VBLC: Visibility Boosting and Logit-Constraint Learning for Domain Adaptive Semantic Segmentation under Adverse Conditions

Mingjia Li[1*], Binhui Xie[1*], Shuang Li[1†], Chi Harold Liu[1], Xinjing Cheng[2,3]

[1]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[2]School of Software, BNRist, Tsinghua University, Beijing, China
[3]Inceptio Technology, Shanghai, China
{mingjiali, binhuixie, shuangli, chiliu}@bit.edu.cn, cnorbot@gmail.com

Abstract

Generalizing models trained on normal visual conditions to target domains under adverse conditions is demanding in the practical systems. One prevalent solution is to bridge the domain gap between clear- and adverse-condition images to make satisfactory prediction on the target. However, previous methods often reckon on additional reference images of the same scenes taken from normal conditions, which are quite tough to collect in reality. Furthermore, most of them mainly focus on individual adverse condition such as nighttime or foggy, weakening the model versatility when encountering other adverse weathers. To overcome the above limitations, we propose a novel framework, Visibility Boosting and Logit-Constraint learning (VBLC), tailored for superior normal-to-adverse adaptation. VBLC explores the potential of getting rid of reference images and resolving the mixture of adverse conditions simultaneously. In detail, we first propose the *visibility boost module* to dynamically improve target images via certain priors in the image level. Then, we figure out the over-confident drawback in the conventional cross-entropy loss for self-training method and devise the *logit-constraint learning*, which enforces a constraint on logit outputs during training to mitigate this pain point. To the best of our knowledge, this is a new perspective for tackling such a challenging task. Extensive experiments on two normal-to-adverse domain adaptation benchmarks, i.e., Cityscapes → ACDC and Cityscapes → FoggyCityscapes + RainCityscapes, verify the effectiveness of VBLC, where it establishes the new state of the art. Code is available at https://github.com/BIT-DA/VBLC.

Introduction

The past few years have witnessed predominance of deep learning based methods in fundamental vision tasks, where scene understanding under extreme vision conditions has been attracting substantial research interest (Ma et al. 2022; Sakaridis, Dai, and Gool 2022). In many outdoor applications, adverse weather conditions are frequently encountered, causing poor visibility and performance degradation. For a safer, smoother operating environment, a desirable perception system should be trustworthy under a wide variety of scenarios (Zhang et al. 2021b; Sakaridis, Dai, and Van Gool 2021). But, existing studies are mostly centered around datasets consisting of clean images, yet ignore the challenge of driving in varying adverse weather conditions, making them vulnerable in practice. Meanwhile, it is implausible to collect a dataset that fully reflects all situations and then separate data into discrete domains, since the visual appearance changes overtime and depends on the specific location, season and many other factors, all of which introduce a natural domain shift between any training (source) and test (target) distributions.

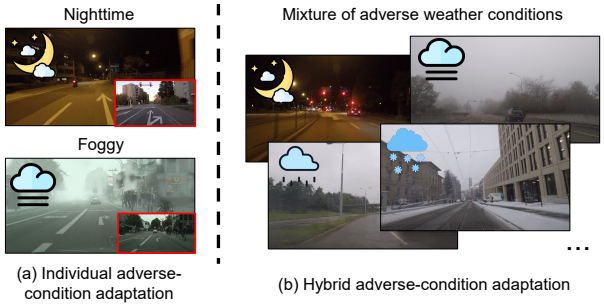


Figure 1: Problem comparison. (a) Individual adverse-condition adaptation: reference images (shown in the red box at bottom right corner) depicting a similar scene are leveraged as an intermediate domain to assist in a specific adverse condition, e.g., nighttime or foggy. (b) Hybrid adverse-condition adaptation: the mixture of images from multiple adverse conditions are used without any reference.

Accordingly, the emergence of more robust models for adverse conditions is vital to paving the way for their real-world utility. Unsupervised domain adaptation (UDA) is an alternative method (Ganin et al. 2016; Tzeng et al. 2017; Long et al. 2018; Liu et al. 2020; Park et al. 2020; Li et al. 2022) to adapt models trained with well-labeled clear (source) images to adverse (target) images without access to target annotations. Until now, nighttime image segmentation and foggy scene understanding are two mainstream tasks. Given the difficulties of both specific problems, a great deal of works, such as (Sakaridis, Dai, and Gool 2022; Dai et al. 2020; Wu et al. 2021; Ma et al. 2022), are carefully designed and highly customized, with the significant prior knowledge, e.g., additional clear-condition images. In Fig. 1(a),

*These authors contributed equally.
†Corresponding author.

reference images (red boxes), depicting the similar scenes in correspondence with adverse images, are meant to boost segmentation performance. Unfortunately, it is no picnic to gather exactly paired images in the rapidly changing driving scenes. On the other hand, such a clear and specific distinction among adverse conditions is hard to define, e.g., test images are continually varying, which could be collected in composite conditions.

Driven by the above analysis, we advocate a new approach without extra clear images for reference, dubbed as Visibility Boosting and Logit-Constraint learning (VBLC). It's worth noting that, this setting is generally regarded as under-constrained, making it quite difficult and rarely researched into. Take it a step further, we concentrate on a much more practical scenario where input images feature a hybrid of adverse conditions, i.e., low-light and flare characteristics of nighttime, veiling effects formed by heavy rain, dense foggy, snow, and so on (see Fig. 1(b)).

To begin with, we introduce the *visibility boost module* in the input space to close the gap between normal- and adverse-condition images without the reliant on normal-adverse image pairs for reference. The absence of such weak supervision urges us to make the most of the priors as a replacement. We provide a saturation-based prior to adaptively heighten the visibility of incoming images. On top of that, boosted images are incorporated during training to bridge the immense gap brought about by adverse conditions.

Second, for the self-training schemes prevailing in UDA, we observe the insufficient exploitation of predictions on unlabeled target samples for fear of overconfidence (Wei et al. 2022). To resolve this, we further propose the *logit-constraint learning* to relieve the stringent demand on the quality of pseudo labels. Through gradient analysis, we demonstrate that the constraint on logits can slow down the trend towards overconfidence and capitalize on predictions.

