{C}}, \boldsymbol{R})), \tag{5}$$

where F_d and F_c are the feature extraction operators, $\boldsymbol{S}_{D_c}$ is the semantic map of each category, $\otimes$ means the point-wise multiplication, and $\hat{\boldsymbol{C}}$ denotes the final deraining result. Each divided path in SAM can learn category-specific feature representations, so as to adaptive deraining. In fact, the SAM can be regarded as a refinement of the coarse deraining results. The final deraining is still the restoration loss:

$$\mathcal{L}_{res_f} = ||\hat{\boldsymbol{C}} - \boldsymbol{C}_{gt}||^2. \tag{6}$$

Implementation Details

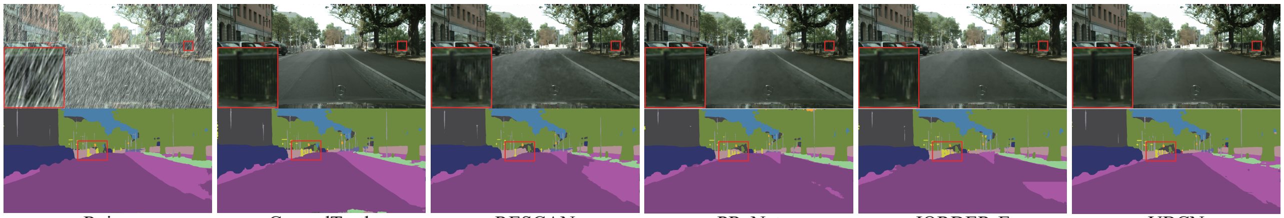
Thus, the overall loss of the proposed UBCN is

$$\mathcal{L}_{overall} = \mathcal{L}_{res_c} + \lambda\mathcal{L}_{adv} + \alpha\mathcal{L}_{CSeg} + \beta\mathcal{L}_{CSeg} + \gamma\mathcal{L}_{res_f}, \tag{7}$$

where $\lambda = 1e^{-3}$, $\alpha = 1$, $\beta = 1$, $\gamma = 10$ are the hyper-parameters. For different datasets, due to the different number of the category, the SAM would be slightly different. The total number of the iteration step our UBCN is set to be 3. As for network training, we adopt SGD with 0.9 momentum as network optimizer and first pre-train image-level derain subnetwork and feature-level adaptation segmentation subnetwork separately for 100 epochs with learning rate 1e-3 and 0.9 poly coefficient, and then fine-tune whole network for 50 epochs with an initial learning rate 1e-5. Random crop and mirror are utilized to perform data argumentation. We adopt SGD with 0.9 momentum as network optimizer.

Difference between UBCN with Existing Works

Here, we further clarify the differences between UBCN with existing works from following aspects. First of all, key ideas of UBCN are very different from other methods. For example, goal of PRRNet (Zhang et al. 2020) is image deraining only, with aid of semantic and stereo information for unidirectional promotion. UBCN is the first to consider synergy relationship to solve joint deraining and segmentation problem, where each subtask is of equal importance and benefits greatly from each other. As for overall network architecture, previous methods all adopt unidirectional pipeline for top-down segmentation→derain (Zhang et al. 2020) or bottom-up derain→segmentation (Wang and Zhang 2021). Instead, we propose the first unified bidirectional cooperation network, considering bidirectional promotion deraining↔segmentation in an iterative feedback manner. Lastly, the methodologies of UBCN are different from other methods. On one hand, previous deraining methods usually utilize semantics with simple concatenation (PRRNet), while semantic attentive module is proposed to adaptively



Rainy GroundTruth RESCAN PReNet JORDER-E UBCN

Figure 6: Visual comparison of rain removal and semantic segmentation results on *Cityscapes* dataset.

Methods	Low-level Metric		+ PSPNet		+ HRNet	
	PSNR	SSIM	mIoU	PA	mIoU	PA
Rain	24.43	0.5652	71.08	95.05	61.48	93.85
DDN	33.66	0.9006	72.60	95.42	62.26	93.68
RESCAN	40.33	0.9712	73.70	95.66	65.70	94.36
PReNet	42.13	0.9823	75.03	95.82	67.87	94.24
JORDER-E	42.33	0.9778	76.01	95.74	67.99	94.72
RCDNet	42.85	0.9810	76.13	95.83	68.08	94.71
UBCN	**43.22**	**0.9831**	**77.09**	**95.97**	**68.65**	**94.89**

Table 1: Quantitative comparison results on *Cityscapes*.

