High-Level Task-Driven Single Image Deraining: Segmentation in Rainy Days

Mengxi Guo[1], Mingtao Chen[1], Cong Ma[2], Yuan Li[2], Xianfeng Li[1], and Xiaodong Xie[2](✉)

[1] School of Electronic and Computer Engineering, Peking University, Shenzhen 518055, China
{nicolasguo,mingtaochen,lixianfeng.sz}@pku.edu.cn
[2] School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
{Cong-Reeshard.Ma,yuanli,donxie}@pku.edu.cn

Abstract. Deraining driven by semantic segmentation task is very important for autonomous driving because rain streaks and raindrops on the car window will seriously degrade the segmentation accuracy. As a pre-processing step of semantic segmentation network, a deraining network should be capable of not only removing rain in images but also preserving semantic-aware details of derained images. However, most of the state-of-the-art deraining approaches are only optimized for high PSNR and SSIM metrics without considering objective effect for high-level vision tasks. Not only that, there is no suitable dataset for such tasks. In this paper, we first design a new deraining network that contains a semantic refinement residual network (SRRN) and a novel two-stage segmentation aware joint training method. Precisely, our training method is composed of the traditional deraining training and the semantic refinement joint training. Hence, we synthesize a new segmentation-annotated rain dataset called Raindrop-Cityscapes with rain streaks and raindrops which makes it possible to test deraining and segmentation results jointly. Our experiments on our synthetic dataset and real-world dataset show the effectiveness of our approach, which outperforms state-of-the-art methods and achieves visually better reconstruction results and sufficiently good performance on semantic segmentation task.

Keywords: Single image deraining · Semantic segmentation · High-level task driven application · Deep learning

1 Introduction

Semantic segmentation is a key task in automatic driving. However, complicated rain weather can dramatically affect the robustness of the segmentation. In recent years, some efforts concentrate on heightening accuracy of the segmentation system in rainy days [24,29]. All their methods are training a deraining

H. Yang et al. (Eds.): ICONIP 2020, LNCS 12532, pp. 350–362, 2020.
https://doi.org/10.1007/978-3-030-63830-6_30

network independently and use it as a pre-processing module before the semantic segmentation network to improve the performance of segmentation systems affected by rain.

However, this kind of approach is not a perfect solution, because the separate deraining models are inclined to pursue higher image-quality metrics [11], but seldom systematically considered whether this is helpful for advanced tasks. Some works [8,18,22] demonstrated that only consider image-quality metrics con not help with the high-level tasks. That is, there is a gap between the low-level and high-level tasks. Although some efforts [5,8,17,20,30] have been made to close the gap through joint training, we observe that there are still some deficiencies: (1) A dataset with rainy-clean pair images and the annotations of high-level task is not enough. (2) The traditional way to close the gap is the one-stage joint training for the low-level and high-level vision tasks networks. However, it often leads to the PSNR/SSIM degradation when getting good performance for high-level tasks [8,17,20,30].

To address above limitations, firstly, we focus on the single image-based deraining of the urban road scenes and synthesize a high resolution rainy dataset based on Cityscapes [2] to simulate the real-world autonomous driving in rainy day. In particular, different from the datasets of [12,21,24] which only contain rain streaks or raindrop, our dataset simulates a complete process of rainy day with rain streaks of different densities and raindrops of different sizes. This makes our dataset more real and more challenging. And this dataset has semantic segmentation labels, which is convenient for us to explore the combination of deraining task and segmentation task simultaneously.

Secondly, in order to better integrate the deraining with the high-level task, we propose a novel high-level task-driven framework consisting of the low-level deraining network, the semantic refinement network and the high-level segmentation network, which follows a two-stage training process. In particular, at the Stage I, the low-level deraining network is trained to remove rain roughly. Subsequently, at the Stage II, the parameters of the previous deraining network are fixed, but the semantic refinement network trains with the high-level vision task network jointly to gets rid of residual rain while retaining high-level semantic information within the images. To achieve this goal, we present a semantic refinement residual network (SRRN) which repairs the semantic information of the image while keeping its vision quality. Our experiments show that compared to traditional one-stage joint training, our proposed model with the two-stage training method, as a bridge of deraining and segmentation task, can balance the two to enhance the performance of segmentation while maintaining relatively high image-quality metrics and achieve better results on both two tasks.

