Semantic Segmentation on Raindrop Degraded Images Using Two-Stage Dual Teacher-Student Learning

Xin Yang[1], Wending Yan[2], Yuan Yuan[2], Michael Bi Mi[2], Robby T. Tan[1]

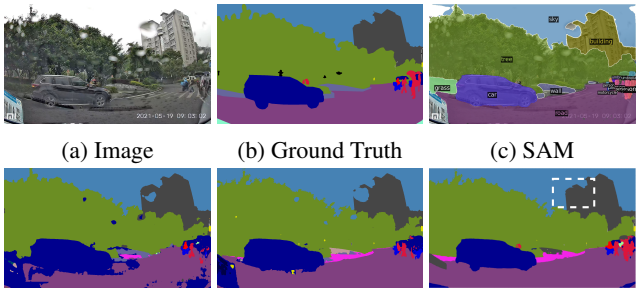
[1]National University of Singapore
[2]Huawei International Pte Ltd
e0674612@u.nus.edu, {yan.wending,yuanyuan10@huawei}.com, michaelbimi@yahoo.com, robby.tan@nus.edu.sg

Abstract

Existing semantic segmentation methods face challenges when processing input images degraded by raindrops on the lens or windshield. Unlike other adverse conditions such as fog and nighttime, which degrade visual quality, raindrops not only impair visual appearances but also introduce misleading occlusion, leading to significant performance drops in current models. The novelty of our approach lies in our two-stage, dual teacher-student framework. We tackle the complex problem of raindrop degradation by dividing it into two distinct challenges: degraded visual appearance and raindrop occlusion. These challenges are then addressed individually in two stages, utilizing two pairs of teacher-student networks. This division enables the networks to develop specialized expertise in handling each aspect of raindrop degradation, enabling their collaboration to achieve superior performance. In the first stage, one teacher-student pair focuses on learning to extract information from visual degraded areas. Building on this, the second teacher-student pair focuses specially on the raindrop occlusion. As such, unlike the existing methods, our approach employs a collaborative approach to decompose and address raindrop-induced degradations. In the second stage, we introduce a mask-based recovery technique to identify and rectify areas that likely contain misleading information, thus further refining the predictions. Additionally, this stage encourages both pairs to expand knowledge by swapping their specialized expertise. Our method achieves a performance of 60.3 mIoU on Rainy WCity (Zhong et al. 2022) and 72.8 mIoU on ACDC Rainy (Sakaridis, Dai, and Van Gool 2021), representing an improvement of +4.4 mIoU and +2.3 mIoU over the existing state-of-the-art methods, respectively.

Introduction

The presence of raindrops on camera lenses or windshields can significantly degrade images, posing challenges for semantic segmentation. One possible approach to address this issue is to remove raindrops from the input image (e.g., (Shao et al. 2021; Özdenizci and Legenstein 2023)) before applying semantic segmentation. However, raindrop removal itself remains a challenging problem, as these methods may generate artifacts that impact the performance of the subsequent semantic segmentation.



(a) Image (b) Ground Truth (c) SAM

(d) Wea. Diff. (e) MIC (f) Ours

Figure 1: (a) Input rainy image. (b) Ground-truth annotated by human. (c) Result from Weather Diffusion (Özdenizci and Legenstein 2023) with a semantic segmentation model (Hoyer, Dai, and Van Gool 2022a) trained on clear images. (d) Result from SAM (Kirillov et al. 2023), which uses a different class protocol as ours. (e) Result from MIC (Hoyer et al. 2023), indicating incorrect predictions due to adherent raindrops. (f) Our method shows better performance particularly in the white rectangular area. Zoom in for better visualization.

A few unsupervised domain adaptation (UDA) methods (e.g., (Hoyer et al. 2023; Kennerley et al. 2023)) address adverse weather degradations, such as rain streaks, haze, nighttime, etc., without relying on annotated data for the weather-degraded targets. However, they are ineffective in handling raindrops. This is largely attributed to the unique characteristics of raindrops: transparent, blurred, varying in size, and unevenly distributed across an image. These features pose challenges for UDA methods to accurately distinguish regions affected by adherent raindrops from the background image. Moreover, transparent yet blurred raindrops occlude the background scene, introducing misleading visual cues that can misguide the methods.

In this paper, we propose a novel approach to tackle complex raindrop degradation by incorporating two pairs of teacher-student networks, each with different expertise, across two stages, along with a mask-based raindrop recovery strategy. Learning to extract correct information from the complex raindrop degradation presents a challenge. We address this problem with two pairs of teacher-student net-

works, one starts by tackling the degraded visual appearance, while another learns to specifically address the raindrop occlusion with guidance from the first pair.

The first pair, referred to as the "reference pair", employs standard augmentation on the target images, serving as the initial step for our framework to extract information from the areas with degraded visual appearances. This augmentation includes color jitters, masking, and others, as recommended in (Hoyer, Dai, and Van Gool 2022a,b; Hoyer et al. 2023). The second pair, which we call the "raindrop pair", applies raindrop-specific augmentations to the target raindrop images. Unlike the standard augmentation, which only blocks and alters visual appearance, in our raindrop-specific augmentation, we synthesize raindrop-like occlusion based on a raindrop physics model. This not only obscures semantic information but also introduces misleading refracted background information. Moreover, our raindrop-like occlusion has varying sizes and ratios and are randomly distributed on the target image, making it more challenging for the model to identify and interpret the augmented regions. This specialized augmentation compels the network to extract correct information from the raindrop occlusion, hence enhancing its robustness against such challenges.