Eventually, we show that VBLC establishes state-of-the-art performance on two challenging benchmarks. Compared to the current SOTA method, VBLC improves the relative performance by 8.9% and 4.9% (mIoU) on Cityscapes $\rightarrow$ ACDC and Cityscapes $\rightarrow$ FoggyCityscapes + RainCityscapes, respectively. We summarize contributions below:

- We tackle a more realistic and challenging task of domain adaptive semantic segmentation under adverse conditions without the aid of extra image counterparts to form a clear-adverse pair.
- We desire to fill the blank through making adjustments at both ends of the network. The *visibility boost module* is proposed to narrow the visibility gap in the input space, while the *logit-constraint learning* is included in the output space to handle the overconfidence issue.
- We justify the effectiveness of our VBLC and explore the mechanism behind its success via extensive experiments.

Related Work

Normal-to-Adverse Domain Adaptation. Domain adaptation has been well investigated in both theory (Ben-David et al. 2010) and practice (Wang and Deng 2018). Here, we are particularly interested in semantic segmentation task. *Adversarial training* is the most examined method that narrows the domain gap via style transfer (Hoffman et al. 2018) or learning indistinguishable representations (Tsai et al. 2018; Vu et al. 2019; Kim and Byun 2020). Recently, *self-training* methods turn to pseudo labels to acquire extra supervision, reaching better performance. Advanced practices are to improve the quality of pseudo labels (Zou et al. 2018, 2019), stabilize the training process (Tranheden et al. 2021; Hoyer, Dai, and Gool 2022), or utilize auxiliary task (Wang et al. 2021a; Xie et al. 2022).

Despite the rising interest in developing domain adaptation models, existing works mostly concentrate on handling domain shifts introduced by the limitations of scene synthesis or by visual differences due to the variation in shooting locations. Considerably fewer attempts have been made to mitigate the shifts posed by adverse conditions, which is especially critical in reality (Sakaridis, Dai, and Van Gool 2021; Liu et al. 2022). Equipped with abundant data from various domains, several methods resort to curriculum-based schemes to realize a progressive adaptation towards a distant target domain (Wulfmeier, Bewley, and Posner 2018; Sakaridis et al. 2018). The dilemma of this scheme is that manually assigned intermediate domains may be suboptimal or arduous to design. Another promising direction is to make the best of the corresponding image pairs in dataset. Ma et al. (2022) decouple style factor, fog factor and dual factor to cumulatively adapt these three factors. Alternatively, pixel-level warping is employed to benefit the prediction of static classes (Wu et al. 2021), enable multi-view prediction fusion (Sakaridis, Dai, and Gool 2022), or guide the subsequent label correction (Bruggemann et al. 2022).

In general, the above methods require corresponding clear images, while the setting excluding image correspondences has rarely been explored. In this work, we are capable of addressing arbitrary adverse conditions without leveraging such weak supervision for adaptation.

Multi-Target Domain Adaptation/Generalization. The goal is to extend domain adaptation to multiple target domains, which is relevant to our work. Works in this field usually employ domain transfer (Lee et al. 2022), knowledge distillation (Isobe et al. 2021), curriculum learning (Liu et al. 2020) or meta-learning (Gong et al. 2021) to bootstrap generalization across domains. To name a few, Park et al. (2020) decompose a hard problem into multiple easy single-target adaptation problems. Lee et al. (2022) perform style transfer in the input space and utilize a direct adaptation strategy towards multiple domains. Our approach differs from these methods in the way we treat the target samples, where images under widely varied adverse scenarios are viewed as a mixture of multiple domains, and domain labels in the test set are unavailable.

Poor Visibility Image Enhancement. There is a significant body of works whose goal is increasing image visibility. Research interest in the mechanism of haze formation (McCartney 1976) has emerged long before the neural nets prevail. Since then, numerous conditions featuring poor visibility are described including low-light (Land 1977), fog (Narasimhan and Nayar 2003), rain (Li, Cheong, and

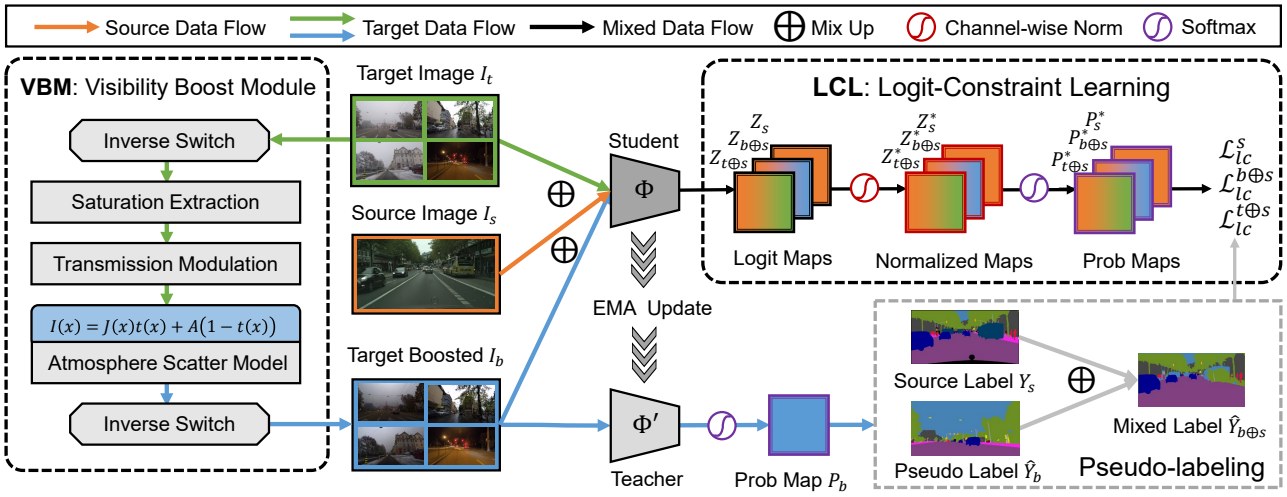


Figure 2: Overview of VBLC. Our framework enhances the capability of self-training schemes at both ends of the pipeline. In the input space, the *visibility boost module* is incorporated to ameliorate target images and generate more reliable pseudo labels. In the output space, the specialized *logit-constraint learning* is devised to conquer the erroneous prediction brought about by tremendous domain gap. Together with slight modifications to the training scheme, a simple, competitive approach is proposed.