Methods	Low-level Metric		+ PSPNet	
	PSNR	SSIM	mIoU	PA
Rain	21.93	0.6617	80.23	95.29
DDN	26.17	0.7880	80.48	95.39
RESCAN	31.07	0.9006	82.31	95.93
PReNet	30.77	0.9100	82.77	96.07
JORDER-E	31.36	0.9124	82.80	96.06
RCDNet	31.03	0.9057	82.42	95.95
UBCN	**32.18**	**0.9173**	**84.13**	**96.45**

Table 2: Quantitative comparison results on *VOC2012*.

utilize semantic prior with physical meaning. On the other hand, the previous methods usually obtain segmentation from either derain image or rain-free feature domain (Wang and Zhang 2021), while we propose to obtain accurate semantic prediction from both image & feature domain via image-level and feature-level adaptation module.

Experiments

Experimental Settings

We test the proposed methods on two widely used datasets: *Cityscapes* (Cordts et al. 2016) and *VOC2012* (Everingham et al. 2015). We simulate rain via screen blend model (Luo, Xu, and Ji 2015). The state-of-the-arts segmentation models (Zhao et al. 2017; Wang et al. 2021) and deraining methods (Fu et al. 2017; Li et al. 2018; Ren et al. 2019; Yang et al. 2019; Wang et al. 2020a) are employed for comparison. They are combined to successively perform deraining and segmentation. *Note that all segmentation models are fine-tuned with corresponding deraining methods for fair comparison*. We employ PSNR and SSIM for deraining evaluation, and pixel-wise mean intersection over union (mIoU) and pixel accuracy (PA) for segmentation evaluation.

Methods	PSNR	SSIM	mIoU	PA
w/o SAM	41.81	0.9782	75.10	95.48
w/o ILD	—	—	71.08	95.05
w/o FLD	43.22	**0.9831**	76.81	95.93
UBCN	**43.23**	**0.9831**	**77.09**	**95.97**

Table 3: The effectiveness of each components in UBCN.

Step	PSNR	SSIM	mIoU	PA
1	38.86	0.9595	69.99	94.80
2	42.45	0.9789	73.98	95.71
3	**43.23**	**0.9831**	**77.09**	**95.97**

Table 4: Quantitative results of each iteration on *Cityscapes*.

Quantitative and Qualitative Evaluations

The quantitative results on *Cityscapes* and *VOC2012* are shown in Table 1 and Table 2, respectively. We can observe that UBCN consistently outperforms competing methods both on low-level restoration results and high-level segmentation index by a large margin. Both PSNR value and mIoU on different datasets have been significantly improved, which strongly supports effectiveness of the proposed unified bottom-up and top-down paradigm.

In Fig. 6, we show the visual deraining and segmentation results on *Cityscapes*. *Due to space limitation, more results are shown in the supplementary.* Although the competing methods can satisfactorily remove the rain, the proposed UBCN can better preserve the subtle image structures such as the fences and the wall of distant buildings. This phenomenon could validate the effectiveness of the semantic information serving as efficient prior. Moreover, the segmentation result of the UBCN is more structural and meaningful.

Ablation Study

The effectiveness of image- and feature-level deraining for segmentation. In bottom-up stream, the key is to get rid of negative effects of rain on segmentation. To demonstrate the effectiveness of both image- and feature-level deraining, in Table 3, we implement UBCN without image-level deraining (ILD) and feature-level deraining (FLD). Without ILD would cause a dramatic drop in segmentation (second row). Without FLD, the performance would also slightly decrease (third row). The proposed UBCN learns rain-robust representation from both image and feature domain, which

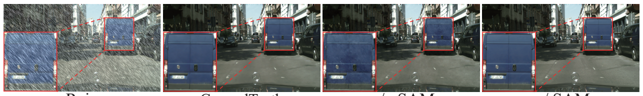


Figure 7: Visualization of deraining results w/ or w/o SAM.

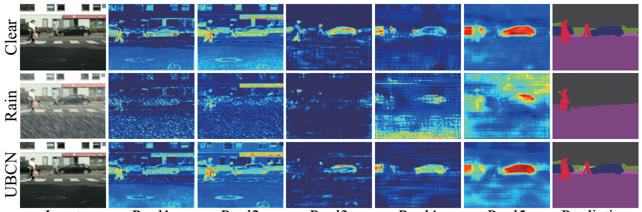


Figure 8: The effectiveness of UBCN to improve the semantic feature representation. We show the feature visualizations of clear, rainy, and UBCN in different layers, respectively.

would better facilitate segmentation.