In the initial stage, students progressively learn from the pseudo-labels provided by teachers. However, a common problem in existing teacher-student networks is that the model learns effectively in areas with high confidence scores but overlooks the areas with low confidence scores. This issue is particularly severe in the complex and challenging context of raindrop degradation. To overcome this limitation, our framework introduces a second stage designed to enhance the quality of pseudo-labels in these low-confidence areas. This stage employs a mask-based recovery technique. The complex raindrop degradations introduce misleading information from the surrounding environment through refraction, resulting in the corresponding pseudo-labels having low confidence scores. To deal with this problem, we mask these uncertain regions, eliminating the misleading occluded information. The altered image, with occlusion minimized, is then fed into our network, facilitating the generation of predictions that are not affected by refracted information. Following this, we incorporate the predictions derived from the modified image to refine the pseudo-labels, hence reducing the influence of misleading information caused by transparent but blurry raindrops. Furthermore, by swapping the augmentations in the second stage, we encourage both pairs to expand their knowledge, which in turn further enhances the model's performance. Fig. 1 shows the effectiveness of our overall method in comparison with the state-of-the-art methods. In summary, our contributions are as follows:

- We propose a novel learning approach that addresses the complex raindrop degradation by decomposing it and tackling each aspect through collaboration between two pairs of teacher-student networks, each with different expertise. One pair focuses on learning from degraded visual appearances, while the other concentrates on raindrop occlusion.
- We introduce a mask-based recovery technique along with a two-stage updating strategy. Our mask-based recovery technique improves the pseudo-labels within the raindrop-affected regions to address the misleading information introduced by raindrops. Meanwhile, our two-stage updating strategy aims to expand the knowledge of both teacher-student pairs, hence enhancing the overall predictions by of our dual student-teacher framework.
- We present a novel method for semantic segmentation designed to mitigate raindrop degradations, without requiring ground-truth labels for raindrop-degraded images. Our method achieves a performance of **60.3 mIoU** on Rainy WCity (Zhong et al. 2022) and **72.8 mIoU** on ACDC Rainy (Sakaridis, Dai, and Van Gool 2021), indicating an improvement of +4.4 mIoU and +2.3 mIoU over the existing state-of-the-art method, respectively.

Related Work

Raindrop Removal Raindrops attached to the camera lenses or windshields exhibit diverse sizes, shapes, and can coalesce into water residues and flows (You et al. 2013), Their presence can significantly reduce the performance of computer vision systems. Numerous studies have attempted to address this challenge by detecting and eliminating raindrops, often involving the creation of masks (Qian et al. 2018; Quan et al. 2021; Shao et al. 2021; Zhang et al. 2021; Pizzati, Cerri, and de Charette 2023), synthesis of raindrops (Hao et al. 2019), and image restoration techniques (Lin et al. 2024b,a; Chen et al. 2025; Zamir et al. 2022; Özdenizci and Legenstein 2023). However, these techniques necessitate corresponding clear images for supervised training, a requirement often difficult to fulfill. While it's possible to train raindrop removal methods on specific datasets with clear images, their generalization to unlabeled raindrop datasets is hindered by inherent domain gaps of various scenes.

Domain Adaptation Methods Domain adaptation techniques find widespread application across diverse computer vision tasks. These methods employ strategies like adversarial training (Vu et al. 2019; Sindagi et al. 2020) and self-training with pseudo-labels (Deng et al. 2021; Hoyer, Dai, and Van Gool 2022a,b; Hoyer et al. 2023; Yi et al. 2024; Bi et al. 2024; Bi, You, and Gevers 2024a,c,b). These approaches adapt models from a labeled source domain (e.g., clear weather) to an unlabeled target domain (e.g., rainy). Consequently, the model can achieve commendable performance in the target domain, even in the absence of corresponding ground truth. While several domain adaptation methods address rainy conditions (Sindagi et al. 2020; Li et al. 2023b,a), their primary focus lies in addressing degradations caused by rain streaks and accumulation effects.

Augmentation Strategies Previous studies have shown that enforcing consistency between a model's predictions on images with and without augmentation can enhance the efficiency of semi-supervised learning in acquiring representations (He et al. 2022; Hoyer et al. 2023). Nevertheless, their findings also indicated that complex augmentation could result in a decline in performance. This phenomenon, is known as the negative impact of augmentation (Yang et al.

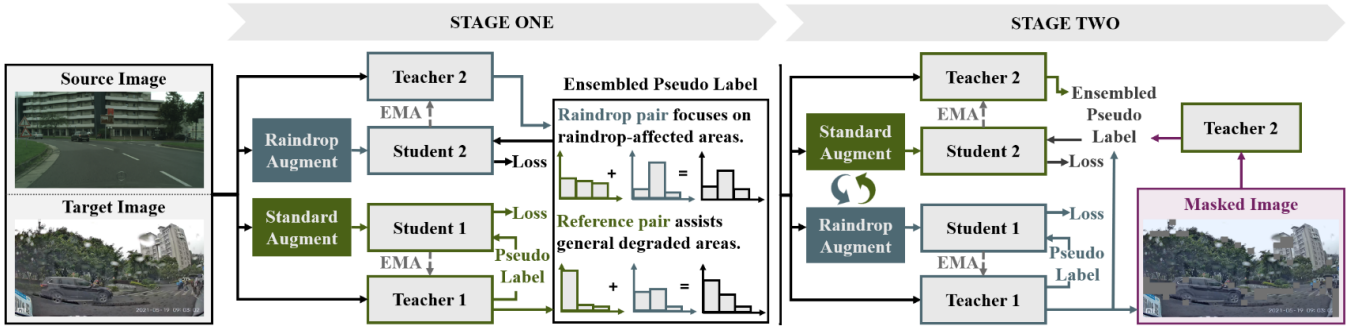


Figure 2: Our architecture for adapting a model from clear weather (source) to rainy condition with raindrops (target). The architecture consists of several key components: (1) Dual teacher-student framework, where two pairs of teacher-student networks are collaborated. (2) Second stage with mask-based recovery, where we incorporate the predictions from the masked images to correct the pseudo-label, and further improve the teacher-student pairs by swapping the augmentations.