Tan 2019), and snow (Chen et al. 2020). Subsequent deep learning-based works either alleviate the burden of hyperparameter tuning via a neural net (Liu et al. 2022), or leave priori factors completely behind and rely on neural nets to automatically adapt to different physical models (Valanarasu, Yasarla, and Patel 2022). Despite these efforts, they always need paired images that are hard or even impossible to acquire due to the dynamic scenes in reality, to cater to network optimization. By contrast, we only utilize the single image with poor visibility and manage to design a dynamic module as a substitute for the manual tuning process.

Method

Given source images with pixel-level annotations from a normal-condition source domain $\mathcal{S}$ (e.g., good weather, favorable illumination), and unlabeled target images from an adverse-condition target domain $\mathcal{T}$ (e.g., nighttime, fog, rain, snow, etc.), the goal is to predict high-quality segmentation maps for the target domain. Note that the target domain described above could turn out to be a combination of several domains, but we blur their boundaries and regard them as featuring diversity in a single target domain.

The full pipeline of our method is illustrated in Fig. 2. Overall, it contains two major parts: (i) a basic framework that is composed of a teacher model for pseudo-labeling and a student model for online learning; (ii) two dedicated modules which encourage reducing domain differences in imaging conditions and output configurations. In the following, we provide a detailed description of *visibility boost module* as well as *logit-constraint learning*. After that, the overall optimization and algorithm are introduced.

Visibility Boost Module (VBM)

Previous methods generally utilize the normal-adverse image correspondences, which may put stress on data collection and annotation. Here, we merely exploit the images under adverse conditions in the target domain. At first, an appropriate prior should come in place to break out of the dilemma, and the atmosphere scatter model (ASM) (Nayar and Narasimhan 1999; Narasimhan and Nayar 2003) is a powerful candidate to describe haze formation:

$$I(x) = J(x)t(x) + A(1 - t(x)) \,, \tag{1}$$

where $I(x)$ is the observed hazy image, $J(x)$ is the scene radiance, namely restored image, and A is the atmospheric light estimated globally. $t(x)$ represents the transmission map describing the portion of light that survives scattering and reaches the camera. It is represented as: $t(x) = e^{-\beta d(x)}$, where β is the scattering coefficient and $d(x)$ denotes pixel-wise scene depth. As depicted in (Li, Cheong, and Tan 2019), the vanilla ASM can already model the veiling effect usually observed in fog, heavy rain, or even snowy scenes.

We consider such effect as the major obstacle lying in the way to a clear vision, and intend to alleviate the problem in an adaptive way. Motivated by (He, Sun, and Tang 2009; Liu et al. 2022), we propose the *visibility boost module* (VBM) to ameliorate images in various adverse conditions. We estimate the atmospheric light A from the 1,000 brightest pixels in the image, and the transmission map can be given as:

$$t(x) = 1 - \min_{c} \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right) , \tag{2}$$

where $c \in \{r, g, b\}$ is the color channel and $\Omega(x)$ is the local patch surrounding position x. To make the restored image more natural in appearance, we further devise a non-parametric coefficient ω_s to control the dehaze extent. Hereafter, the $t(x)$ is reformulated as

$$t(x) = 1 - \omega_s \min_{c} \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right) , \tag{3}$$

where ω_s is an adaptive coefficient for transmission modulation. Through observation, it can be summarized that the

images with veiling effect tend to be gray and dull, which can be approximately described as low saturation. On the contrary, the images with clear vision can be more vivid and colorful, resulting in greater saturation. By introducing this coefficient, the restored appearance can be constrained to some extent. To be precise, ω_s is dynamically calculated from the mean value of saturation $mean_s$ within an image:

$$\omega_s = e^{-mean_s \times \gamma}\,, \tag{4}$$

where γ is a scaling factor that is experimentally fixed to 4.0. And $mean_s$ is calculated by:

$$mean_s = \frac{1}{HW}\sum_{h,w}\frac{\max\limits_c I^c_{h,w} - \min\limits_c I^c_{h,w}}{\max\limits_c I^c_{h,w}}\,, \tag{5}$$

where H, W are height and width of the image I, and $I^c_{h,w}$ is the c^{th} color channel of the pixel indexed (h, w) in I.

However, one exception is the nighttime condition. Luckily, as claimed by Zhang et al. (2012), the reverted low-light images can be viewed as hazy images. Thus, we additionally add a pair of *Inverse Switch* to extend VBM to low-light condition, by which only night images are inverted and other types of images remain unchanged. We formulate this as:

$$1 - I(x) = (1 - J(x))t(x) + A(1 - t(x))\,. \tag{6}$$

With conditional inversion, one can handle all cases without modifying the core procedure of visibility enhancement. Note that Eq. (3) may seem close to the ones proposed in (He, Sun, and Tang 2009) and (Liu et al. 2022), but is different as it is neither manually tuned for each image nor reliant on paired image to learn a parameter. Actually, the scale factor γ is globally assigned, and then the coefficient ω_s can dynamically adapt to different images.

The complete pipeline of VBM is illustrated in the left of Fig. 2. For an arbitrary adverse-condition image, we first decide the application of *Inverse Switch* according to its lighting condition, and the mean saturation value $mean_s$ is extracted form the original/inverted image. After that, the coefficient ω_s is yielded by Eq. (4), which is then used to perform transmission modulation in Eq. (3). The image is then enhanced by ASM. Another *Inverse Switch* is applied in parallel with the aforementioned one. Through the above process, visibility enhanced target images can be obtained.

Logit-Constraint Learning (LCL)

In the literature, it is almost common practice to combine the self-training strategy with cross-entropy (CE) loss for both source and target samples (Zou et al. 2019, 2018; Tranheden et al. 2021; Xie et al. 2022). For simplicity, we take a pixel as an example, whose CE loss is formulated as follows:

$$\mathcal{L}_{ce} = -\sum_{k=1}^{K} y_k \log(p_k),\ \text{where}\ \ p_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}\,, \tag{7}$$

where K represents the number of classes, y is a one-hot ground-truth (pseudo) label, p_i is the probability for the i^{th} class, and z_i is the i^{th} element of the logit output.

