Semantic Segmentation in Multiple Adverse Weather Conditions with Domain Knowledge Retention

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

Semantic segmentation's performance is often compromised when applied to unlabeled adverse weather conditions. Unsupervised domain adaptation is a potential approach to enhancing the model's adaptability and robustness to adverse weather. However, existing methods encounter difficulties when sequentially adapting the model to multiple unlabeled adverse weather conditions. They struggle to acquire new knowledge while also retaining previously learned knowledge. To address these problems, we propose a semantic segmentation method for multiple adverse weather conditions that incorporates adaptive knowledge acquisition, pseudo-label blending, and weather composition replay. Our adaptive knowledge acquisition enables the model to avoid learning from extreme images that could potentially cause the model to forget. In our approach of blending pseudo-labels, we not only utilize the current model but also integrate the previously learned model into the ongoing learning process. This collaboration between the current teacher and the previous model enhances the robustness of the pseudo-labels for the current target. Our weather composition replay mechanism allows the model to continuously refine its previously learned weather information while simultaneously learning from the new target domain. Our method consistently outperforms the state-of-the-art methods, and obtains the best performance with averaged mIoU (%) of **65.7** and the lowest forgetting (%) of **3.6** against 60.1 and 11.3 (Hoyer et al. 2023), on the ACDC datsets for a four-target continual multi-target domain adaptation.

Introduction

Semantic segmentation methods face challenges in adverse weather conditions, as these conditions significantly degrade images. Adapting models to multiple adverse weather conditions in an unsupervised and successive manner causes additional difficulties due to substantial domain gaps among various weather conditions, and potentially leads to forgetting previously learned knowledge.

Continual unsupervised domain adaptation emerges as a potential solution to address these challenges by adapting the model from the labeled source domain to the unlabeled target domains in a sequential manner, e.g., ((Lin et al. 2022; Saporta et al. 2022)). However, these methods are designed
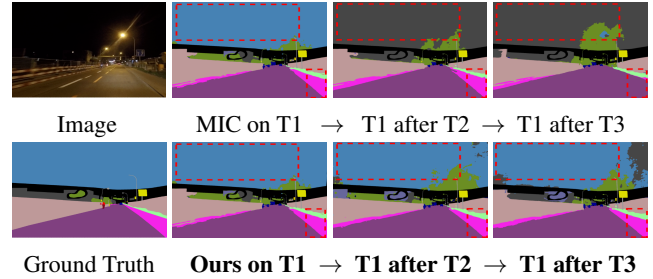
Figure 1: The illustration of our method, where the model adapts to each target domain sequentially. MIC (Hoyer et al. 2023) fails to retain previously learned knowledge, as its performance on the first target gradually deteriorates, e.g., the sky and the side walk in Target 1 are disappearing after the method learns Targets 2 and 3. Our method retain previously learned knowledge while adapting to new targets.

to acquire all information from the new target, without considering whether this information might lead to a forgetting of previously learned knowledge. Moreover, as mentioned in (Kalb and Beyerer 2023), domain shifts across distinct adverse weather conditions are primarily induced by the deteriorated information within low-level features extracted from the early convolutional layers. Consequently, there is a necessity to develop a method that takes this factor into consideration to effectively address the challenges posed by adverse weather conditions in domain adaptation.

In this paper, we present a semantic segmentation method that sequentially adapts the model to multiple unlabeled adverse weather domains, by progressively learning a new domain at a time while retaining the previous learned knowledge. Our method conceives three novel concepts: adaptive knowledge acquisition, pseudo-label blending, and weather composition replay.

In contrast to single-target unsupervised domain adaptation, our sequential domain adaptation aims to both learn from the new target and retain previously acquired knowledge simultaneously. Because of this, our model needs to identify potentially detrimental input regions that could introduce significant domain gaps and lead to forgetting of previously learned knowledge (Yang et al. 2022a; Kalb and Beyerer 2023). To achieve this, we introduce adaptive

knowledge acquisition by utilizing the previous model and class-wise feature representations, resulting in a dynamic weighting map. This dynamic weighting map acts as a constraint, preventing the current model from learning potentially detrimental areas.

Various adverse weather conditions may exhibit similar degradation effects (Kalb and Beyerer 2023; Li et al. 2023). For instance, both fog and the rain veiling effect visually look alike. Based on this similarity, models trained under different adverse weather conditions can collaborate to enhance learning from the shared degradation patterns (Allen-Zhu and Li 2020). Motivated by this idea, we propose a pseudo-label blending strategy. This involves employing the previous model as an auxiliary model to identify images from the new target that share similarities with those from the past targets. We then involve the auxiliary model (i.e., the previous model) to enhance our learning process in an ensemble manner. Note that, since we use a teacher-student framework, the term "previous model" refers to the previous teacher.