2022; Kong et al. 2023; Kennerley et al. 2023). Existing methods empirically choose augmentation based on tasks and datasets (He et al. 2022; Hoyer et al. 2023; Tian et al. 2023). However, when addressing real-world raindrop problems, the raindrops can contain deceptive information and exhibit significant variation in sizes and ratios, which cannot be described properly by the existing augmentation. Consequently, the creation of raindrop-specific augmentation and the development of a method capable of learning from such complex augmentation remain challenges.

Proposed Method

Fig. 2 illustrates the pipeline of our method. In this pipeline, we introduce the novelty of addressing the complex issue of raindrop degradation by dividing it into two separate challenges: the deterioration of visual quality and the obstruction caused by raindrops. We address these issues in a two-stage process, utilizing two pairs of specialized teacher-student networks, with pair dedicated to addressing one of the challenges independently. In the first stage, our "reference pair" comprises Teacher-1 (T_1) and Student-1 (S_1), undergoing training with the standard augmentation. Simultaneously, our "raindrop pair" consists of Teacher-2 (T_2) and Student-2 (S_2), trained with our raindrop-specific augmentation. In the second stage, we incorporate the mask-based raindrop recovery technique to further refine the quality of the pseudo-labels for the uncertain areas. Additionally, upon reaching stage two, we swap the augmentations applied to the dual teacher-student pairs: T_1 and S_1 are now trained with our raindrop-specific augmentation, while T_2 and S_2 are trained with the standard augmentation.

Dual Teacher-Student Framework

Our dual teacher-student framework incorporates two teacher-student pairs: the reference pair utilizing the standard augmentation for the target images; and the raindrop pair employing our raindrop-specific augmentation for the target images. It's important to note that merging all these augmentations into a single teacher-student network can negatively affect performance, leading to suboptimal outcomes. This issue underlies the shortcomings of existing methods. An ablation study detailing this is presented in subsequent sections.

Reference Pair In each iteration, S_1 performs supervised learning for semantic segmentation on a source image. Additionally, S_1 makes predictions for an augmented target image, where the augmentation includes standard augmentation like color jitter, random cropping, and masking. The predictions generated by S_1 are then compared with the pseudo-labels that are generated from T_1 for unsupervised learning on the new domain. This process has been utilized in some existing works to enhance the model's robustness against degraded visual appearances due to various adverse conditions, such as haze and nighttime environments (Hoyer, Dai, and Van Gool 2022a,b; Hoyer et al. 2023; Kennerley et al. 2023). Following the existing works, we define the pseudo-labels as:

$$p_1(x)_{ij} = [c = \underset{c'}{\operatorname{argmax}}(S_1(x)_{ijc'})], \tag{1}$$

where p_1 represents the pseudo-labels, i, j represent the pixel, c represents the ground truth class and c' represents the predicted class for the corresponding pixel. T_1 is regularly updated using the Exponential Moving Average (EMA) of S_1. This EMA process helps in stabilizing and improving the performance of T_1 over time.

Raindrop Pair For the raindrop pair, S_2 undergoes the same supervision process as S_1. However, for the input images fed to S_2, we apply the raindrop-specific augmentation. Unlike the standard augmentation, which involves a single augmented image, our raindrop-specific augmentation generate multiple augmented images, with raindrop occlusion with various sizes and ratios. We render raindrops with dimensions of 64×64 and a 70% ratio to create the first augmented target image. Then, we render raindrops with dimensions 32×32 and a 60% ratio to create the subsequent augmented target image. This procedure is repeated, generating augmented images with raindrops measuring 16×16 and a 50% ratio, as well as 8×8 with a 40% ratio. The effectiveness of having multiple augmented images is evaluated in the ablation studies. We formulate raindrop-specific occlu-

sion based on the adherent raindrop physics model:

$$I_{ij} = (1 - \tau)I_{\mathrm{e}} + \tau I_{\mathrm{r}}, \tag{2}$$

where I_{r} represents information refracted by the raindrop, while I_{e} denotes the clear information in the raindrop area as if there is no raindrop degradation. τ represents the proportion of the refracted information (You et al. 2013, 2015). To simulate the behavior of one raindrop, we first randomly select its location, size, shape and thickness ρ. With this information, we can ascertain the directions of the refracted lights and compute a displacement map (U, V). Hence, for a pixel (x, y) in the raindrop, its refracted information is mapped to $(x + U(x, y)\rho, y + V(x, y)\rho)$. We then integrate this refracted information into the area and apply focus blurring using a Gaussian point spread function. In this augmentation, we set τ to 1, to introduce a high level of degradation.