The contributions of this paper are as follows:

1. We create a large-scale synthetic rainy dataset of road scenes for training and evaluation, which contains both rain streaks and raindrops with well segmentation-annotation. It is worth mentioning that this dataset simulates continuous changes in a rainy process and covers more complex rainy conditions.

2. We propose a novel semantic segmentation task-driven deraining framework with two-stage segmentation aware joint training, in which the semantic refinement residual network (**SRRN**) is the bridge to balance the performance of low-level deraining and high-level segmentation. Experimental results show our deraining framework can contribute to segmentation dramatically while achieving better image quality than the one-stage joint training.
3. We conduct a benchmark test and analysis by evaluating our proposed deraining framework and state-of-the-art methods on our synthetic rainy dataset and the other open real-world rainy dataset. The experimental results show that our method quantitatively and qualitatively outperforms existing works on both visual performance and segmentation accuracy.

2 Related Work

Rain Streak and Raindrop Removal. Recently, many researches focus on single image rain streak removal by adopting image priors [15,19] or deep learning approaches [6,27]. Unfortunately, rain streaks have quite different characteristics from raindrops, the state-of-the-art rain streak removal methods cannot be used for raindrop removal directly [18]. Most traditional de-raindrop methods exploit physical characteristics and context information of multiple frames [16,32]. As for a deep learning based method, Eigen et al. [4] first separate raindrops via a simple CNN. Popular image-to-image translation methods like Pix2pix [13] are used for reference in learning raindrop-free images [24,25]. Other mainstream methods are inclined to combine their models with attention mechanism (Attentive-GAN [25], shape and channel attention [26], depth-guided attention [12]). Furthermore, Liu et al. [21] proposes a versatile Dual Residual Network, which can be applied to both rain-streak and raindrop removal tasks.

Bridging Low-Level and High-Level Vision Task. Lately, Some researches begin to combine low-level task with high-level vision task. [5,9,31] fuse high-level semantic information into the low-level network to help it to obtain better visual effects. [17,18,22,24] utilize the low-level network as an independent preprocessor before the high-level network. However, [18,22] show that, in this independent pipeline, the output of the low-level network will degrade the performance of subsequent high-level vision system. Therefore, some efforts have been made to close the gap between them using an end-to-end joint training method. Liu et al. [20] train a denoising model by cascading a fixed pre-trained segmentation network and tune with a joint loss function of segmentation awareness and MSE loss. Wang et al. [30] propose an unsupervised segmentation-aware denoising network using joint training without needing segmentation labels. Similar high-level application-driven method [8,17] have been proposed in the fields of super-resolution and dehazing. However, to the best of our knowledge, there is no related deraining method using our proposed two-stage semantic refinement joint training so far.

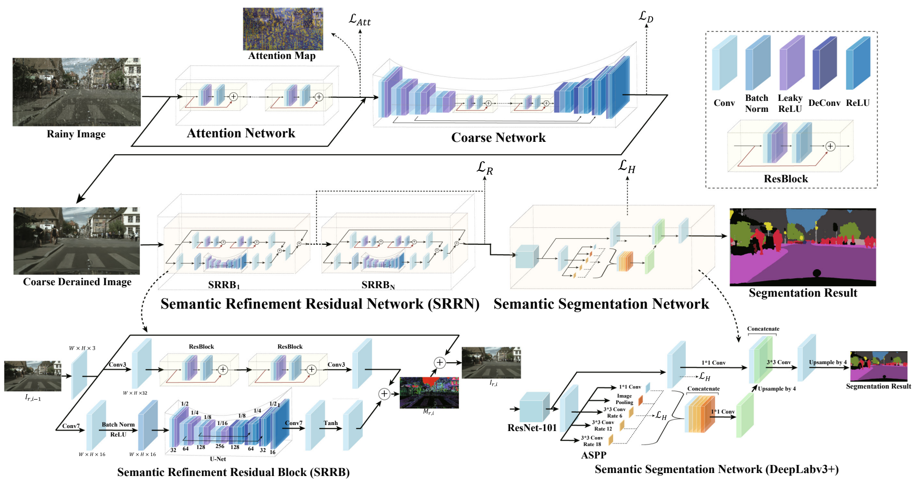

Fig. 1. The detail architectures of our networks.