When operating in a sequential manner, we do not assume accessibility to the images from previously learned targets. In such a scenario, to maintain the previously acquired knowledge, we introduce a replay technique that serves as a continuous reminder of the weather degradation patterns encountered in the past. To implement this, we retain the weather information acquired from different targets in the previous steps, Once a new target is introduced, for each target image, we randomly augment each of the past weather information instances and integrate them into random segments of the current target images. Through training on these composite images, the model sustains and revises its comprehension of various weather degradations over time, even when direct access to previously learned target domain images is unavailable. In a summary, our contributions are as follows:

- We present an adaptive knowledge acquisition method that guides the model to refrain from learning new contents that could potentially result in forgetting.
- We introduce the concept of incorporating the previous model into the current learning process to enhance our method's overall performance in an ensemble manner.
- To retain the past weather information, we propose a continuous replay of previously learned weather degradation, by randomly augmenting and integrating it into the present target images.

Our method consistently outperforms the state-of-the-art methods, and obtains the best performance with averaged mIoU (%) of **65.7** and the lowest forgetting (%) of **3.6** against 60.1 and 11.3 (Hoyer et al. 2023), on the ACDC datsets for a four-target continual multi-target domain adaptation.

Related Work

Unsupervised domain adaptation (UDA) has been explored extensively in recent years, with many applications ranging from different vision tasks (Ganin and Lempitsky 2015; Chen et al. 2018, 2021; Vu et al. 2019; Saito et al. 2019;

Zou et al. 2018; Li, Yuan, and Vasconcelos 2019). However, UDA settings are limited to one source and one target, meaning that the trained domain adaptive model can only work on a certain target domain and will fail if more target domains are involved. Hence, some researchers start to explore methods to adapt a model into multiple target domains (Peng et al. 2019; Chen et al. 2019; Yu, Hu, and Chen 2018; Gholami et al. 2020; Nguyen-Meidine et al. 2021; Yao et al. 2022; Roy et al. 2021).

(Isobe et al. 2021; Saporta et al. 2021; Lee et al. 2022) adapt the model into multi-target domains in a parallel way, where all the target domains are involved in every iteration. (Isobe et al. 2021) maintains an expert model for each target domain, they are trained on many augmented images and teach a common student model via a knowledge distillation loss. (Saporta et al. 2021) proposes two MTDA methods, Multi-Dis and MTKT. For each target, Multi-Dis uses two types of domain classifiers, a source vs. target classifier and a target vs. all the other targets classifier. MTKT has a target domain-specific decoder, and a corresponding target-specific domain classifier for each target domain, the knowledge learned from different decoders is passed to a common decoder with knowledge distillation. (Lee et al. 2022) disentangles the input images into semantic contents and style contents, they obtain source images in target styles by swapping the contents. The drawbacks of the parallel approaches are, (1) They need additional modules/images for every target domain in every iteration, and the real-time memory consumption could be a bottleneck when the number of target domains becomes larger. (2) Similar to the single-target UDA model, when a new target domain is introduced, there is no way to update a pretrained model, and hence a new model will need to be trained from scratch with all the targets.

(Saporta et al. 2022) proposes a continual way to adapt a model to multiple target domains, where the target domains are adapted through multiple steps, and the model can be updated at any time when a new target domain is introduced. The paper also points out that due to this problem, the continual approach usually performs worse than the parallel approach.

Replay has been proven effective in addressing this preserving previously learned knowledge for adverse weather conditions (Kalb and Beyerer 2023). Existing replay techniques including storing representative exemplar images (Rebuffi et al. 2017; Hayes et al. 2020; Kang, Park, and Han 2022) and adversarially generating images in the style of previous domains (Shin et al. 2017). However, the weather degradation is closely coupled to the image's physics properties (e.g., depth) (Sindagi et al. 2020; Hu et al. 2021; Yang et al. 2022b; Li et al. 2023), so the degradation in each image is unique and it is hard to identify a representative exemplar image. Moreover, generating synthetic weather degradation is still an unsolved problem (Sakaridis, Dai, and Van Gool 2018; Anoosheh et al. 2019), and there does not exist a solution to continuously generate different adverse weather conditions.
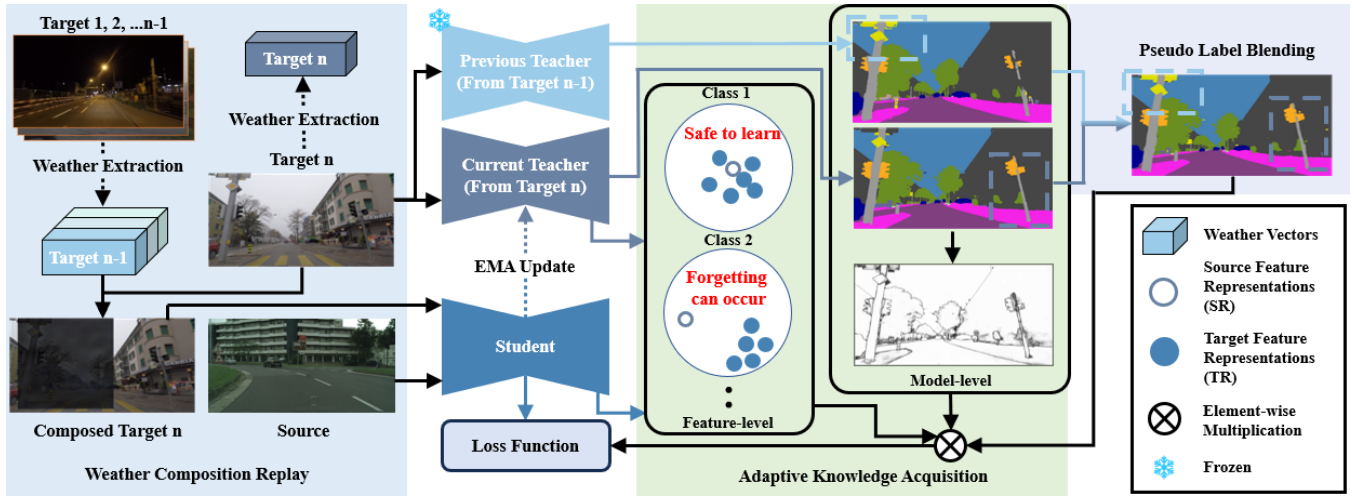
Figure 2: Our architecture for adapting a model to n adverse weather conditions in n steps in a sequential manner. The architecture consists of several key components: (1) Adaptive knowledge acquisition, where the model is guided to avoid learning the areas that could lead to a forgetting problem. (2) Pseudo-label blending, where the previous teacher is involved for enhancing the pseudo-label. (3) Weather composition replay, where the weather vectors from previous steps are composed into the current target image for revising on previously learned knowledge.