After generating the raindrop-specific augmented images, we input them into S_2. Utilizing the pseudo-labels produced by T_2, S_2 learns from the more challenging augmented data, compelling it to learn more. Since, employing different raindrop-specific augmentation provides the model with an opportunity to learn different types of raindrop degradations and their corresponding clear backgrounds. Existing methods fail in the combined challenges of degraded visual appearances and raindrop occlusion, resulting in suboptimal performance. In contrast, our approach decomposes the problem and allocates it to two pairs of teacher-student models, encouraging them to solve the problem through collaboration. To facilitate this collaboration, we integrate predicted logits from the teacher of the reference pair, T_1, to refine and calibrate the pseudo-labels generated by T_2, expressed as:

$$p(x)ij = [c = \operatorname{argmax} c'(\alpha T_1(x)_{ijc'} + (1 - \alpha)T_2(x)_{ijc'})], \tag{3}$$

where α is a parameter controlling the contributions of predicted logits from each pair. These ensembled and calibrated pseudo-labels are utilized to instruct the student from the raindrop pair, S_2.

Since T_1 is only exposed to the standard augmentation, it is less susceptible to raindrop occlusion. Hence, when T_1 produces highly confident predictions on general visual degradation (indicated by the largest $T_1(x)_{ijc'}$ being significantly larger than other classes), these confident and reliable predictions are given priority in the combination with $T_2(x)_{ijc'}$. As a result, p becomes equal to p_1. This implies that the reference pair serves as a reference or anchor to assist the raindrop pair in mitigating the negative effects from the visual degraded areas.

Note that, compared to the raindrop pair (T_2 and S_2), the reference pair (T_1 and S_1) has limited potential due to the simpler standard augmentation. Their predictions for raindrop-occluded areas in the target images may be less confident and potentially incorrect (indicated by the largest $T_1(x)_{ijc'}$ for p_1 being only slightly larger than other classes). In contrast, T_2 is updated by the EMA of S_2 and exposed to more challenging raindrop-specific augmentation; hence, $T_2(x)_{ijc'}$ has the opportunity to contain predictions with high confidence scores on the raindrop-occluded areas. Therefore, the combined pseudo-label p has the chance to outperform p_1, particularly in raindrop-occluded areas.

When a conflict arises in the pseudo-labels between the two pairs, where each pair exhibits high confidence in their predictions but may have different label predictions, our ensemble pseudo-label will yield a lower combined confidence score. This results in a reduced learning weight on the corresponding regions, allowing the raindrop pair to reevaluate and reconsider the conflicted regions.

Mask-Based Raindrop Recovery

Due to the severity, the model's predictions for some raindrops-affected regions may be uncertain, indicated by low confidence scores. To address this issue, we introduce the second stage with a mask-based recovery technique for the raindrop pair to identify such areas and rectify the corresponding pseudo-labels.

Our mask-based recovery technique involves two rounds in each training iteration. In the first round, T_1 receives the target image and generates predictions. Some areas in the predictions are affected by raindrops, resulting in low confidence scores. We compute the confidence of the predictions generated in this round as follows:

$$\mathrm{con}_{\mathrm{rain},ij} = \max(T_1(x)_{ijc'}). \tag{4}$$

In the second round, based on the predictions from the first round, we apply masks to the target image in areas where confidence scores are lower than the average confidence score for the entire image. The masked image, x_{mask}, is then fed to T_2 to generate new predictions. The confidence of the predictions obtained in this round is computed as:

$$\mathrm{con}_{\mathrm{mask},ij} = \max(T_2(x_{\mathrm{mask}})_{ijc'}). \tag{5}$$

Subsequently, we obtain a confidence mask to select the areas that $T_2(x_{\mathrm{mask}})$ is more confident than $T_1(x)$, following $M_{\mathrm{con}} = \mathrm{con}_{\mathrm{rain},ij} < \mathrm{con}_{\mathrm{mask},ij}$. By leveraging M_{con}, we adjust the pseudo-labels from T_{rain}, by injecting the raindrop-free $T_2(x_{\mathrm{mask}})_{ijc'}$ to $T_1(x)_{ijc'}$ as follow:

$$\phi_{\mathrm{rec},ijc'} = T_1(x)_{ijc'} + \theta M_{\mathrm{con}} T_2(x_{\mathrm{mask}})_{ijc'}, \tag{6}$$

$$p_{\mathrm{rec}} = \underset{c}{\operatorname{argmax}}(\alpha T_2(x)_{ijc'} + (1 - \alpha)\phi_{\mathrm{rec},ijc'}) \tag{7}$$

We incorporate the reliable logits from $T_2(x_m)_{ijc'}$ with high confidence and utilize them to generate new pseudo-labels denoted as p_{rec} based on the calibrated logits $\phi_{\mathrm{rec},ijc'}$. θ is a parameter that prevents $T_2(x_{\mathrm{mask}})_{ijc'}$ from being dominant. Moreover, we replace the previous pseudo-labels p with the newly generated p_{rec} to guide the raindrop pair. An example of this process is presented in Fig. 3. This iterative pseudo-label adjustment step effectively prevents the students from learning distorted pseudo-labels from the teachers, allowing for a progressive recovery of the raindrop-affected areas.

Two-Stage Updating Strategy

With the current configuration, the raindrop pair continuously learns from both the reference pair and the raindrop pair. However, the learning of both pairs is constrained by their exclusive focus on their specifically assigned augmentations. Consequently, their collective knowledge may limited over time, as each pair has not been exposed to the alternate augmentation strategy.