3 Method

3.1 Model Architecture

As shown in Fig. 1, the architecture of our cascaded network consists of three parts: the deraining network, semantic refinement network and the semantic segmentation network for high-level vision task. The deraining network is first applied to generate the coarse derained image. And then the derained image is fed into the semantic refinement network to further repair the semantic information. Finally, the network for semantic segmentation gets the refined derained image as input and generates the segmentation result.

Deraining Network: The detailed structure of the whole deraining network can be seen in Fig. 1. It consists of two networks: (1) the attention Network **A** takes a rainy image I_r as input and generates an attentive map M which could localize the regions with rain streaks and raindrops in an image; (2) the coarse Network **C** is responsible for the preliminary rain removal of the rainy image. It takes M and I_r as input and generates a coarse derained image I_C.

Attention Network. We use a full-resolution ResNet [23] as the attention network to generate a binary attention map M. As the supervisory information, the binary attention map, M_{gt}, is simply generated by subtracting the gray-scale rainy image $I_{r,g}$ with the gray-scale clean image$I_{G,g}$. We set the threshold as 5 to determine whether a pixel is rainy region. The loss function is defined as the mean squared error (MSE) between M and M_{gt}:

$$\mathcal{L}_{Att} = MSE(M, M_{gt}) = MSE(A(I_r), M_{gt}) \tag{1}$$

Coarse Network. The role of the coarse network is to perform single image deraining. The input is the concat of the rainy image I_r and the attention map M generated by the attention network. The output is a coarse derained image I_C. We define the loss function as:

$$\mathcal{L}_D(I_C, I_{gt}) = \alpha(1 - SSIM(I_C, I_{gt})) + (1 - \alpha) \left\| I_C - I_{gt} \right\|_1 \tag{2}$$

where α is the weight for balancing the two losses, and we set α as 0.85.

Semantic Refinement Network. The semantic refinement Network **R** takes I_C as the input and conducts joint training with semantic segmentation network to generate the final derained images I_R with more semantic information. The purpose of our refinement network is to further improve the quality while restoring the high-level semantic-aware information of the derained image to make it better adapted to the subsequent task. For this, we propose the semantic Refinement residual block (**SRRB**) and use it to construct our refinement network. As shown in Fig. 1, the whole SRRB is composed of two parallel branches: One U-Net structure branch and one full-resolution convolutional branch. It allows the network to acquire semantic information from a large receptive field. In order to make the output of SRRB not only as much as possible keep the pixel-level detail, but also further improve the semantic-aware information of the local pixels, we use three tips: (1) We use skip connections in the U-Net structure, which has been proved to be helpful in keeping most of the details of the output images; (2) We add a parallel branch, which is composed of two full-resolution convolutional layers and two Res-blocks. Such a branch can maintain the image details and obtain the local semantic information; (3) The outputs of the two branches of SRRB are added to generate a 3-channel residual semantic map M_r with rich semantic and pixel-level restoring information.

We connect multiple SRRBs to make up our semantic refinement residual network (**SRRN**). SRRB can accomplish local and global information restoring. However, we find that one SRRB can not be enough for restoring well. There are two purposes we connect multiple SRRBs: first, in the refinement stage, the superposition of the blocks can maintain a certain depth of the network, which ensure the quality of the images. Second, what we find is that if only one SRRB is used for the refinement network, it is still difficult to achieve better consistency between the quality of images and the performance of high-level tasks. Therefore, we use N blocks to form the SRRN and use a step process to generate the final output. In the SRRN, the i-th SRRB takes the output of $i-1$-th SRRB $I_{R,i-1}$ as the input and generates the more refined image $I_{R,i}$. When $i = 0$, the input of SRRB is the output of coarse network I_C. The reconstruction step loss of the SRRN has two parts: One is the sum of mean squared error between the outputs of the $N-1$ SRRBs and the clean image. The other is a L_D loss between the output of the N-th SRRB and the clean image. It can be expressed as:

$$\mathcal{L}_R = 0.5 \sum_{i=0}^{N-1} (I_{R,i-1} - I_{gt})^2 + \mathcal{L}_D(I_{R,N}, I_{gt}) \tag{3}$$

Finally, we set N = 2 because what we find in our experiment is that when N > 2, the performance doesn't increase significantly.