As we all known, the CE loss forces predictions to resemble the corresponding one-hot labels, which makes it effective for supervised training paradigms. However, when it comes to unlabeled target samples, this characteristic can be a mixed blessing: reliable pseudo labels can compensate the deficiency of ground-truth labels, but erroneous pseudo labels can be catastrophic as a strong supervision. Existing self-training methods are well aware of such risk, and have attempted to address the issue through loss reweighting (Olsson et al. 2021; Tranheden et al. 2021), confidence thresholding (He, Yang, and Qi 2021), or pseudo label refinement (Zhang et al. 2021a). Nevertheless, these methods either generate pseudo labels regardless of their confidence, or just ignore the pixels under the threshold of confidence policy, failing to make the most of precious predictions. Furthermore, in the task of normal-to-adverse adaptation, trustworthy predictions can be rather scarce, thus blindly ignoring the unconfident pixels will result in low data efficiency.

To address this issue, we seek to push the utilization of model prediction to a new height through the enhancement of loss term, for its close relation with predictions. As the unconfident samples are promising providers of extra information, we contend that inter-class relationship within the prediction of a pixel should be emphasized. When taking the derivability into consideration, ℓ_2-norm is an ideal candidate as it does link all elements in an equal manner. Therefore, we integrate the ℓ_2-norm into the original CE loss as an expansion, forming a new *logit-constraint learning* loss:

$$\mathcal{L}_{lc} = -\sum_{k=1}^{K} y_k \log(p^*_k),\ \text{where}\ \ p^*_i = \frac{e^{z_i/\|z\|}}{\sum_{k=1}^{K} e^{z_k/\|z\|}}\,, \tag{8}$$

where $\|\cdot\|$ means ℓ_2-norm. The name *logit-constraint learning* comes from the fact that every logit element is rescaled by dividing the norm term, whose optimization is thus constrained by the other elements constituting the logit. We will theoretically reveal the inter-class constraint through gradient analysis in the following part.

Gradient Analysis. It's worth mentioning that this new loss function is not confined to the confident portion of predictions, but can be applied to unconfident predictions with a unified form. To justify our claim, let's take a closer look at the gradient during back-propagation. For the vanilla CE loss, the gradient of loss to the j^{th} logit element is:

$$\frac{\partial \mathcal{L}_{ce}}{\partial z_j} = p_j - y_j\,. \tag{9}$$

Just as discussed above, this gradient merely narrows the gap between the prediction and the corresponding label, making it inevitable to ruin the prediction if a wrong label is given. On the contrary, the gradient of our proposed *logit-constraint learning* loss to the j^{th} logit element is:

$$\frac{\partial \mathcal{L}_{lc}}{\partial z_j} = \frac{1}{\|z\|}\left((p^*_j - y_j) - \sum_{k=1}^{K}\frac{z_j z_k}{\|z\|^2}(p^*_k - y_k)\right)\,. \tag{10}$$

In this formula, the gradient is made up of two parts. The former part is essentially identical to the gradient of vanilla CE loss, while the latter part undoubtedly reflect the connection built across different classes. More derivations can be found in the Appendix.

Let us analysis the gradient from both confident and unconfident conditions. Assuming the prediction is confident, then the coefficient of the second term, namely $\frac{z_j}{\|z\|} * \frac{z_k}{\|z\|}$,

should be relative small except for $k = j$, and the gradient can be approximate to

$$\frac{1}{\|z\|}\left((p_j^* - y_j) - (\frac{z_j}{\|z\|})^2(p_k^* - y_k)\right). \tag{11}$$

If there is space left for optimization, i.e., $z_j < \|z\|$, this term is still capable of providing gradient; otherwise, the gradient degrades to zero and the optimization stops. When confronted with unconfident predictions, the coefficient of the second term would be relatively larger for all classes whose prediction p_k is close to that of p_j, given that z is directly related to p, thus reducing the gradient and slowing down the optimization towards the assigned pseudo label.

In a nutshell, the above gradient analysis not only reflects the ability of *logit-constraint learning* to follow the guidance of confident pseudo label, but highlights its potential to explore the knowledge hidden in unconfident predictions without fear of overconfidence.

Overall Optimization

During the training stage, images from both source and target domains are first randomly sampled, i.e., $I_s, I_t \in R^{H\times W\times 3}$, respectively. The target image is then passed into *visibility boost module* to obtain boosted target image I_b with better visibility. Subsequently, the pseudo label $\hat{Y}_b$ of I_b is predicted from the teacher model Φ'. Next, both original target image and boosted one are separately mixed up with the source image using Classmix (Olsson et al. 2021) for online learning. The blended images are noted as $I_{t\oplus s}$ and $I_{b\oplus s}$, which are then passed through the student model Φ to get logit outputs $Z_{t\oplus s}$ and $Z_{b\oplus s}$, respectively. And they share the same mixed label $\hat{Y}_{b\oplus s}$ that is mixed up between source ground-truth label Y_s and target pseudo label $\hat{Y}_b$. Before participating in the final loss calculation, logits are processed through channel-wise norm and softmax function to get the final prediction maps $P_s^*, P_{t\oplus s}^*, P_{b\oplus s}^*$. Eventually, for any image, the *logit-constraint learning* loss is given by:

$$\mathcal{L}_{lc}(Y, P^*) = -\frac{1}{HW}\sum_{h,w}\sum_{k=1}^{K} Y_{h,w,k}\log P_{h,w,k}^*, \tag{12}$$

where H, W are height and width of an input image, $Y_{h,w,k}$ is the k^{th} element in a one-hot label of pixel indexed (h, w), and $P_{h,w,k}^*$ is the corresponding pixel-level prediction.