The effectiveness of high-level semantics for deraining. In Table 3, we also conduct experiments to investigate the influence of the proposed SAM and semantic information for deraining. The proposed SAM greatly improves deraining performance with 1.42dB in PSNR. Moreover, the segmentation result is also improved in turn, due to the better deraining results. In Fig. 7, we visualize the deraining results with or without SAM. The result with SAM is visually pleasing with less rain residual and clearly sharp edges.

The effectiveness of unified bidirectional paradigm. To verify iterative cooperation and refinement strategy in our method, we analyze how deraining and segmentation results change in each iteration. In Table 4, we can observe the performance of both tasks getting better synchronously. Better deraining would alleviate segmentation difficulty; accurate semantic guidance would in turn facilitate deraining.

Discussion and Analysis

Visualization of the semantic features. To further verify the effectiveness of image- and feature-level deraining in facilitating segmentation, in Fig. 8, we visualize the intermediate features in the segmentation network. From the top to bottom row are the features extracted from clear, rain, and derain images by UBCN. For clean image, the extracted feature is discriminative. On the contrary, the activation from rainy image is quite noisy (Pool 1-2 layer) and less informative (Pool 3-5 layer), leading to obvious segmentation performance drop. The features extracted from UBCN are very similar to clear image features. This indicates that deraining in image and feature domain indeed alleviates degradation and promotes segmentation performance in rainy scenes.

Real-world image derain and segmentation. To further illustrate the generalization of UBCN in real-world rain conditions, we evaluate the performance of UBCN on the real-world rainy images. We collect several real rainy images on city road scene, which has less domain gap between the real images and the *Cityscapes*. The model trained on the *Cityscapes* dataset is directly employed for testing on the real rainy images. In Fig. 9, we show the real rainy image deraining and segmentation results along with the results of RCDNet. Interestingly, the UBCN can not only achieve satisfactory deraining results but also good segmentation performance compared with other methods. We believe that the unsupervised domain adaption would be a good choice for real-world rainy image segmentation in our future work.

Figure 9: Visualization of deraining and segmentation results on real-world rainy images.

Method	PSNR	SSIM	mIoU	PA
Image domain w/o FT	41.81	0.9782	75.10	95.48
Image domain w/ FT	42.10	0.9794	76.07	95.83
Feature domain adaption	—	—	72.64	95.38
UBCN	**43.22**	**0.9831**	**76.81**	**95.93**

Table 5: Different strategies for rainy image segmentation.

The advantage of unified bidirectional paradigm. There are several techniques introduced in Fig. 2 for rainy image segmentation. Here, we compare the typical methods in Table 5. We choose the bottom-up methods Liu *et al.* (Liu et al. 2020) (image domain without fine-tune first row, with fine-tune second row), feature domain adaptation methods (Dai et al. 2020) (second row), and the proposed method UBCN. We can observe that the unsupervised feature domain adaptation method is difficult to get rid of rain degradation effect. The supervised image domain deraining strategies could offer better deraining and segmentation results. Compared with these unidirectional methods, our unified bidirectional UBCN allows interaction between two tasks in each iterative step cooperatively with better performance.

Conclusion

In this work, we aim at a very practical problem rainy image segmentation. We first provide a comparison between existing unidirectional methods and the proposed bidirectional paradigm. Then, we analyze how rain degradation influences segmentation and how semantic information facilitates deraining. Based on these analysis, we propose a unified bidirectional cooperation network within a unified bottom-up and top-down paradigm. On one hand, image and feature domain adaption scheme is presented for better segmentation in bottom-up stream; on the other hand, a novel SAM is designed for physical meaning adaptive deraining in top-down stream. The two streams are iteratively performed with mutual promotion. Extensive experiments verify the superiority of UBCN in both image deraining and segmentation.

Acknowledgement

This work was supported in part by National Natural Science Foundation of China under Grant 61971460, in part by JCJQ Program under Grant 2021-JCJQ-JJ-0060, in part by China Postdoctoral Science Foundation under Grant 2020M672748, and in part by National Postdoctoral Program for Innovative Talent under Grant BX20200173.

References

Chang, Y.; Yan, L.; and Zhong, S. 2017. Transformed low-rank model for line pattern noise removal. In *Int. Conf. Comput. Vis.*, 1726–1734.

Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Eur. Conf. Comput. Vis.*, 801–818.

Chen, Y.; and Hsu, C. 2013. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Int. Conf. Comput. Vis.*, 1968–1975.

Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; and Schiele, B. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3213–3223.

Dai, D.; Sakaridis, C.; Hecker, S.; and Van Gool, L. 2020. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *Int. J. Comput. Vis.*, 128(5): 1182–1204.

Deng, S.; Wei, M.; Wang, J.; Feng, Y.; Liang, L.; Xie, H.; Wang, F.; and Wang, M. 2020. Detail-recovery Image Deraining via Context Aggregation Networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 14560–14569.

Everingham, M.; Eslami, S.; Van Gool, L.; Williams, C.; Winn, J.; and Zisserman, A. 2015. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.*, 111(1): 98–136.

Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017. Removing rain from single images via a deep detail network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3855–3863.

Fu, X.; Liang, B.; Huang, Y.; Ding, X.; and Paisley, J. 2020. Lightweight Pyramid Networks for Image Deraining. *IEEE Trans. Neural Netw. Learn. Syst.*, 31(6): 1794–1807.

Guo, M.; Chen, M.; Ma, C.; Li, Y.; Li, X.; and Xie, X. 2020. High-Level Task-Driven Single Image Deraining: Segmentation in Rainy Days. In *Int. Conf. Neural Inf. Process.*, 350–362.

Halder, S.; Lalonde, J.; and Charette, R. 2019. Physics-based rendering for improving robustness to rain. In *Int. Conf. Comput. Vis.*, 10203–10212.

Hu, X.; Fu, C.; Zhu, L.; and Heng, P. 2019. Depth-attentional features for single-image rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 8022–8031.

Huang, S.; Le, T.; and Jaw, D. 2021. DSNet: Joint Semantic Learning for Object Detection in Inclement Weather Conditions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(8): 2623–2633.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 8346–8355.

Kamann, C.; and Rother, C. 2020. Benchmarking the robustness of semantic segmentation models. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 8828–8838.

Kang, L.; Lin, C.; and Fu, Y. 2011. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4): 1742–1755.

Li, R.; Cheong, L.; and Tan, R. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 1633–1642.

Li, S.; Araujo, I.; Ren, W.; Wang, Z.; Tokuda, E.; Junior, R.; Cesar-Junior, R.; Zhang, J.; Guo, X.; and Cao, X. 2019. Single image deraining: A comprehensive benchmark analysis. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3838–3847.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Eur. Conf. Comput. Vis.*, 254–269.

Li, Y.; Tan, R.; Guo, X.; Lu, J.; and Brown, M. 2016. Rain streak removal using layer priors. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2736–2744.

Liu, D.; Wen, B.; Jiao, J.; Liu, X.; Wang, Z.; and Huang, T. 2020. Connecting image denoising and high-level vision tasks via deep learning. *IEEE Trans. Image Process.*, 29: 3695–3706.

Luo, Y.; Xu, Y.; and Ji, H. 2015. Removing rain from a single image via discriminative sparse coding. In *Int. Conf. Comput. Vis.*, 3397–3405.

Ma, C.; Jiang, Z.; Rao, Y.; Lu, J.; and Zhou, J. 2020. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 5569–5578.

Pei, Y.; Huang, Y.; Zou, Q.; Zhang, X.; and Wang, S. 2021. Effects of Image Degradation and Degradation Removal to CNN-Based Image Classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(4): 1239–1253.

Porav, H.; Bruls, T.; and Newman, P. 2019. I can see clearly now: Image restoration via de-raining. In *Int. Conf. Robot. Autom.*, 7087–7093.

Qian, R.; Tan, R. T.; Yang, W.; Su, J.; and Liu, J. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2482–2491.

Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; and Meng, D. 2019. Progressive image deraining networks: A better and simpler baseline. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3937–3946.

Ren, W.; Zhang, J.; Xu, X.; Ma, L.; Cao, X.; Meng, G.; and Liu, W. 2018. Deep video dehazing with semantic segmentation. *IEEE Trans. Image Process.*, 28(4): 1895–1908.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Semantic foggy scene understanding with synthetic data. *Int. J. Comput. Vis.*, 126(9): 973–992.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2020. Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*

Scheirer, W.; VidalMata, R.; Banerjee, S.; RichardWebster, B.; Albright, M.; Davalos, P.; McCloskey, S.; Miller, B.; Tambo, A.; Ghosh, S.; Nagesh, S.; Yuan, Y.; Hu, Y.; Wu, J.; Yang, W.; Zhang, X.; Liu, J.; Wang, Z.; Chen, H. T.; Huang, T. W.; Chin, W. C.; Li, Y. C.; Lababidi, M.; and Otto, C. 2020. Bridging the Gap Between Computational Photography and Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*

Sharma, V.; Diba, A.; Neven, D.; Brown, M.; Van Gool, L.; and Stiefelhagen, R. 2018. Classification-driven dynamic image enhancement. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 4033–4041.