Semantic Segmentation Network: The detailed structure of the semantic segmentation network can be seen in Fig. 1. We use DeepLabv3+ [1] as our semantic segmentation network and use Resnet-101 [10] as a backbone. The network is well trained in the rainless settings and the weights of the network are fixed while training the deraining network. The Astrous spatial pyramid pooling (ASPP) which contains several parallel atrous convolution kernels with different ratios is used to extract semantic information at different scales in DeepLabv3+. To some extent, the feature maps from ASPP have multi-scale high-level semantic information and the feature maps generated by the last layer of the backbone contain the low-level semantic information. We regard these feature maps as important semantic supervised information and use them in our training process. The loss of the semantic segmentation network can be defined as:

$$\mathcal{L}_H = \frac{1}{N_{AF}} \sum_{i=1}^{N_{AF}} \left\| SS(I_{gt})_{AFi} - SS(I_{R,N})_{AFi} \right\|_1 + \lambda_{LF} \left\| SS(I_{gt})_{LF} - SS(I_{R,N})_{LF} \right\|_1 \tag{4}$$

where SS is the semantic segmentation network. $SS(I)_{AF}$ means the feature maps generated from ASPP. N_{AF} is the number of parallel channels of ASPP and it is 5 in this model. $SS(I)_{LF}$ represents the feature maps from the last layer of the backbone. λ_{LF} is the weight for balancing low-level and high-level semantic losses and we set it as 0.1 in our experiments.

It is worth mentioning that, unlike the previous methods [5,20], L_H can be calculated without the actual need of any segmentation ground truth during the training process which is similar to minimizing the perceptual loss [14] for image super-resolution [3,7].

3.2 Training Method

For training our cascaded network, we design a brand new corresponding two-stage segmentation aware joint training method.

Stage I: Traditional Deraining Training. At this training stage, we use a normal deraining training method to train our attentive network and coarse network. We use the end-to-end training method and the pixel-level image processing loss functions to adjust the parameters of our network. The loss function in this stage is expressed as:

$$\mathcal{L}_1 = \mathcal{L}_{Att} + \mathcal{L}_D \tag{5}$$

Stage II: Semantic Refinement Joint Training. After the stage I training process, we leave the parameters of the attention network and the coarse network fixed. On the basis of stage I, stage II cascades SRRN for refining the coarse derained image and the semantic segmentation network for the high-level vision task, aiming to simultaneously further reconstruct visually pleasing results of deraining and attain sufficiently good performance in semantic segmentation task. The training at this stage is achieved by minimizing the joint loss function and updating the parameters of SRRN:

$$\mathcal{L}_2 = \mathcal{L}_R + \mathcal{L}_H \tag{6}$$

4 A New Raindrop-Cityscapes Dataset

It is almost impossible to capture a pair of photos with and without rain in real life. Many researches [12,24,25,33] try to construct rainy datasets through simulation methods. However, the existing datasets are not good enough to simulate the complex rain streaks and raindrops in real life. Moreover, except [12,18,24], there is no open rainy dataset with complete annotation information of high-level vision task which makes it difficult to test cascaded high-level network performance. We model a rain streak and raindrop degraded image as the combination of a clean image and effect of the raindrops and a rain streaks layer:

$$I_r = (1 - M) \odot I_{gt} + R + S \tag{7}$$

where I_r is the synthetic rain image, I_{gt} is the clean image, M is the binary rain map, R is the effect brought by the raindrops and S is the rain streak layer.

In this paper, we aim at a task-driven rain removal problem in road scenes, therefore, based on Cityscapes [2], we use the synthetic model in [24,28] to generate raindrops and use model in [33] to generate rain streaks in images. On the basis, we add the rain streak layer which randomly changes with each timestep. It is worth mentioning that, unlike other synthesis rainy dataset, our dataset contains a complete process of rainy change from no rain to gradual occurrence of rain streaks, to accumulation of small raindrops on the car window and slow generation for large raindrops which finally flows down. This allows our dataset to cover more rainy day situations, making it more complex and challenging. Altogether, our Raindrop-Cityscapes dataset has 3,479 high-resolution (1024×2048) images that cover various rainy scenes.

5 Experiments

Datasets. We use our Raindrop-Cityscapes as a synthetic dataset and RobotCar(R) [24] as a real-world dataset in our experiment. Specially. The Raindrop-Cityscapes contains 3,479 images and we randomly select 479 images as the testing dataset and the remaining 3,000 images as the training dataset.

Semantic Segmentation Test. For semantic segmentation test, following the evaluation protocol of Cityscapes [2], 19 output of 30 semantic labels are used for evaluation. The final performance is measured in terms of pixel intersection-over-union (mIoU) averaged across these classes.