(a) Image (b) Original p_{rain} (c) Mask M_{con} (d) Refined p_{rec} (e) Final Prediction

Figure 3: The illustration of the mask-based raindrop recovery in one iteration. In (b), the pseudo-labels are misled by raindrops. In the second round, upon inputting the masked image into the network, we rectify the wrong pseudo-labels by employing the confidence mask, as illustrated in (c). This process aims to correct the erroneous pseudo-labels that were influenced by raindrops, shown in (d). Over many iterations, the entire erroneous region will be filled, resulting in the final prediction in (e).

To address this limitation, we introduce a two-stage updating strategy. Stage one involves exposing the reference pair (T_1 and S_1 as illustrated in Fig. 2) to the standard augmentation, while the raindrop pair (T_2 and S_2) is exposed to both the raindrop-specific augmentation and guidance from the reference pair. After the learning completes a certain number of iterations, in stage two, we swap the raindrop and reference pairs: T_2 and S_2 are now exposed to the standard augmentation, while T_1 and S_1 are exposed to the raindrop-specific augmentation.

In stage two, the raindrop pair (T_2 and S_2) guides the reference pair (T_1 and S_1). The calibrated pseudo-labels, $p(x)_{ij}$, calculated using Eq. (3), are fed to the reference pair, while the raindrop pair relies solely on its own logits for guidance and learning. Note that we do not utilize the combined pseudo-labels for both pairs, as we aim to preserve their diversities for enhanced ensemble pseudo-labels. In contrast to stage one, both teacher-student pairs are now updated, resulting in improved combined pseudo-labels.

Experiments

Datasets For our source domain, we use the Cityscapes dataset (Cordts et al. 2016), which comprises real-world images captured under clear weather conditions. As for the target domain, we utilize both the ACDC Rainy set (Sakaridis, Dai, and Van Gool 2021) and the Rainy WCity dataset (Zhong et al. 2022), which consists of real-world images captured under diverse rain conditions. In our experiments, we evaluate our model's robustness against rain conditions with severe raindrops using the Rainy WCity dataset as the target dataset. To demonstrate the versatility of our approach, we also evaluate our model on the ACDC Rainy set, illustrating that our method extends beyond raindrops.

Baseline Models In our experiments, we conduct comparisons with both raindrop removal methods and unsupervised domain adaptation methods. For the raindrops removal methods, we consider Restormer (Zamir et al. 2022) and Weather Diffusion (Özdenizci and Legenstein 2023). For Restormer, the pretrained model is not trained on raindrop removal tasks. Therefore, we fine-tune it on the Raindrop dataset (Qian et al. 2018), following the settings suggested by (Özdenizci and Legenstein 2023). In contrast, Weather Diffusion offers a pretrained model for raindrop removal tasks, which we utilize directly as provided by the authors. After processing the images with the raindrop removal methods, we proceed to feed these processed images into a semantic segmentation model. This segmentation model has been trained using the Cityscapes dataset in a supervised manner. We use DAFormer (Hoyer, Dai, and Van Gool 2022a) as the backbone of our semantic segmentation model. The combined pipeline enables us to evaluate the performance of the raindrop removal methods in the context of semantic segmentation tasks.

As for the domain adaptation methods, we compare with DAFormer, HRDA (Hoyer, Dai, and Van Gool 2022b), and MIC (Hoyer et al. 2023). When training these models, we follow the same training configurations as suggested in these paper, and we also use DAFormer as the backbone of these methods. To make the comparison fair, we use the same DAFormer pretrained model as suggested in MIC, for all the experiments. We train two models using our method, one is adapting from Cityscapes to Rainy WCity to evaluate the effectiveness of dealing with raindrops, another is adapting from Citysacpes to ACDC Rainy for showcasing our methods performance under diverse rain conditions.

As for our method's parameters, we set the initial value of α to be 0.8, gradually reducing it to 0.2 as the number of iterations progresses. θ is assigned a value of 0.1 to prevent it from overpowering the learning process. For the standard augmentation, we adopt the usually used augmentation as recommended in (Hoyer, Dai, and Van Gool 2022a,b; Hoyer et al. 2023). On the other hand, for the raindrop augmentation, we utilize four raindrop types with varying sizes of 64×64, 32×32, 16×16, and 8×8, along with corresponding ratios of 0.7, 0.6, 0.5, and 0.4, respectively. We gradually introducing more raindrops with smaller sizes and ratios to encourage the model to learn small areas.

Quantitative Results

As shown in Tab. 1, our models outperform other methods on Rainy WCity. Our model surpasses Weather Diffusion and MIC by 18.4 % and 4.4 % in mIoU Avg., respectively, signifying substantial advancements. The consistently elevated mIoU values across all classes and the notable enhancement observed in most classes underscore the robustness of our model in mitigating the severe occlusion arising from raindrops.

Furthermore, our evaluation extends to light rain conditions with minor raindrop occlusion, as demonstrated in Tab. 2. As raindrop occlusion is not consistently present in light rain conditions, we have excluded raindrop-removal