Shen, Z.; Lai, W.; Xu, T.; Kautz, J.; and Yang, M. 2020. Exploiting semantics for face image deblurring. *Int. J. Comput. Vis.*, 128(7): 1829–1846.

Sun, K.; Wu, W.; Liu, T.; Yang, S.; Wang, Q.; Zhou, Q.; Ye, Z.; and Qian, C. 2019. Fab: A robust facial landmark detection framework for motion-blurred videos. In *Int. Conf. Comput. Vis.*, 5462–5471.

Valada, A.; Vertens, J.; Dhall, A.; and Burgard, W. 2017. Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In *Int. Conf. Robot. Autom.*, 4644–4651.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(11): 2579–2605.

Wang, F.; and Zhang, Y. 2021. A De-raining semantic segmentation network for real-time foreground segmentation. *J. Real-Time Image Process.*, 18(3): 873–887.

Wang, H.; Xie, Q.; Zhao, Q.; and Meng, D. 2020a. A Model-driven Deep Neural Network for Single Image Rain Removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3103–3112.

Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; Liu, W.; and Xiao, B. 2021. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(10): 3349–3364.

Wang, L.; Li, D.; Zhu, Y.; and Shan, Y. 2020b. Dual Super-Resolution Learning for Semantic Segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3774–3783.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 12270–12279.

Wang, Z.; Chang, S.; Yang, Y.; Liu, D.; and Huang, T. 2016. Studying very low resolution recognition using deep networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 4792–4800.

Xu, K.; Tian, X.; Yang, X.; Yin, B.; and Lau, R. 2021. Intensity-Aware Single-Image Deraining With Semantic and Color Regularization. *IEEE Trans. Image Process.*, 30: 8497–8509.

Yang, M.; Yu, K.; Zhang, C.; Li, Z.; and Yang, K. 2018. Denseaspp for semantic segmentation in street scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 3684–3692.

Yang, W.; Tan, R. T.; Feng, J.; Guo, Z.; Yan, S.; and Liu, J. 2019. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(6): 1377–1393.

Yang, W.; Wang, S.; Xu, D.; Wang, X.; and Liu, J. 2020. Towards scale-free rain streak removal via self-supervised fractal band learning. In *AAAI Conf. Artif. Intell.*, volume 34, 12629–12636.

Yasarla, R.; and Patel, V. 2019. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 8405–8414.

Yu, C.; Wang, J.; Gao, C.; Yu, G.; Shen, C.; and Sang, N. 2020. Context prior for scene segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 12416–12425.

Zhang, H.; Dana, K.; Shi, J.; Zhang, Z.; Wang, X.; Tyagi, A.; and Agrawal, A. 2018. Context encoding for semantic segmentation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 7151–7160.

Zhang, H.; and Patel, V. 2018. Density-aware single image deraining using a multi-stream dense network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 695–704.

Zhang, H.; Yang, J.; Zhang, Y.; Nasrabadi, N.; and Huang, T. 2011. Close the loop: Joint blind image restoration and recognition with sparse representation prior. In *Int. Conf. Comput. Vis.*, 770–777.

Zhang, K.; Li, R.; Yu, Y.; Luo, W.; and Li, C. 2021. Deep Dense Multi-Scale Network for Snow Removal Using Semantic and Depth Priors. *IEEE Trans. Image Process.*, 30: 7419–7431.

Zhang, K.; Luo, W.; Ren, W.; Wang, J.; Zhao, F.; Ma, L.; and Li, H. 2020. Beyond Monocular Deraining: Stereo Image Deraining via Semantic Understanding. In *Eur. Conf. Comput. Vis.*, 71–89.

Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J. 2017. Pyramid scene parsing network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2881–2890.

Zhu, H.; Peng, X.; Zhou, J.; Yang, S.; Chanderasekh, V.; Li, L.; and Lim, J. 2019. Singe image rain removal with unpaired information: A differentiable programming perspective. In *AAAI Conf. Artif. Intell.*, volume 33, 9332–9339.

Zhu, L.; Fu, C.; Lischinski, D.; and Heng, P. 2017. Joint bi-layer optimization for single-image rain streak removal. In *Int. Conf. Comput. Vis.*, 2526–2534.