Settings. The proposed network is implemented by PyTorch and executed on four NVIDIA TITAN X GPUs. In our experiments, the semantic segmentation network-DeepLabv3+ is trained in their default setting. The two-stage training for the deraining network and semantic segmentation network share the same training setting: the patch size is 512×512, and the batch size is 6 for the limitations by the GPU memory. For gradient descent, Adam is applied with an initial learning rate of 1×10^{-3}, and ends after 100 epochs. When reaching 30, 50 and 80 epochs, the learning rate is decayed by multiplying 0.2.

5.1 Ablation Studies

All the ablation studies are conducted on our Raindrop-Cityscapes dataset. We discuss the effects of the refinement network and the two-stage joint training method. As shown in Table 1, A/C means we only use attention network and coarse network to do deraining and use traditional deraining training as Eq. 5. A/C/R means adding the semantic refinement network R to A/C and training same as A/C. In addition to discussing the role of the refinement network, we also test the effect of high-level task aware jointing training. We abbreviate the traditional one-stage joint training [20] and our two-stage joint training to OJT and TJT in the experiments section. It should be noted that because R is not introduced into A/C, the joint training of A/C can only be trained with OJT.

Ablation for the Semantic Refinement Network. The comparisons between A/C and A/C/R, and between A/C + OJT and A/C/R + OJT show the importance of the refinement network. As shown in Table 1, the models with R get higher PSNR and SSIM which means that the semantic refinement network indeed has a function of further optimizing the quality of the derained image. Moreover, the higher mIoU shows that the semantic refinement network strengthens the link between deraining and semantic segmentation tasks.

Ablation for Joint Training. Table 1 lists the quantitative evaluations of these four different tests. Unsurprisingly, A/C + OJT, A/C/R + OJT, A/C/R + TJT outperform A/C and A/C/R in terms of the performance of derained images in semantic segmentation task. However, PSNR/SSIM of OJT group are lower than these without joint training. Particularly, A/C + OJT are significantly lower than A/C. Similar phenomenon also occurs in another works [8,17,20,30]. This shows the disadvantage of OJT. We think there are two reasons for this phenomenon: (1) PSNR/SSIM depend on the reconstruction error between the derained output and the clean image. Compared with the traditional image optimized loss function which always is MSE loss or L1 loss, the introduction of joint loss from the cascaded high-level task will make it difficult for the single deraining network to reach the maximum PSNR/SSIM with a one-stage training process. (2) A single small deraining network limits the optimal performance of the network in both tasks. The comparison between A/C/R, A/C/R + OJT and A/C/R + TJT can support our first explanation to some extent, because compared to A/C/R + OJT, A/C/R + TJT shows slight PSNR decrease while mIoU increase which shows the advantage of our TJT. And from the higher

PSNR and SSIM of A/C/R + OJT than A/C + OJT, we can see that increasing the number of network parameters can make up for this shortcoming of OJT, but it is still not as good as TJT.

Table 1. The results of the ablation studies.

Method	Rainy image	A/C	A/C+OJT	A/C/R	A/C/R+OJT	A/C/R+TJT
PSNR	23.23	33.88	28.84	34.61	31.26	34.57
SSIM	0.8002	0.9573	0.9324	0.9581	0.9521	0.9583
mIoU(%)	34.50	66.60	67.54	66.79	67.50	67.60

5.2 Evaluation on Synthetic Dataset

Our proposed network is evaluated on our Raindrop-Cityscapes datasets. Four competing methods: AGAN [25], ICSC [24], PReNet [27] and DuRN [21] are considered. Our model is the whole network with attention network, coarse network and SRRN. As we can see the top table in Fig. 2, our network not only achieves significant PSNR and SSIM gains over all the competing methods but also shows the best performance in the successor semantic segmentation task. As can be seen from Fig. 2, our method has fewer residual traces on the derained image. In terms of semantic segmentation results, our network, especially trained with semantic refinement joint training method, retains more complete semantic information. We also test the semantic segmentation performance through training segmentation network directly, but the performance is not good as ours.