For the source data, we have $\mathcal{L}_{lc}^s = \mathcal{L}_{lc}(Y_s, P_s^*)$. For the target data, a quality estimation is produced for the pseudo labels following Tranheden et al. (2021). Here an adaptive weight λ_δ is calculated from the proportion of confident pixel-level predictions ($\max_k P_{h,w,k}^* > \delta$) and weighted on losses involving target images. Subsequently, the losses for blended images $I_{t\oplus s}, I_{b\oplus s}$ can be respectively computed as $\mathcal{L}_{lc}^{t\oplus s} = \lambda_\delta\mathcal{L}_{lc}(\hat{Y}_{b\oplus s}, P_{t\oplus s}^*)$ and $\mathcal{L}_{lc}^{b\oplus s} = \lambda_\delta\mathcal{L}_{lc}(\hat{Y}_{b\oplus s}, P_{b\oplus s}^*)$. Overall, the training objective can be formulated as:

$$\min_{\Phi}\ \mathcal{L}_{lc}^s + \mathcal{L}_{lc}^{t\oplus s} + \mathcal{L}_{lc}^{b\oplus s}. \tag{13}$$

By default, loss weighting coefficients are set to 1.0. For a detailed schedule, please refer to Alg. 1 in the Appendix.

Experiments

In this section, we assess the effectiveness of VBLC under several adverse conditions. For each task, we first give a brief introduction to the datasets and architectures involved. Following up are experimental results and insight analyses. Limited by space, more details are left for the Appendix[1].

Normal-to-Adverse Domain Adaptation

Datasets and Architectures. We first take out the experiments on two challenging semantic segmentation tasks, i.e., Cityscapes (Cordts et al. 2016) → ACDC (Sakaridis, Dai, and Van Gool 2021) and Cityscapes (Cordts et al. 2016) → FoggyCityscapes (Sakaridis, Dai, and Van Gool 2018) + RainCityscapes (Hu et al. 2019). We further testify the generality of VBLC by performing object detection on the latter. Among these tasks, Cityscapes (source domain) serves as a collection of clear images, while images from other datasets (target domain) all feature degraded visibility to some extent. For this part, we experiment on both CNN-based DeepLab-v2 (Chen et al. 2017) and Transformer-based SegFormer (Xie et al. 2021) to give a whole picture of the segmentation quality of our method. As to object detection task, following Wang et al. (2021b), Deformable DETR (Zhu et al. 2021) is adopted as the basic architecture.

For semantic segmentation task, we utilize per class Intersection-over-Union (IoU) (Everingham et al. 2015) and mean IoU (mIoU) over all classes as an evaluation. For object detection task, we report the standard average precision (AP) result under different IoU thresholds and object scales.

Experimental Results. The comparison of our VBLC to relative methods on Cityscapes → ACDC segmentation task is listed in Table 1. Generally, the SegFormer-based methods substantially outperform the DeepLab-based ones. The previous state-of-the-art method built on DeepLab-v2 is FDA with a mIoU of 45.7%, but our VBLC takes a step further and achieves 47.8% mIoU, gaining a large boost of +2.1% and could even rank among the SegFormer-based counterparts. When integrated with the stronger backbone, VBLC still yields a leading result of 64.2% mIoU, outperforming DAFormer by a huge margin of +8.9%. We also provide qualitative semantic segmentation results in Fig. 3. We can observe clear improvement against both Source-only and state-of-the-art adaptation (DAFormer) models, especially in the prediction of sky, light, and sign.

Table 2 shows the segmentation results on Cityscapes → FoggyCityscapes + RainCityscapes. Due to slight domain shift, outcomes of the Source-only models are already high, however, our VBLC is still capable of providing consistent performance gain, surpassing DeepLab-v2 and SegFormer by +13.0% mIoU and +8.8% mIoU, respectively. This complementary experiment explores the robustness of our VBLC on tasks containing synthetic datasets, and proves the scalability of the proposed modules.

To showcase the flexibility of VBLC, we further combine it with state-of-the-art UDA detection methods (Wang

[1]Appendix can be found at https://arxiv.org/abs/2211.12256

Method	road	side.	buil.	wall	fence	pole	light	sign	veg.	terr.	sky	pers.	rider	car	truck	bus	train	mbike	bike	mIoU
DeepLab-v2	71.9	26.2	51.1	18.8	22.5	19.7	33.0	27.7	67.9	28.6	44.2	43.1	22.1	71.2	29.8	33.3	48.4	26.2	35.8	38.0
AdaptSegNet	69.4	34.0	52.8	13.5	18.0	4.3	14.9	9.7	64.0	23.1	38.2	38.6	20.1	59.3	35.6	30.6	53.9	19.8	33.9	33.4
ADVENT	72.9	14.3	40.5	16.6	21.2	9.3	17.4	21.2	63.8	23.8	18.3	32.6	19.5	69.5	36.2	34.5	46.2	26.9	36.1	32.7
BDL	56.0	32.5	68.1	20.1	17.4	15.8	30.2	28.7	59.9	25.3	37.7	28.7	25.5	70.2	39.6	40.5	52.7	29.2	38.4	37.7
CLAN	**79.1**	29.5	45.9	18.1	21.3	22.1	35.3	40.7	67.4	29.4	32.8	42.7	18.5	73.6	42.0	31.6	55.7	25.4	30.7	39.0
CRST	51.7	24.4	67.8	13.3	9.7	30.2	38.2	34.1	58.0	25.2	**76.8**	39.9	17.1	65.4	3.7	6.6	39.6	11.8	8.6	32.8
FDA	73.2	34.7	59.0	24.8	**29.5**	28.6	43.3	44.9	**70.1**	28.2	54.7	47.0	28.5	74.6	**44.8**	**52.3**	**63.3**	28.3	39.5	45.7
DACS	58.5	34.7	76.4	20.9	22.6	31.7	32.7	46.8	58.7	39.0	36.3	43.7	20.5	72.3	39.6	34.8	51.1	24.6	38.2	41.2
VBLC	49.6	**39.3**	**79.4**	**35.8**	**29.5**	**42.6**	**57.2**	**57.5**	69.1	**42.7**	39.8	**54.5**	**29.3**	**77.8**	43.0	36.2	32.7	**38.7**	**53.4**	**47.8**
SegFormer	66.9	25.8	71.3	20.9	22.2	41.1	47.2	46.6	**74.2**	44.9	75.6	50.4	23.5	73.1	30.3	36.8	55.8	29.4	37.1	45.9
DAFormer	56.9	45.4	84.7	**44.7**	35.1	48.6	44.8	57.4	69.5	52.9	45.8	57.1	28.2	82.8	57.2	63.9	84.0	40.2	50.5	55.3
VBLC	**89.2**	**59.8**	**85.9**	44.0	**37.2**	**53.5**	**64.5**	**63.2**	72.4	**56.3**	**84.1**	**65.5**	**37.7**	**85.1**	**60.1**	**71.8**	**85.2**	**47.7**	**56.3**	**64.2**