Proposed Method

Our goal is to adapt our model sequentially to multiple adverse weather conditions, focusing on one weather domain at a time, all the while retaining the knowledge gained previously. This sequential learning is necessary since we do not assume that once we have learned a domain, the data from that particular domain is accessible. To achieve this goal, we begin by adapting our model to an initial weather domain. Subsequently, we utilize the knowledge gained from the first domain (previous domain), which includes the teacher model and weather vectors, to aid the model's adaptation to the second weather domain (current domain). This process of acquiring and retaining knowledge is iterated for each weather domain.

Fig. 2 shows our pipeline. We progressively adapt our model to n distinct adverse weather domains over n steps. We feed the source images to the student model in each step for supervised learning on semantic segmentation. As for the target parts, from step 1 to step $n - 1$, our weather composition replay extracts and preserves weather vectors for every previously encountered target domain. Then, in step n, we compose the extracted weather vectors into the current target image. This composite image is subsequently fed into the student model for making predictions.

We combine pseudo-labels generated by both the current teacher model and the previous teacher model to synergistically improve the quality of pseudo-labels. Our adaptive knowledge acquisition assesses pixel-wise domain shifts from both model-level and feature-level viewpoints. This dynamic reassessment enables us to adjust the learning process to the current target image, helping the model avoid incorporating detrimental information that may result in significant forgetting.

Adaptive Knowledge Acquisition

In this process of adaptive knowledge acquisition, we identify image regions with significant domain shifts that could potentially trigger model forgetting. Subsequently, we dynamically adjust the weighting of these areas, ensuring that while acquiring new knowledge, the model can still retain past knowledge.

Model-Level Adjustment Following the standard unsupervised domain adaptation methods (Kennerley et al. 2023; Hoyer et al. 2023), in each step n (weather domain n), we obtain a teacher model through the EMA (Exponential Moving Average) of a student model. We then adapt the student model to the new target domain by learning from the soft pseudo-labels generated by the teacher.

In this standard process, the student will quickly adapt to the new target domain. However, certain images in the new target domain can result in forgetting previous knowledge, especially for images that differ significantly from those in the previous target domain. To mitigate this potential risk, we incorporate the teacher model from a previous target domain as an auxiliary model. This teacher model has not been exposed to the new target images during training. We refer to this teacher model as *Previous-Teacher*.

Our Previous-Teacher makes a prediction for each image in the current step. Consequently, the confidence of our Previous-Teacher in areas affected by the new adverse weather degradation can be used as an indication of the degree of the domain shift. Since, its confidence is low for areas that have never been learned before. This allows our model to selectively learn from the new target, avoiding harmful (low confidence) areas and thereby reducing the risk of forgetting previously acquired knowledge (Kalb and Beyerer 2023). To implement this idea, we define confidence as

q:

$$q(x)_{ij} = \max(g(x)_{ij}), \qquad (1)$$

where, g is the model and x is the input.

We compute a pixel-wise dynamic weighting mask for the model-level adjustment, denoted as M^{mod}, as follows:

$$M_{ij}^{\mathrm{mod}} = (1-\alpha)q_{\mathrm{cur}}(x^T)_{ij} + \alpha q_{\mathrm{pre}}(x^T)_{ij}, \qquad (2)$$

where x^T represents the target image, q_{pre} and q_{cur} represent the confidence scores from the previous teacher and the current teacher, respectively. α is a parameter to control the weights of each term, where it is large initially, and reduces along the number of iterations. For regions where Previous-Teacher exhibits significant low confidence, M^{mod} dynamically reduces the learning weight in those areas. As a result, the model is encouraged to prioritize exploration in the regions that are less likely to lead to forgetting. As the model progressively strengthens its resilience to new weather degradations, the confidence $q_{\mathrm{cur}}(x^T)_{ij}$ in these challenging regions increases. Simultaneously, with a decrease in the parameter α, the model becomes more inclined to adaptively learn from these areas.