5.3 Evaluation on the Real-World Dataset

Since we focus on a rain removal problem of the urban street scene driven by segmentation task, We use RobotCar(R) dataset [24], which is a real-world raindrop dataset of road scenes constructed by using a double-lens stereo camera mounted on a vehicle. However, this dataset doesn't have strictly annotated semantic segmentation labels. Therefore, we use it in well-trained semantic segmentation network on Cityscapes [2] to generate the loosely annotated ground truth on the rainless images of RobotCar(R), and we pick 500 images manually as the dataset based on intuition. Of these, 352 images are the training dataset and 148 images are the testing dataset. The competing methods are similar to those in Sect. 5.2 and our model is the whole networks trained by our two-stage training method. The top table in Fig. 3 presents the qualitative results of these five methods. It is seen that our method achieves the best result. As shown in Fig. 3, the visual quality improvement by our method is also significant in this dataset and the semantic segmentation performance achieves the best among the five methods.

Method	Rainy image	AGAN	PReNet	DuRN	ICSC	Ours	Our+TJT
PSNR	23.23	30.17	31.36	31.88	30.25	34.61	34.57
SSIM	0.8002	0.9520	0.9542	0.9563	0.9406	0.9581	0.9583
mIoU(%)	34.50	66.75	65.25	66.37	62.76	66.79	67.60

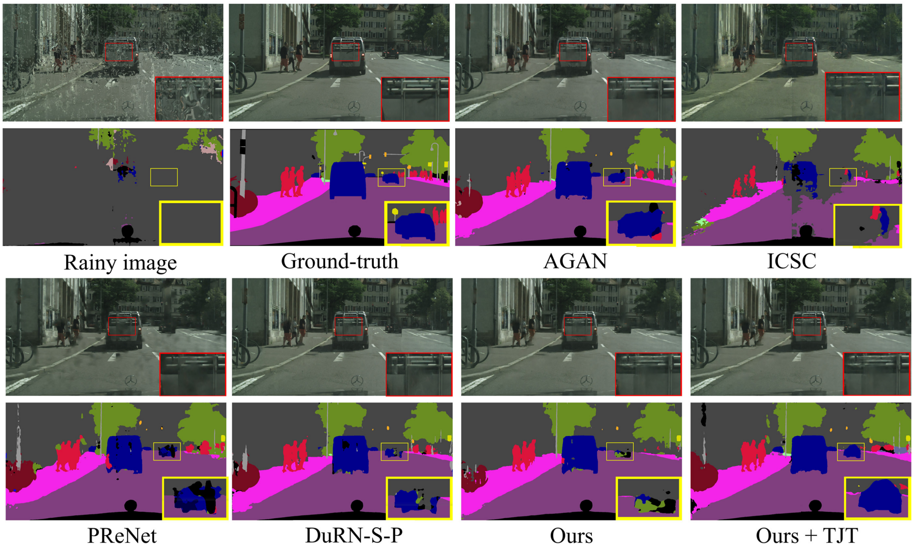

Fig. 2. Examples of deraining and semantic segmentation results obtained by ours and others on our Raindrop-Cityscapes dataset. **Top:** Mean PSNR/SSIM and mIoU results of our model and other state-of-arts. **Bottom:** Deraining and semantic segmentation examples of our model and other state-of-arts.

Method	Rainy image	AGAN	PReNet	DuRN-S-P	ICSC	Ours
PSNR	14.20	20.87	19.02	21.69	20.15	22.01
SSIM	0.5927	0.7870	0.7252	0.8047	0.7269	0.8085
mIoU(%)	13.86	33.32	17.10	33.39	28.52	35.68

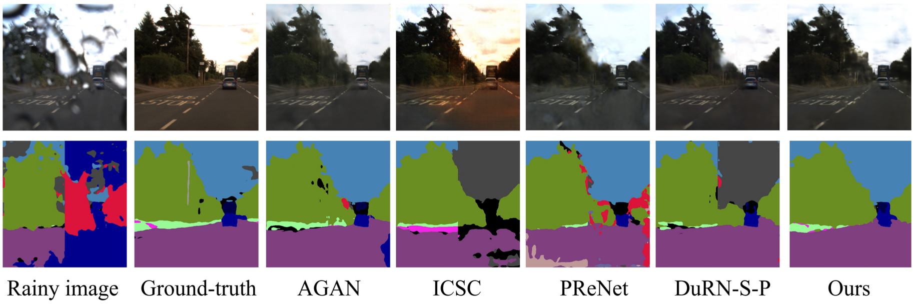

Fig. 3. Examples of deraining and semantic segmentation results obtained by five methods including ours on RobotCar(R). **Top:** Mean PSNR/SSIM and mIoU of our model and other state-of-arts. **Bottom:** Deraining and semantic segmentation examples of our model and other state-of-arts.