Table 1: Comparison with the state-of-the-arts on Cityscapes → ACDC semantic segmentation task. IoU score of each class and the mIoU score are reported on ACDC testing set. The bests results are highlighted in bold.

Method	road	side.	buil.	wall	fence	pole	light	sign	veg.	terr.	sky	pers.	rider	car	truck	bus	train	mbike	bike	mIoU
DeepLab-v2	96.7	72.4	74.1	28.6	41.4	42.2	49.8	67.6	72.6	62.5	80.6	70.4	54.4	88.4	56.1	72.4	33.7	42.7	70.1	61.9
FDA	87.0	56.9	82.1	4.3	11.6	36.3	41.8	60.4	80.6	51.6	70.6	66.7	50.3	86.0	46.4	63.7	26.2	41.4	66.3	54.2
DACS	97.9	82.3	**88.7**	40.8	42.4	41.0	53.5	67.3	**89.2**	58.2	**90.8**	70.8	54.4	91.3	62.9	82.5	**56.4**	47.0	72.4	67.9
VBLC	**98.6**	**86.9**	87.2	**62.1**	**55.3**	**54.2**	**65.1**	**77.8**	86.9	**66.8**	90.1	**77.5**	**63.2**	**93.7**	**77.3**	**86.6**	55.0	**59.4**	**79.5**	**74.9**
SegFormer	97.8	81.6	86.9	54.3	48.3	49.2	57.3	71.6	86.9	65.5	83.4	71.9	57.1	91.8	67.9	80.1	73.1	49.9	74.6	71.0
DAFormer	98.5	87.0	90.8	55.1	53.7	56.3	62.8	73.6	91.5	70.7	90.0	75.6	56.8	92.7	65.9	88.3	79.9	56.9	77.6	74.9
VBLC	**98.7**	**88.4**	**91.9**	**66.3**	**65.2**	**62.7**	**69.1**	**79.6**	**92.2**	**72.4**	**92.3**	**80.0**	**66.0**	**94.6**	**79.9**	**90.9**	**81.8**	**64.0**	**80.6**	**79.8**

Table 2: Comparison with the state-of-the-arts on Cityscapes → FoggyCityscapes + RainCityscapes semantic segmentation task. IoU score of each class and the mIoU score are reported. The bests results are highlighted in bold.

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Deformable DETR	13.4	22.7	13.4	3.4	17.0	26.8
SFA	14.3	24.6	14.6	4.2	17.4	28.2
SFA + **VBLC**	**15.6**	**26.0**	**16.4**	**5.2**	**17.8**	**30.9**

Table 3: Comparison with the state-of-the-arts on Cityscapes → FoggyCityscapes + RainCityscapes object detection task with Deformable DETR (Zhu et al. 2021).

et al. 2021b) on Cityscapes → FoggyCityscapes + RainCityscapes object detection task. To be specific, images from the target domain is first boosted with the designed *visibility boost module*, which performs coarse alignment between normal and adverse conditions. Then, the *logit-constraint learning* is integrated with class prediction. And the results are reported in Table 3. We can observe that VBLC boosts the performance of SFA by a substantial 1.3 AP, validating that our method can indeed generalize well under the variation of weather conditions, both for segmentation and detection.

Multi-Target Domain Adaptation

Datasets and Architectures. Now we turn to multi-target domain adaptation (MTDA), adapting from Cityscapes to IDD (Varma et al. 2019) and Mapillary (Neuhold et al. 2017). All datasets are captured in reality without specific inclusion of images under adverse conditions. We investigate this as a special case and compare our VBLC with other well-established MTDA methods. All methods mentioned in this part are built on DeepLab-v2 for a fair comparison.

# Class	Method	IDD(mIoU)	Mapillary(mIoU)	Avg. mIoU
7	MTKT	68.3	69.3	68.8
	ADAS	70.4	**75.1**	72.7
	VBLC	**73.9**	71.7	**72.8**
19	CCL	**53.6**	51.4	52.5
	ADAS	48.3	53.6	50.5
	VBLC	52.9	**57.8**	**55.3**

Table 4: Comparison with the state-of-the-arts on Cityscapes → IDD + Mapillary semantic segmentation task. mIoU score of each domain and their average are reported.

Experimental Results. In accordance with previous attempts, we report the segmentation results on Cityscapes → IDD + Mapillary with both 19 classes and 7 super classes settings in Table 4. On either of both, VBLC takes the lead regarding the average mIoU over two target domains, attaining 55.3%/72.8% mIoU for 19/7 classes, respectively. Note that VBLC neither intends to deal with MTDA directly nor to enforce the explicit dispersion of multiple target domains, yet is still comparable to specially designed counterparts.

Ablation Studies

The Effect of Each Component on Cityscapes → ACDC. We report the contribution of each component in Table 5. The first line presents the Source-only (SegFormer) model trained only on Cityscapes, which serves as the baseline with 45.9% mIoU. When combined with conventional self-training on source-target blended image ($\mathcal{L}_{ce}^{t \oplus s}$), the perfor-

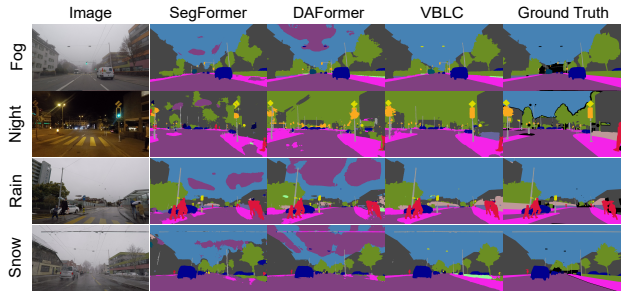


Figure 3: Visualization of segmentation results on ACDC validation set. From left to right: target images under distinct adverse conditions, results of SegFormer, results of DAFormer, results of our VBLC, and ground-truth labels.