Clear-to-Rain: Cityscapes → Rainy WCity																			
Method	DA	Road	Side W.	Build.	Wall	Fence	Pole	Traf.L.	Sign	Vege.	Sky	Person	Rider	Car	Truck	Bus	Motor.	Bike	mIoU
Restormer	✗	43.9	13.5	48.4	11.4	42.1	24.7	36.0	52.9	76.3	91.3	28.5	24.9	46.3	**1.8**	63.7	8.7	18.2	37.2
Wea. Diff.	✗	60.8	9.4	60.7	**17.7**	46.3	30.6	39.8	63.6	78.9	91.5	31.9	31.3	46.0	1.6	65.3	14.3	23.2	41.9
DAFormer	✓	84.6	28.3	70.7	7.8	66.3	38.7	41.7	75.6	83.8	92.7	34.9	34.7	73.9	0.2	80.8	3.2	37.4	50.3
HRDA	✓	84.7	27.3	72.5	9.8	59.1	33.3	35.1	77.0	**85.6**	93.0	43.1	38.5	75.0	0.0	84.0	3.3	38.8	53.4
MIC	✓	83.6	26.9	71.7	7.8	**69.3**	40.2	39.8	76.1	85.3	94.3	53.0	39.2	67.9	0.0	84.7	9.9	**51.3**	55.9
Ours	✓	**87.4**	**32.4**	**73.4**	6.6	68.6	**45.1**	**45.2**	**77.9**	**85.6**	**94.5**	**55.5**	**48.8**	**75.3**	0.6	**86.7**	**35.8**	51.1	**60.3**

Table 1: Quantitative results of Ours compare to the existing raindrop removal and domain adaptation methods, evaluated against Rainy WCity. **Bold** numbers are the best scores, and underlined numbers are the second-best scores. The IoU (%) of each class, the mIoU of all the classes are presented. Our method outperforms the best existing method over 4.4 mIoU (%).

Clear-to-Light Rain: Cityscapes → ACDC Rainy																			
Method	Road	Side W.	Build.	Wall	Fence	Pole	Traf.L.	Sign	Vege.	Terr.	Sky	Person	Rider	Car	Truck	Bus	Train	Motor.	Bike	mIoU
DAFormer	92.3	71.4	89.5	39.1	33.1	57.2	**79.2**	**72.8**	91.6	60.5	97.6	76.0	55.2	92.4	44.3	68.4	37.8	**59.6**	62.8	67.4
HRDA	**92.7**	**72.0**	89.6	41.1	33.9	**58.5**	79.0	72.4	91.8	**61.6**	97.6	**76.8**	**62.9**	**92.6**	46.6	73.0	39.0	53.3	65.6	68.4
MIC	91.4	68.6	88.5	40.0	31.0	54.0	78.8	66.4	92.0	60.5	97.9	75.8	48.5	92.1	54.6	95.0	87.3	55.9	61.2	70.5
Ours	91.4	69.4	**90.3**	**41.4**	**34.1**	56.0	79.0	65.4	**92.5**	59.1	**98.1**	76.6	61.0	91.7	**69.2**	**95.3**	**88.2**	56.9	**68.0**	**72.8**

Table 2: Quantitative results of Ours compare to the existing domain adaptation methods, evaluated against ACDC Rainy (Val.). **Bold** numbers are the best scores, and underlined numbers are the second-best scores. Our method outperforms the best existing method over 2.3 mIoU (%).

methods from our comparisons. Despite our method's primary design for raindrops, the results reaffirm our model's ability to generalize effectively to lighter rain scenarios.

Qualitative Results

We show our qualitative results against Rainy WCity and ACDC Rainy in Fig. 4, for adapting the model from clear weather to rain with raindrop occlusion. We evaluate Ours against Weather Diffusion with DAFormer trained on clear image, MIC, and the ground truths semantic segmentation maps. From the results, we can observe that the raindrop removal methods suffer from the domain gaps between their trained dataset, and the real-world rainy datasets, resulting a massive hallucination effects and hence a poor performance. As for MIC, since it does not have any occlusion recovery technique, it cannot recover the areas occluded by the raindrops and waterflow. In the first row, existing methods failed to identify the rider and the sidewalk accurately due to the raindrop occlusion. In the second row, the raindrop blurs the bus, causing existing methods to fail in accurately identifying the scenes. Conversely, in our approach, one pair of teacher-student networks is dedicated to extracting information from visually degraded areas, while another pair focuses on addressing raindrop occlusion. Combined with our two-stage updating strategy and mask-based recovery technique, our models exhibit better robustness under raindrop conditions and successfully address the issues present in existing methods, as clearly demonstrated in the qualitative results.

Ablation Studies

Dual Teacher-Student Settings Ablation studies of the dual teacher-student framework with various configurations:

- No Dual TS(4): One pair of teacher-student using both standard and raindrop augmentations, consisting of four raindrop types (64×64, 32×32, 16×16, and 8×8), each with corresponding ratios (0.7, 0.6, 0.5, and 0.4).
- No Dual TS(1): One pair of teacher-student with the standard augmentation, which serves as the MIC baseline and our reference pair in stage one.
- Dual TS(2): Application of the dual teacher-student algorithm with one reference pair (No Dual TS(1)) and one raindrop pair with two raindrop types (64×64 and 32×32) and a fixed α of 0.5.
- Dual TS(2a): Dual TS(2) system, but with an adaptive α.
- Dual TS(3a): An extension of Dual TS(2a) with an additional raindrop type (16×16) and adaptive α.
- Dual TS(4a) (Ours): Further extending Dual TS(3a) with another raindrop type (8×8) and adaptive α.