Feature-Level Adjustment While model-level adjustment focuses on utilizing teacher models to guide the model's learning, feature-level adjustment aims to utilize class-specific feature representations to quantify the domain shift caused by weather degradation.

When a model attempts to adapt to a new adverse weather domain, the low-level features it extracts from the early convolutional layers are prone to degradation due to the new weather conditions. Consequently, predictions relying on these imprecise features could result in erroneous outcomes and induce substantial domain shifts within the feature space. In contrast, a model capable of accurately extracting information should avoid notable domain shifts in the feature space. Therefore, we leverage this feature information to assess the extent of domain shifts.

To implement this, we first define the feature representations. Based on the ground-truth of a source image, we partition the feature maps for different classes and calculate their average for each class, yielding class-wise source feature vectors. We then compute the exponential moving average (EMA) of the source feature vectors, denoted as SR. Compared to the source feature vectors, SR can filter out images that are different from the typical source images, and thus can provide more representative features.

Similarly, we partition and calculate the average of the target feature maps according to the corresponding pseudo-labels, resulting in target feature vectors. As our objective involves dynamically assigning weights to each target image, we directly employ the target feature vectors as the target feature representation, denoted as TR, for every image. Following this, we construct a weighting map based on the distances between SR and TR:

$$M^{\mathrm{feat}} = \max(0, 1 - \frac{(\mathrm{TR}_{c_1} - \mathrm{SR}_{c_1})^2}{\sum_{c=1}^{C}(\mathrm{TR}_c - \mathrm{SR}_{c_1})^2}), \qquad (3)$$

where c_1 represents the predicted class, and C is the number of all the predicted classes. In this equation, we calculate the relative distance between SR and TR for class c_1.

Intuitively, as shown in Fig. 2, within the feature space, for the target feature representations of Class 1, if they closely resemble the source feature representations of the same class, the domain shift is probably minor. This implies that learning in these areas would result in lesser chances of significant forgetting. Conversely, when we notice that the target feature representations are considerably distant from the source feature representations of the same class (as illustrated in Class 2 in the figure), it indicates that the model might have captured degraded information from the image. Training within these regions can lead to substantial alterations in the feature space, potentially affecting the pretrained feature structures (Hoyer, Dai, and Van Gool 2022; Su et al. 2023). In such scenarios, our objective is to mitigate the possibility of forgetting by constraining the model from learning based on this information.

Pseudo-Label Blending

By integrating Previous-Teacher with the current teacher model, we can effectively ensemble them to enhance the robustness of the pseudo-labels (Allen-Zhu and Li 2020). We achieve this using both Previous-Teacher and its target feature representations. First, we utilize both the Previous-Teacher and the current teacher model to generate predictions for the current target image. Based on the confidences of these predictions, we generate a binary mask, denoted as M^{con}, which indicates the regions where the Previous-Teacher exhibit higher confidence compared to the current teacher,

$$M_{ij}^{\mathrm{con}} = \begin{cases} 1, & \text{if } q_{\mathrm{pre}}(x^T)_{ij} > q_{\mathrm{cur}}(x^T)_{ij} \\ 0, & \text{otherwise} \end{cases}, \qquad (4)$$

As discussed in the feature-level adjustment, if Previous-Teacher demonstrates greater robustness to an area than the current teacher for a specific image segment, the corresponding extracted target feature representations should be closer to the source feature representations. Hence, we also compute a weighting map using the equation mentioned in Eq. (3), but using the target feature representations generated from Previous-Teacher. We denote this weighting map as $M_{\mathrm{pre}}^{\mathrm{feat}}$. We can then obtain the refined pseudo-label, denoted as p.

$$p(x^T)_{ij} = \underset{c}{\arg\max}(q_{\mathrm{cur}}(x^T)_{ijc} + M^{\mathrm{con}}M_{\mathrm{pre}}^{\mathrm{feat}}q_{\mathrm{pre}}(x^T)_{ijc}). \qquad (5)$$

We incorporate reliable predictions from the previously learned models into the current learning process. This blending of pseudo-labels enables the model to effectively learn the similar patterns in different targets, leveraging the knowledge and expertise accumulated by both Previous-Teacher and the current teacher model.

Weather Composition Replay

We propose a replay technique utilizing the weather information obtained from different weather domains. As mentioned in (Kalb and Beyerer 2023), the domain shifts introduced by weather degradation can be captured in the frequency domain by the amplitude spectrum. Hence, when

Figure 3: Examples of composing night and fog weather vectors into snow and rain images, respectively.

adapting to a new target domain, we begin by translating each new target image into the frequency domain, and separate the amplitude from the phase. Although each target image may contain some image-specific information, they are all subjected to the same adverse weather conditions. By averaging the extracted amplitude, we can then suppress image-specific information while preserving the weather vectors, in frequency domain.