Method	VBM	$\mathcal{L}_{ce}^{s}$	$\mathcal{L}_{ce}^{t\oplus s}$	$\mathcal{L}_{ce}^{b\oplus s}$	$\mathcal{L}_{lc}^{s}$	$\mathcal{L}_{lc}^{t\oplus s}$	$\mathcal{L}_{lc}^{b\oplus s}$	mIoU
Src-only		✓						45.9
Ours		✓	✓					52.6 (6.7↑)
	✓	✓	✓	✓				57.0 (4.4↑)
	✓	✓				✓	✓	63.2 (6.2↑)
	✓				✓	✓	✓	**64.2 (1.0↑)**

Table 5: Ablation study on Cityscapes → ACDC semantic segmentation task.

mance is significantly boosted with a gain of +6.7% mIoU, validating the huge potential of self-training scheme.

Next, the *visibility boost module* (VBM) is integrated to ease the generation of pseudo labels. Concretely, we utilize the prediction of boosted target image from teacher model to guide both original and boosted target images. A moderate improve of performance can be witnessed in this process, and we attribute this to the fact that prediction of boosted target images are appropriately constrained (VBM + $\mathcal{L}_{ce}^{b\oplus s}$).

The penultimate line highlights the power of our *logit-constraint learning*, which further brings a substantial increase of +6.2% mIoU, leading to a competitive performance of 63.2% mIoU ($\mathcal{L}_{lc}^{t\oplus s}$ + $\mathcal{L}_{lc}^{b\oplus s}$). Finally, we also apply the *logit-constraint learning* to the source domain, and obtain a bonus of +1.0% mIoU ($\mathcal{L}_{lc}^{s}$), yielding the ultimate score of 64.2% mIoU. In summary, we can learn that VBLC mainly enhances the performance from two critical aspects, i.e., visibility boost and logit-constraint learning.

Analysis on $\mathcal{L}_{lc}$. In Fig. 4, we visualize confidence distributions from models trained by vanilla CE loss ($\mathcal{L}_{ce}$) or *logit-constraint learning* loss ($\mathcal{L}_{lc}$), on ACDC validation set. We opt to adopt max softmax as the confidence, without test-time logit constraint for a fair comparison. The left chart shows confidence distribution on the whole validation set. We can observe that, the model trained with $\mathcal{L}_{ce}$ tends to be confident with a majority of predictions, while the one trained with $\mathcal{L}_{lc}$ remains skeptical to a handful of them. Indeed, the $\mathcal{L}_{lc}$ allows the coexistence of especially confident predictions and rather unconfident ones, which is in line with our analysis. The right chart, on the other hand, illustrates

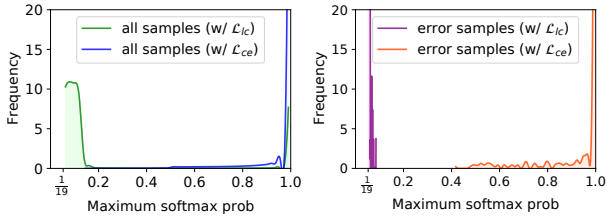
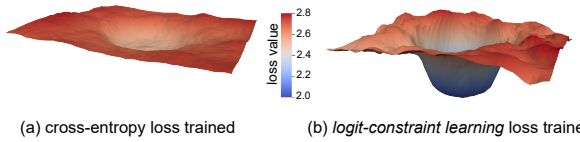


Figure 4: Confidence distribution over all (left chart) or erroneous (right chart) predictions on ACDC validation set.



(a) cross-entropy loss trained (b) *logit-constraint learning* loss trained

Figure 5: The loss surfaces of models trained with $\mathcal{L}_{ce}$ and $\mathcal{L}_{lc}$. Our $\mathcal{L}_{lc}$ is more advanced for parameter optimization.

the confidence distribution on erroneous predictions only. This chart further reflects whether the predictions are overconfident. Obviously, predictions from the model trained with $\mathcal{L}_{ce}$ is much unreliable, as higher confidence could indicate greater error rate. By contrast, the model trained with $\mathcal{L}_{lc}$ seldom predicts wrong with a high confidence, showing its strong capacity to mitigate overconfidence.

Loss Landscape Visualization. To delve into the optimization potential of *logit-constraint learning*, we plot the loss landscape (Li et al. 2018) of models train with $\mathcal{L}_{ce}$ or $\mathcal{L}_{lc}$ in Fig. 5. The figures are drawn from loss variation with model parameter perturbation, and the statistics is collected on the whole target train set with ground-truth labels. It is clear that the model trained by $\mathcal{L}_{lc}$ is able to achieve much lower loss in regions where that trained by $\mathcal{L}_{ce}$ remains confused. Furthermore, our $\mathcal{L}_{lc}$ expands the model's potential to pursue superior prediction quality on the target domain, as is illustrated by the blue region featuring minor error.

Influence of Hyperparameters. We traverse hyperparameters (pseudo threshold δ, scaling factor γ, and momentum-update ratio α) around their optimal values $[\delta^*, \gamma^*, \alpha^*]$. Results are provided in the Appendix, in which we observe that VBLC is much less sensitive to its hyperparameters.

Conclusion

In this paper, we propose VBLC, an new framework especially designed for better normal-to-adverse adaptation, to explore the possibility of getting rid of reference images. This method contributes in both input and output space to enable an improved prediction quality even in poor visibility scenarios. This simple yet effective approach provides the best of both worlds: *visibility boost module* dynamically ameliorates incoming images via certain priors, while *logit-constraint learning* relieves the pain of overconfidence in the self-training paradigm. Our method can be trained end-to-end in one stage, leading to considerable performance gains on many challenging adverse conditions.

