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

Yi Li¹, Yi Chang¹²*, Changfeng Yu¹, Luxin Yan¹

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. 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
Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
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-down 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
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.

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.

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?**
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 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 de-raining. 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.
specifically, explicit deraining is first performed via image-tation network in bottom-up stream, as shown in Fig. 5(a). Feature-domain adaptation method could well extract rainy-feature may inevitably be damaged. On the contrary, the feature-domain deraining could better visually remove the rain, the artifacts either in the image domain or the feature domain so effect in segmentation. Previous methods suppressed the rain 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:

\[ \mathcal{L}_{res} = ||\hat{C} - C_{gt}||^2, \]  

where \( \hat{C} \) is the estimated derain image, \( C_{gt} \) 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 \( \hat{C} \) and clean image \( C_{gt} \) are also indistinguishable in the feature domain with adversarial loss:

\[ \mathcal{L}_{adv} = \log(1 - D(G_{de}(\hat{C}))) + \log(D(G_{ce}(C_{gt}))), \]  

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:

\[ \mathcal{L}_{C\text{Seg}} = -\sum_c S_{gt}^{(c)} \log(S_{\hat{C}}^{(c)}), \]  

\[ \mathcal{L}_{DSeg} = -\sum_c S_{gt}^{(c)} \log(S_{\hat{C}}^{(c)}), \]  

where \( c \) is the number of class, \( S_{D}, S_{C}, S_{gt} \) 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:

\[ \hat{C} = \sum_{c=1}^C F_d(S_{D_{c}} \odot F_d(\hat{C}, R)), \]  

where \( F_d \) and \( F_c \) are the feature extraction operators, \( S_{D_{c}} \) is the semantic map of each category, \( \odot \) means the point-wise multiplication, and \( \hat{C} \) 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:

\[ \mathcal{L}_{rest} = ||\hat{C} - C_{gt}||^2. \]  

Implementation Details

Thus, the overall loss of the proposed UBCN is

\[ L_{overall} = L_{res} + \lambda L_{adv} + \alpha L_{C\text{Seg}} + \beta L_{DSeg} + \gamma L_{rest}, \]  

where \( \lambda = 1e^{-3}, \alpha = 1, \beta = 1, \gamma = 10 \) are the hyperparameters. 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 augmentation. 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
Figure 6: Visual comparison of rain removal and semantic segmentation results on Cityscapes dataset.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Low-level Metric</th>
<th>+ PSPNet</th>
<th>+ HRNet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>SSIM</td>
<td>mIoU</td>
</tr>
<tr>
<td>Rain</td>
<td>24.43</td>
<td>0.5652</td>
<td>71.08</td>
</tr>
<tr>
<td>DDN</td>
<td>33.66</td>
<td>0.9006</td>
<td>72.60</td>
</tr>
<tr>
<td>RESCAN</td>
<td>40.33</td>
<td>0.9712</td>
<td>73.70</td>
</tr>
<tr>
<td>PReNet</td>
<td>42.13</td>
<td>0.9823</td>
<td>75.03</td>
</tr>
<tr>
<td>JORDER-E</td>
<td>42.33</td>
<td>0.9778</td>
<td>76.01</td>
</tr>
<tr>
<td>RCDNet</td>
<td>42.85</td>
<td>0.9810</td>
<td>76.13</td>
</tr>
<tr>
<td>UBCN</td>
<td>43.22</td>
<td>0.9831</td>
<td>77.09</td>
</tr>
</tbody>
</table>

Table 3: The effectiveness of each components in UBCN.

<table>
<thead>
<tr>
<th>Step</th>
<th>PSNR</th>
<th>SSIM</th>
<th>mIoU</th>
<th>PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.86</td>
<td>0.9595</td>
<td>80.23</td>
<td>95.29</td>
</tr>
<tr>
<td>2</td>
<td>42.45</td>
<td>0.9789</td>
<td>82.31</td>
<td>95.93</td>
</tr>
<tr>
<td>3</td>
<td>43.23</td>
<td>0.9831</td>
<td>84.13</td>
<td>96.45</td>
</tr>
</tbody>
</table>

Table 4: Quantitative results of each iteration on Cityscapes.

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.

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
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.

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. (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 JCQ Program under Grant 2021-JCQ-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