As illustrated in Fig. 2, once the weather vectors are extracted, we keep them to future steps. In step n, we apply augmentations on the stored weather vectors from step 1 to step $n - 1$, and randomly inject them into the current target images as follow:

$$x_r^{T2} = M_r x^{T2} + (1 - M_r)\text{iFT}(P^{T2}, \sigma A^{T1}), \quad (6)$$

where, x^{T2} is the current target image, x_r^{T2} is the composed images for the model to learn, P^{T2} and A^{T1} represent the phase component of the current image and the weather vectors of the previous domain, respectively. σ represents a randomly determined volume of A^{T1}, where a larger volume indicates a stronger dominance of the previous weather information in the area. iFT represents an inverse Fourier Transformation to translate back the images from frequency domain. M_r represents a random segment in that image, as augmenting a segment can be more efficient than augmenting the whole image (Yun et al. 2019; Olsson et al. 2021). This process is repeated for each weather vector we obtained, for a total of $n - 1$ times.

This replay technique, utilizing the composed images, ensures that the model continuously updates its understanding of adverse weather conditions while retaining and refining its knowledge from past experiences.

Experimental Results

In this section, we present a comprehensive evaluation of our method under multiple adverse weather conditions in a sequential setting. We begin by describing the datasets used in our experiments, followed by the architectures and parameters employed. Next, we assess our model's performance both quantitatively and qualitatively, demonstrating its effectiveness in handling diverse weather conditions. Lastly, we conduct ablation studies to evaluate the importance of each factor in our method.

Datasets We utilize Cityscapes (Cordts et al. 2016) as our source domain, consisting of real-world street scene images captured under daytime, clear weather conditions. For the target dataset, we employ ACDC (Sakaridis, Dai, and

Van Gool 2021), which contains real-world street scene images captured under four adverse weather conditions: nighttime, rain, fog, and snow. In our experiments, each of these four adverse weather conditions is treated as a separate target, and our models are adapted to these targets sequentially.

Baseline Models In our experiments, we conduct comparisons with the state-of-the-art unsupervised domain adaptation method, MIC (Hoyer, Dai, and Van Gool 2022). To ensure a fair comparison, we use the same architecture, DAFormer, in all comparisons. Additionally, we employ identical optimization strategy, including the number of epochs, batch sizes, domain adaptation techniques and the pretrained backbone, as suggested in MIC.

Since our method is not limited by the model's architecture, we also compare it with two parallel and one continual general-purpose multiple target domain adaptation methods, MTKT, Multi-Dis, and MuHDi (Saporta et al. 2021, 2022). Note, for the parallel setting methods (Saporta et al. 2021), they learn all the targets simultaneously, their models are not affected by the forgetting problem. For the continual setting method (Saporta et al. 2022), it provides a forgetting prevention technique, but this method is not designed specifically for adverse weather conditions. Once again, to maintain fairness in the comparison, we utilize the same architecture, DeeplabV2 (Chen et al. 2017), and employ identical optimization strategy, domain adaptation techniques and pretrained backbone, following (Saporta et al. 2022).

Regarding our method's parameters, we initialize the value of α to be 0.8 and gradually decrease it to 0.2 as the number of iterations progresses. When injecting the previous weather information into the new target image, we randomly generate σ between 0.2 and 1.2, and the sizes of the affected area are randomly selected in the target image, ranging from one-third to half of the image size. All these parameters are decided empirically.

Evaluation Metrics Cityscapes and ACDC have the same segmentation class protocols, hence we apply the same protocol in our evaluation. We use the percentage of Intersection over Union (IoU %) as our evaluation metric for the effectiveness of the knowledge acquisition from a new target, the higher the better ($\uparrow$). As for the knowledge retention, we use *Accumulated Forgetting*, which is calculated as follow,

$$A.F. = \sum_{k=1}^{K-1} (mIoU_{k,k} - mIoU_{k,K}), \quad (7)$$

where, K is the number of targets. We adapt the model to K targets in K steps. $mIoU_{k,k}$ represents the initial performance of target k at step k, and $mIoU_{k,K}$ represents the final performance of target k at the last step, K. A smaller Accumulated Forgetting indicates a lesser degree of forgetting in the model ($\downarrow$).

Quantitative Results

We conduct experiments on four ACDC adverse weather conditions: nighttime, rain, fog, and snow. As shown in Tab. 1, our models outperformed other methods on all targets. Specifically, our models surpassed AdvEnt (Vu et al.