As shown in Fig. 5, No Dual TS(4) without reference pair has inferior performance compared to the baseline that employs only the standard augmentation. This outcome indicates that complex raindrop degradation pose a significant challenge for a single teacher-student pair to effectively handle, highlighting the importance of dual networks. Upon implementing the dual teacher-student algorithm in Dual TS(2), a noticeable performance improvement is observed

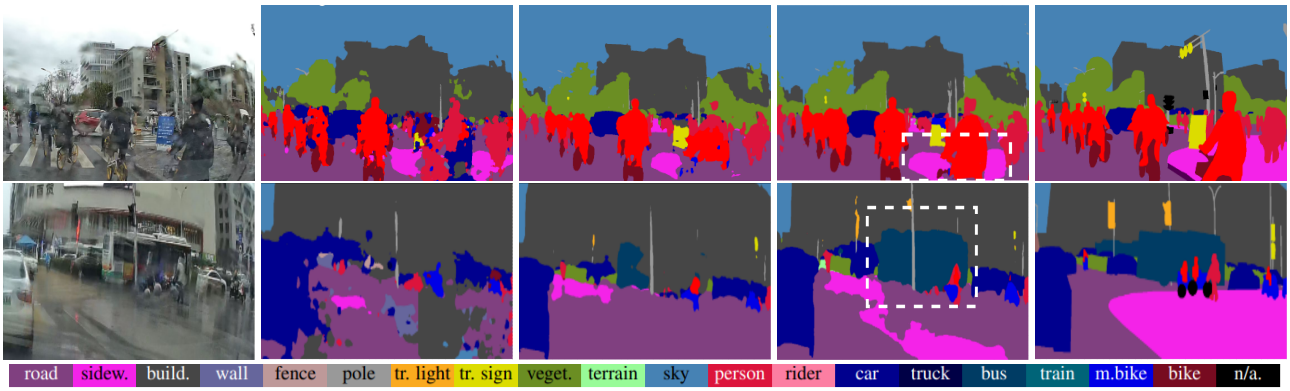


Figure 4: Comparisons on the semantic segmentation performance with Weather Diffusion (Özdenizci and Legenstein 2023), MIC (Hoyer et al. 2023), Ours, and ground truths on Rainy WCity and ACDC Rainy. Different colors represent different classes.

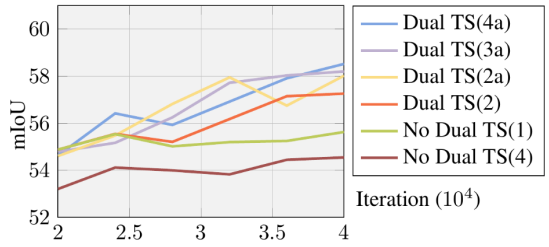


Figure 5: Ablation analysis of models with different augmentations. Dual TS(4) is our chosen configuration.

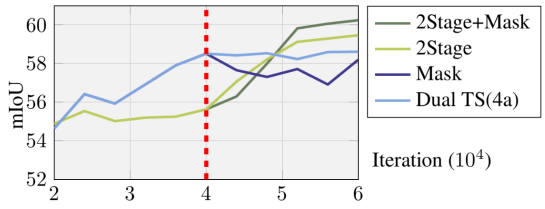


Figure 6: Performance change for the two pairs of teacher-student in stage one (left of the red dashed line), and stage two (right). Ours (2Stage+Mask) outperforms the others.

compared to No Dual TS(1), indicating the effectiveness of evolving multiple raindrop augmented images. Furthermore, introducing an adaptive α from large to small enhances the performance, resulting in an approximate 1 mIoU improvement from Dual TS(2) to Dual TS(2a). The differences in performance among Dual TS(2a), Dual TS(3a), and Dual TS(4a) are minimal. Empirical evaluation led us to select the optimal setting, Dual TS(4a), as our best-performing model.

Two-Stage Updating and Mask-Based Recovery We conducted ablation studies investigating the two-stage updating and mask-based recovery aspects of our method:

- Dual TS(4a): We directly use Dual TS(4a) from stage one, without updating or mask-based recovery.
- Mask: We directly employing Dual TS(4a) and No Dual TS(1) for mask-based recovery, without updating them.
- 2Stage: We apply a two-stage updating strategy, without mask-based recovery.
- 2Stage+Mask (Ours): Both two-stage updating strategy and mask-based recovery are applied.

As shown in Fig. 6, through the two-stage updating, both 2Stage+Mask (Ours) and 2Stage exhibit enhancements, with 2Stage+Mask (Ours) showcasing a more performance gain attributable to the mask-based recovery technique. Notably, it becomes evident that direct adoption of mask-based recovery without updating the pairs can lead to a slight performance drop. This outcome is primarily attributed to the constrained performance of No Dual TS(1) from stage one, which limits the efficacy of the recovery process.

Conclusion

We have introduced an unsupervised domain adaptation method designed to address the challenges posed by adherent raindrops. The novelty of our approach lies in our two-stage, dual teacher-student framework. We tackle the complex problem of raindrop degradation by dividing it into two distinct challenges: degraded visual appearance and raindrop occlusion. These challenges are then addressed individually in two stages, utilizing two pairs of teacher-student networks. This division enables the networks to develop specialized expertise in handling each aspect of raindrop degradation, enabling their collaboration to achieve superior performance. Through comprehensive evaluations on a rainy dataset with severe raindrop occlusion, our model has demonstrated superiority over the finest existing raindrop removal and domain adaptation methods. Additionally, we have extended our assessment to a diverse rain dataset that features raindrop occlusion varying from light to severe, thereby showcasing the generalization capability of our method.

References

Bi, Q.; Yi, J.; Zheng, H.; Zhan, H.; Huang, Y.; Ji, W.; Li, Y.; and Zheng, Y. 2024. Learning Frequency-Adapted Vision Foundation Model for Domain Generalized Semantic Segmentation. In *Annual Conference on Neural Information Processing Systems*.

Bi, Q.; You, S.; and Gevers, T. 2024a. Generalized Foggy-Scene Semantic Segmentation by Frequency Decoupling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1389–1399.