6 Conclusion

In this paper, we combine the deraining task with the semantic segmentation task. We create a new synthetic rainy dataset named Raindrop-Cityscapes and propose a novel semantic segmentation task-driven deraining model with a two-stage training method. We demonstrate that our proposed method is capable of not only deraining but also semantic segmentation and achieves the state-of-the-art on both deraining visual performance and semantic segmentation accuracy.

Acknowledgement. This work is partially supported by the National Science Foundation of China under contract No. 61971047 and the Project 2019BD004 supported by PKU-Baidu Fund.

References

1. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
2. Cordts, M., et al.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3213–3223 (2016)
3. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
4. Eigen, D., Krishnan, D., Fergus, R.: Restoring an image taken through a window covered with dirt or rain. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 633–640 (2013)
5. Fan, Z., Sun, L., Ding, X., Huang, Y., Cai, C., Paisley, J.: A segmentation-aware deep fusion network for compressed sensing MRI. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 55–70 (2018)
6. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3855–3863 (2017)
7. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 349–356. IEEE (2009)
8. Haris, M., Shakhnarovich, G., Ukita, N.: Task-driven super resolution: object detection in low-resolution images. arXiv preprint arXiv:1803.11316 (2018)
9. Harley, A.W., Derpanis, K.G., Kokkinos, I.: Segmentation-aware convolutional networks using local attention masks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5038–5047 (2017)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
11. Hore, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369. IEEE (2010)

12. Hu, X., Fu, C.W., Zhu, L., Heng, P.A.: Depth-attentional features for single-image rain removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8022–8031 (2019)
13. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
14. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
15. Kang, L.W., Lin, C.W., Fu, Y.H.: Automatic single-image-based rain streaks removal via image decomposition. IEEE Trans. Image Process. **21**(4), 1742–1755 (2011)
16. Kurihata, H., et al.: Rainy weather recognition from in-vehicle camera images for driver assistance. In: IEEE Proceedings of Intelligent Vehicles Symposium, pp. 205–210. IEEE (2005)
17. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: all-in-one dehazing network. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4770–4778 (2017)
18. Li, S., et al.: Single image deraining: a comprehensive benchmark analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3838–3847 (2019)
19. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2736–2744 (2016)
20. Liu, D., Wen, B., Liu, X., Wang, Z., Huang, T.S.: When image denoising meets high-level vision tasks: a deep learning approach. arXiv preprint arXiv:1706.04284 (2017)
21. Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7007–7016 (2019)
22. Pei, Y., Huang, Y., Zou, Q., Lu, Y., Wang, S.: Does haze removal help CNN-based image classification? In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 682–697 (2018)
23. Pohlen, T., Hermans, A., Mathias, M., Leibe, B.: Full-resolution residual networks for semantic segmentation in street scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4151–4160 (2017)
24. Porav, H., Bruls, T., Newman, P.: I can see clearly now: image restoration via de-raining. arXiv preprint arXiv:1901.00893 (2019)
25. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2482–2491 (2018)
26. Quan, Y., Deng, S., Chen, Y., Ji, H.: Deep learning for seeing through window with raindrops. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2463–2471 (2019)
27. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3937–3946 (2019)
28. Roser, M., Kurz, J., Geiger, A.: Realistic modeling of water droplets for monocular adherent raindrop recognition using bézier curves (2010)

29. Valada, A., Vertens, J., Dhall, A., Burgard, W.: Adapnet: adaptive semantic segmentation in adverse environmental conditions. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4644–4651. IEEE (2017)
30. Wang, S., Wen, B., Wu, J., Tao, D., Wang, Z.: Segmentation-aware image denoising without knowing true segmentation. arXiv preprint arXiv:1905.08965 (2019)
31. Wu, J., Timofte, R., Huang, Z., Van Gool, L.: On the relation between color image denoising and classification. arXiv preprint arXiv:1704.01372 (2017)
32. You, S., Tan, R.T., Kawakami, R., Mukaigawa, Y., Ikeuchi, K.: Adherent raindrop modeling, detection and removal in video. IEEE Trans. Pattern Anal. Mach. Intell. **38**(9), 1721–1733 (2015)
33. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 695–704 (2018)