		Cityscapes→Night→Rain→Fog→Snow					
Method	Forgetting Prevention	Night	Rain	Fog	Snow	mIoU Avg. ↑	A.F. ↓
MTKT (Saporta et al. 2021)	-	21.5	39.4	48.8	38.7	37.1	-
Multi-Dis (Saporta et al. 2021)	-	20.5	38.5	43.6	36.8	34.8	-
AdvEnt (Vu et al. 2019)	✗	16.9 (-9.0)	36.9 (-3.2)	40.8 (-10.3)	35.1	32.4	22.4
MuHDi (Saporta et al. 2022)	✓	17.6 (-8.3)	37.2 (-3.4)	44.4 (-6.7)	36.3	33.9	18.4
Ours (DeeplabV2)	✓	**24.0 (-1.9)**	**42.0 (-1.3)**	**50.8 (-1.1)**	**44.0**	**40.2**	**4.3**
MIC (Hoyer et al. 2023)	✗	34.7 (-7.2)	65.8 (-2.8)	78.4 (-1.3)	65.2	60.1	11.3
Ours (DAFormer)	✓	**39.0 (-2.9)**	**70.6 (-0.9)**	**80.4 (-0.2)**	**72.6**	**65.7**	**3.6**

Table 1: Quantitative results of Ours compare to the existing unsupervised domain adaptation methods and continual multiple target domain adaptation methods, evaluated against four targets, ACDC nighttime, rain, fog and snow. Bold numbers are the best scores for different backbones. The mIoU (%) of each target, the mIoU average of all the targets (the higher the better), and the accumulated forgetting, A.F. (the lower the better) are presented. The number in parentheses ')' indicates changes in performance, with a smaller number indicating a more pronounced forgetting effect. Our method outperforms the best existing method with forgetting prevention on DeeplabV2 backbone, by 6.3 mIoU (%) in average across all the targets, and 14.1 in the accumulated forgetting. As for the DAFormer backbone, our method outperforms the best domain adaptation method by 5.6 mIoU (%) in average across all the targets, and 7.7 in the accumulated forgetting.

2019) and MIC (Hoyer et al. 2023) by 8.2 mIoU (%) and 5.6 mIoU (%) in mIoU Avg., and 18.1 and 7.7 in accumulated forgetting, respectively. The highest mIoU and the smallest forgetting across all previous targets indicate the effectiveness of our models' knowledge acquisition and retention when adapting to different targets.

We compared our method to MuHDi (Saporta et al. 2022), which offers a general-purpose forgetting prevention technique for multiple targets. Our method outperformed MuHDi by 6.3 mIoU (%) in mIoU Avg. and 14.1 in accumulated forgetting. Moreover, while the parallel multiple target domain adaptation methods do not suffer from the forgetting problem, our model still outperforms their performance in mIoU Avg. by 3.1 mIoU (%) and 5.4 mIoU (%), respectively. This highlights the importance of our method's weather-specific knowledge acquisition and retention.

Without Source In certain circumstances, access to the source data may not be available in different steps. In this section, we evaluate our method when we can only access the current target domain in each step. Since the source datasets are inaccessible, we do not apply M^{feat} and $M^{\text{feat}}_{\text{pre}}$ in this scenario. We use MIC as the baseline for comparison. Note, for both MIC and our method without source, we provide a model finetuned on the source, following (Kalb and Beyerer 2023). Both methods are required to adapt this model to different adverse weather conditions. For AdvEnt backbone, it requires the source dataset for domain adaptation, so models with this backbone are not involved.

The results are presented in Tab. 2, with the absence of the source image dataset, MIC's accumulated forgetting increased significantly from 11.3 (%) to 23.9 (%), by 12.6 (%), where our model is less affected, with only an increase of accumulated forgetting by 6.7 (%).

Qualitative Results

We first show our qualitative results of our knowledge acquisition ability in Fig. 4, for a four-target sequential do-

		Cityscapes→Night→Rain→Fog→Snow				
Method	Night	Rain	Fog	Snow	mIoU Avg. ↑	A.F. ↓
MIC (w/o)	30.8	63.0	74.5	69.2	57.5	23.9
Ours (w/o)	36.7	69.8	75.9	69.6	63.0	10.3

Table 2: Quantitative comparisons between MIC (Hoyer et al. 2023) and our method, both without source (w/o).

Cityscapes→Night→Rain→Fog→Snow		
Method	mIoU Avg. ↑	A.F. ↓
AdvEnt	32.4	22.4
Blending	34.0	19.6

Table 3: Ablation studies of knowledge acquisition with pseudo-label blending.

main adaptation (on night, rain, fog, and snow). We evaluate our best model Ours (DAFormer) against MIC, and the ground truths semantic segmentation maps. We can see that our method can predict more precise semantic segmentation maps, and also has fewer false positives compared to MIC in all the target domains. More qualitative results are provided in the supplementary materials.

Ablation Studies

Knowledge Acquisition and Retention We begin by evaluating the effectiveness of our adaptive knowledge acquisition and weather composition replay techniques, as they are specifically designed for both knowledge acquisition and knowledge retention. In Tab. 3, we report the performance changes of four models: AdvEnt, AdvEnt with model-level assistance, AdvEnt with both model-level and feature-level assistance, and AdvEnt with adaptive knowledge acquisition and weather composition replay tech-

| road | sidew. | build. | wall | fence | pole | tr. light | tr. sign | veget. | terrain | sky | person | rider | car | truck | bus | train | m.bike | bike | n/a. |