Bi, Q.; You, S.; and Gevers, T. 2024b. Learning content-enhanced mask transformer for domain generalized urban-scene segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 819–827.

Bi, Q.; You, S.; and Gevers, T. 2024c. Learning generalized segmentation for foggy-scenes by bi-directional wavelet guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 801–809.

Chen, S.; Ye, T.; Zhang, K.; Xing, Z.; Lin, Y.; and Zhu, L. 2025. Teaching Tailored to Talent: Adverse Weather Restoration via Prompt Pool and Depth-Anything Constraint. In *European Conference on Computer Vision*, 95–115. Springer.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3213–3223.

Deng, J.; Li, W.; Chen, Y.; and Duan, L. 2021. Unbiased mean teacher for cross-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4091–4101.

Hao, Z.; You, S.; Li, Y.; Li, K.; and Lu, F. 2019. Learning from synthetic photorealistic raindrop for single image raindrop removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0–0.

He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; and Girshick, R. 2022. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000–16009.

Hoyer, L.; Dai, D.; and Van Gool, L. 2022a. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9924–9935.

Hoyer, L.; Dai, D.; and Van Gool, L. 2022b. HRDA: Context-aware high-resolution domain-adaptive semantic segmentation. In *European Conference on Computer Vision*, 372–391. Springer.

Hoyer, L.; Dai, D.; Wang, H.; and Van Gool, L. 2023. MIC: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11721–11732.

Kennerley, M.; Wang, J.-G.; Veeravalli, B.; and Tan, R. T. 2023. 2PCNet: Two-Phase Consistency Training for Day-to-Night Unsupervised Domain Adaptive Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11484–11493.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643*.

Kong, L.; Ma, M. Q.; Chen, G.; Xing, E. P.; Chi, Y.; Morency, L.-P.; and Zhang, K. 2023. Understanding Masked Autoencoders via Hierarchical Latent Variable Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7918–7928.

Li, J.; Xu, R.; Ma, J.; Zou, Q.; Ma, J.; and Yu, H. 2023a. Domain Adaptation for Enhanced Object Detection in Foggy and Rainy Weather for Autonomous Driving. *arXiv preprint arXiv:2307.09676*.

Li, M.; Xie, B.; Li, S.; Liu, C. H.; and Cheng, X. 2023b. VBLC: Visibility Boosting and Logit-Constraint Learning for Domain Adaptive Semantic Segmentation under Adverse Conditions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8605–8613.

Lin, Y.; Fu, Z.; Wen, K.; Ye, T.; Chen, S.; Meng, G.; Wang, Y.; Huang, Y.; Tu, X.; and Ding, X. 2024a. Unsupervised Low-light Image Enhancement with Lookup Tables and Diffusion Priors. *arXiv preprint arXiv:2409.18899*.

Lin, Y.; Ye, T.; Chen, S.; Fu, Z.; Wang, Y.; Chai, W.; Xing, Z.; Zhu, L.; and Ding, X. 2024b. AGLLDiff: Guiding Diffusion Models Towards Unsupervised Training-free Real-world Low-light Image Enhancement. *arXiv preprint arXiv:2407.14900*.

Özdenizci, O.; and Legenstein, R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Pizzati, F.; Cerri, P.; and de Charette, R. 2023. Physics-Informed Guided Disentanglement in Generative Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10300–10316.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2482–2491.

Quan, R.; Yu, X.; Liang, Y.; and Yang, Y. 2021. Removing raindrops and rain streaks in one go. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9147–9156.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10765–10775.

Shao, M.-W.; Li, L.; Meng, D.-Y.; and Zuo, W.-M. 2021. Uncertainty guided multi-scale attention network for raindrop removal from a single image. *IEEE Transactions on Image Processing*, 30: 4828–4839.

Sindagi, V. A.; Oza, P.; Yasarla, R.; and Patel, V. M. 2020. Prior-based domain adaptive object detection for hazy and rainy conditions. In *European Conference on Computer Vision*, 763–780. Springer.

Tian, K.; Jiang, Y.; Diao, Q.; Lin, C.; Wang, L.; and Yuan, Z. 2023. Designing BERT for Convolutional Networks: Sparse and Hierarchical Masked Modeling. *arXiv:2301.03580*.

Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; and Pérez, P. 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2517–2526.

Yang, L.; Zhuo, W.; Qi, L.; Shi, Y.; and Gao, Y. 2022. St++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4268–4277.

Yi, J.; Bi, Q.; Zheng, H.; Zhan, H.; Ji, W.; Huang, Y.; Li, Y.; and Zheng, Y. 2024. Learning Spectral-decomposited Tokens for Domain Generalized Semantic Segmentation. In *ACM Multimedia 2024*.

You, S.; Tan, R. T.; Kawakami, R.; and Ikeuchi, K. 2013. Adherent raindrop detection and removal in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1035–1042.

You, S.; Tan, R. T.; Kawakami, R.; Mukaigawa, Y.; and Ikeuchi, K. 2015. Adherent raindrop modeling, detection and removal in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9): 1721–1733.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Zhang, K.; Li, D.; Luo, W.; and Ren, W. 2021. Dual attention-in-attention model for joint rain streak and raindrop removal. *IEEE Transactions on Image Processing*, 30: 7608–7619.

Zhong, X.; Tu, S.; Ma, X.; Jiang, K.; Huang, W.; and Wang, Z. 2022. Rainy WCity: A real rainfall dataset with diverse conditions for semantic driving scene understanding. In *International Joint Conference on Artificial Intelligence*, 1743–1749.