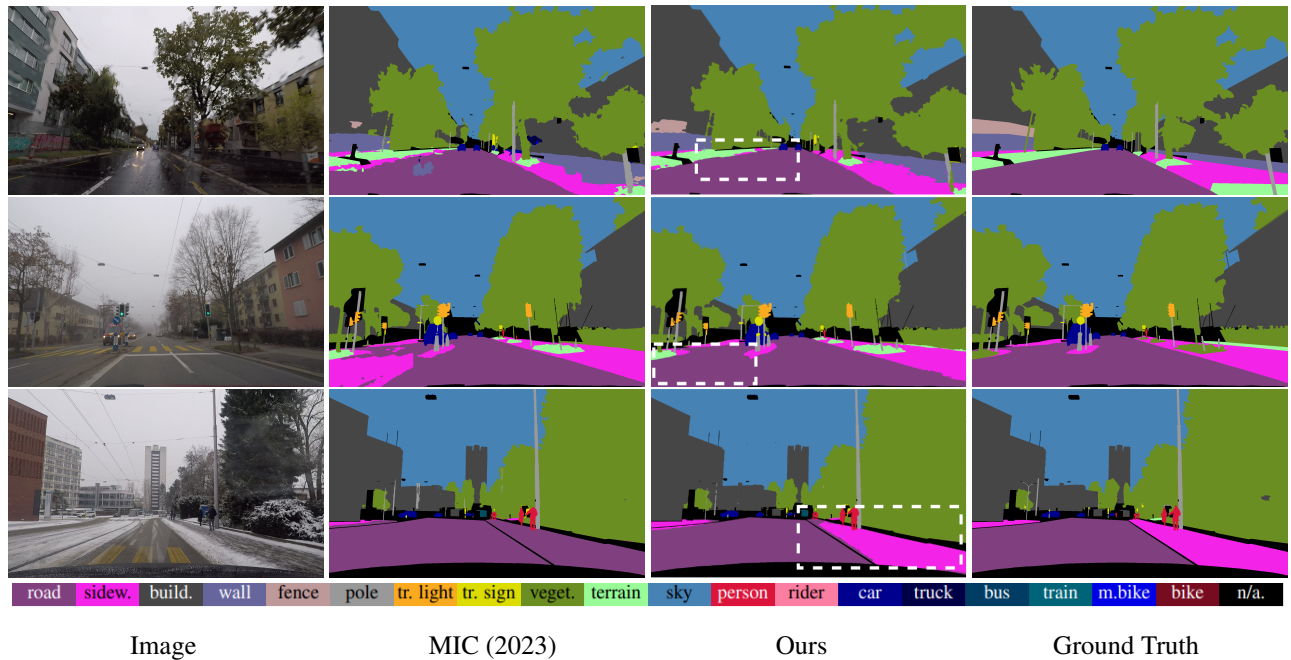| Image | MIC (2023) | Ours | Ground Truth |

Figure 4: Comparisons on the semantic segmentation performance with MIC (Hoyer et al. 2023), Ours (DAFormer), and ground truths on ACDC (Val.) under rainy, foggy, and snowy weather conditions following sequential multitarget domain adaptation.

Cityscapes→Night→Rain→Fog→Snow			
Progressive		Replay	A.F. ↓
Model Level	Feature Level		
			22.4
✓			12.9
✓	✓		9.6
✓	✓	✓	4.5

Table 4: Ablation studies of our knowledge retention techniques, including adaptive knowledge acquisition and weather composition replay.

niques. All these models are trained to adapt to four adverse weather conditions using source data, and it is evident that each component contributes to preventing forgetting.

Pseudo-Label Blending Pseudo-label blending is designed for exploring the similar patterns among different targets to improve the knowledge acquisition on the current target. We evaluate the knowledge acquisition with two approaches: AdvEnt and AdvEnt with pseudo-label blending. The results are presented in Tab. 3. We can observe that pseudo-label blending enhances the average mIoU across the four targets, without causing additional forgetting issues.

Combining all our proposed techniques, we achieve an enhanced performance on the new target, while retain the most knowledge from the previous targets.

Conclusion

We have proposed a novel method that adapts a model to multiple unlabeled adverse weather conditions sequentially. We use both model-level and feature-level knowledge to assist the model avoid learning from the harmful contents from the new target image that can lead to a forgetting of the previously learned knowledge. To support our current learning process, we have also proposed a method to involve the previously obtained model to jointly improving the pseudo-labels for the current target. We propose a weather composition replay technique, which compose the previously learned weather information to the current target image, enabling the model learn from the current target image while revising the previously learned weather information. We train two models using our method on both DeeplabV2 and DAFormer, respectively, to demonstrate that our method can be generalized to different architectures. We evaluate these models on several benchmark adverse weather conditions with different settings and found that our models outperform many methods in different settings.

References

Allen-Zhu, Z.; and Li, Y. 2020. Towards understanding ensemble, knowledge distillation and self-distillation in deep learning. *arXiv preprint arXiv:2012.09816*.

Anoosheh, A.; Sattler, T.; Timofte, R.; Pollefeys, M.; and Van Gool, L. 2019. Night-to-day image translation for retrieval-based localization. In *2019 International Conference on Robotics and Automation (ICRA)*, 5958–5964. IEEE.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4): 834–848.

Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3339–3348.

Chen, Y.; Wang, H.; Li, W.; Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. Scale-aware domain adaptive faster r-cnn. *International Journal of Computer Vision*, 129(7): 2223–2243.

Chen, Z.; Zhuang, J.; Liang, X.; and Lin, L. 2019. Blending-target domain adaptation by adversarial meta-adaptation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2248–2257.

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 conference on computer vision and pattern recognition*, 3213–3223.

Ganin, Y.; and Lempitsky, V. 2015. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, 1180–1189. PMLR.

Gholami, B.; Sahu, P.; Rudovic, O.; Bousmalis, K.; and Pavlovic, V. 2020. Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Transactions on Image Processing*, 29: 3993–4002.

Hayes, T. L.; Kafle, K.; Shrestha, R.; Acharya, M.; and Kanan, C. 2020. Remind your neural network to prevent catastrophic forgetting. In *European Conference on Computer Vision*, 466–483. Springer.

Hoyer, L.; Dai, D.; and Van Gool, L. 2022. 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.; 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.

Hu, X.; Zhu, L.; Wang, T.; Fu, C.-W.; and Heng, P.-A. 2021. Single-image real-time rain removal based on depth-guided non-local features. *IEEE Transactions on Image Processing*, 30: 1759–1770.

Isobe, T.; Jia, X.; Chen, S.; He, J.; Shi, Y.; Liu, J.; Lu, H.; and Wang, S. 2021. Multi-target domain adaptation with collaborative consistency learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8187–8196.

Kalb, T.; and Beyerer, J. 2023. Principles of Forgetting in Domain-Incremental Semantic Segmentation in Adverse Weather Conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19508–19518.

Kang, M.; Park, J.; and Han, B. 2022. Class-incremental learning by knowledge distillation with adaptive feature consolidation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16071–16080.

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.

Lee, S.; Choi, W.; Kim, C.; Choi, M.; and Im, S. 2022. ADAS: A Direct Adaptation Strategy for Multi-Target Domain Adaptive Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19196–19206.

Li, M.; Xie, B.; Li, S.; Liu, C. H.; and Cheng, X. 2023. 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.

Li, Y.; Yuan, L.; and Vasconcelos, N. 2019. Bidirectional learning for domain adaptation of semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6936–6945.

Lin, H.; Zhang, Y.; Qiu, Z.; Niu, S.; Gan, C.; Liu, Y.; and Tan, M. 2022. Prototype-guided continual adaptation for class-incremental unsupervised domain adaptation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, 351–368. Springer.

Nguyen-Meidine, L. T.; Belal, A.; Kiran, M.; Dolz, J.; Blais-Morin, L.-A.; and Granger, E. 2021. Unsupervised multi-target domain adaptation through knowledge distillation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1339–1347.

Olsson, V.; Tranheden, W.; Pinto, J.; and Svensson, L. 2021. Classmix: Segmentation-based data augmentation for semi-supervised learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1369–1378.

Peng, X.; Huang, Z.; Sun, X.; and Saenko, K. 2019. Domain agnostic learning with disentangled representations. In *International Conference on Machine Learning*, 5102–5112. PMLR.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2001–2010.

Roy, S.; Krivosheev, E.; Zhong, Z.; Sebe, N.; and Ricci, E. 2021. Curriculum graph co-teaching for multi-target domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5351–5360.

Saito, K.; Ushiku, Y.; Harada, T.; and Saenko, K. 2019. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6956–6965.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126: 973–992.

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.

Saporta, A.; Douillard, A.; Vu, T.-H.; Pérez, P.; and Cord, M. 2022. Multi-Head Distillation for Continual Unsupervised Domain Adaptation in Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3751–3760.

Saporta, A.; Vu, T.-H.; Cord, M.; and Pérez, P. 2021. Multi-target adversarial frameworks for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9072–9081.

Shin, H.; Lee, J. K.; Kim, J.; and Kim, J. 2017. Continual learning with deep generative replay. *Advances in neural information processing systems*, 30.

Sindagi, V. A.; Oza, P.; Yasarla, R.; and Patel, V. M. 2020. Prior-based domain adaptive object detection for hazy and rainy conditions. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, 763–780. Springer.

Su, W.; Han, Z.; He, R.; Wei, B.; He, X.; and Yin, Y. 2023. Neighborhood-based credibility anchor learning for universal domain adaptation. *Pattern Recognition*, 142: 109686.

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. 2022a. 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.

Yang, Y.; Wang, C.; Liu, R.; Zhang, L.; Guo, X.; and Tao, D. 2022b. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2037–2046.

Yao, C.-H.; Gong, B.; Qi, H.; Cui, Y.; Zhu, Y.; and Yang, M.-H. 2022. Federated multi-target domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1424–1433.

Yu, H.; Hu, M.; and Chen, S. 2018. Multi-target unsupervised domain adaptation without exactly shared categories. *arXiv preprint arXiv:1809.00852*.

Yun, S.; Han, D.; Oh, S. J.; Chun, S.; Choe, J.; and Yoo, Y. 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6023–6032.

Zou, Y.; Yu, Z.; Kumar, B.; and Wang, J. 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, 289–305.