Acknowledgments

This paper was supported by National Key R&D Program of China (No. 2021YFB3301503), and also supported by the National Natural Science Foundation of China under Grant No. U21A20519.

References

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Mach. Learn.*, 79(1): 151–175.

Bruggemann, D.; Sakaridis, C.; Truong, P.; and Van Gool, L. 2022. Refign: Align and Refine for Adaptation of Semantic Segmentation to Adverse Conditions. *arXiv preprint arXiv:2207.06825*.

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 Trans. Pattern Anal. Mach. Intell.*, 40(4): 834–848.

Chen, W.-T.; Fang, H.-Y.; Ding, J.-J.; Tsai, C.-C.; and Kuo, S.-Y. 2020. JSTASR: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *ECCV*, 754–770.

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 *CVPR*, 3213–3223.

Dai, D.; Sakaridis, C.; Hecker, S.; and Gool, L. V. 2020. Curriculum Model Adaptation with Synthetic and Real Data for Semantic Foggy Scene Understanding. *Int. J. Comput. Vis.*, 128(5): 1182–1204.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *CVPR*, 248–255.

Everingham, M.; Eslami, S. M. A.; Gool, L. V.; Williams, C. K. I.; Winn, J. M.; and Zisserman, A. 2015. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.*, 111(1): 98–136.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. S. 2016. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.*, 17(1): 2096–2030.

Gong, R.; Chen, Y.; Paudel, D. P.; Li, Y.; Chhatkuli, A.; Li, W.; Dai, D.; and Gool, L. V. 2021. Cluster, Split, Fuse, and Update: Meta-Learning for Open Compound Domain Adaptive Semantic Segmentation. In *CVPR*, 8344–8354.

He, K.; Sun, J.; and Tang, X. 2009. Single image haze removal using dark channel prior. In *CVPR*, 1956–1963.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*, 770–778.

He, R.; Yang, J.; and Qi, X. 2021. Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In *ICCV*, 6930–6940.

He, Y.; Rahimian, S.; Schiele, B.; and Fritz, M. 2020. Segmentations-leak: Membership inference attacks and defenses in semantic image segmentation. In *ECCV*, 519–535. Springer.

Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A.; and Darrell, T. 2018. Cycada: Cycle-consistent adversarial domain adaptation. In *ICML*, 1989–1998.

Hoyer, L.; Dai, D.; and Gool, L. V. 2022. DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation. In *CVPR*, 9924–9935.

Hu, X.; Fu, C.-W.; Zhu, L.; and Heng, P.-A. 2019. Depth-attentional features for single-image rain removal. In *CVPR*, 8022–8031.

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 *CVPR*, 8187–8196.

Kim, M.; and Byun, H. 2020. Learning texture invariant representation for domain adaptation of semantic segmentation. In *CVPR*, 12975–12984.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

Land, E. H. 1977. The retinex theory of color vision. *Scientific american*, 237(6): 108–129.

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 *CVPR*, 19196–19206.

Li, H.; Xu, Z.; Taylor, G.; Studer, C.; and Goldstein, T. 2018. Visualizing the loss landscape of neural nets. In *NeurIPS*, 6391–6401.

Li, R.; Cheong, L.-F.; and Tan, R. T. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *CVPR*, 1633–1642.

Li, S.; Xie, B.; Lin, Q.; Liu, C. H.; Huang, G.; and Wang, G. 2022. Generalized Domain Conditioned Adaptation Network. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8): 4093–4109.

Li, Y.; Yuan, L.; and Vasconcelos, N. 2019. Bidirectional learning for domain adaptation of semantic segmentation. In *CVPR*, 6936–6945.

Liu, W.; Ren, G.; Yu, R.; Guo, S.; Zhu, J.; and Zhang, L. 2022. Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions. In *AAAI*, 1792–1800.

Liu, Z.; Miao, Z.; Pan, X.; Zhan, X.; Lin, D.; Yu, S. X.; and Gong, B. 2020. Open Compound Domain Adaptation. In *CVPR*, 12403–12412.

Long, M.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018. Conditional Adversarial Domain Adaptation. In *NeurIPS*, 1647–1657.

Loshchilov, I.; and Hutter, F. 2019. Decoupled weight decay regularization. In *ICLR*. OpenReview.net.

Luo, Y.; Zheng, L.; Guan, T.; Yu, J.; and Yang, Y. 2019. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *CVPR*, 2507–2516.

Ma, X.; Wang, Z.; Zhan, Y.; Zheng, Y.; Wang, Z.; Dai, D.; and Lin, C.-W. 2022. Both style and fog matter: Cumulative

domain adaptation for semantic foggy scene understanding. In *CVPR*, 18922–18931.

McCartney, E. J. 1976. Optics of the atmosphere: scattering by molecules and particles. *New York*.

Narasimhan, S. G.; and Nayar, S. K. 2003. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(6): 713–724.

Nayar, S. K.; and Narasimhan, S. G. 1999. Vision in bad weather. In *ICCV*, volume 2, 820–827.

Neuhold, G.; Ollmann, T.; Rota Bulo, S.; and Kontschieder, P. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *ICCV*, 4990–4999.

Olsson, V.; Tranheden, W.; Pinto, J.; and Svensson, L. 2021. Classmix: Segmentation-based data augmentation for semi-supervised learning. In *WACV*, 1369–1378.

Park, K.; Woo, S.; Shin, I.; and Kweon, I. S. 2020. Discover, Hallucinate, and Adapt: Open Compound Domain Adaptation for Semantic Segmentation. In *NeurIPS*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*, 8024–8035.

Sakaridis, C.; Dai, D.; and Gool, L. V. 2022. Map-Guided Curriculum Domain Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(6): 3139–3153.

Sakaridis, C.; Dai, D.; Hecker, S.; and Van Gool, L. 2018. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *ECCV*, 687–704.

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

Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *ICCV*, 10765–10775.

Saporta, A.; Vu, T.-H.; Cord, M.; and Pérez, P. 2021. Multi-target adversarial frameworks for domain adaptation in semantic segmentation. In *ICCV*, 9072–9081.

Tranheden, W.; Olsson, V.; Pinto, J.; and Svensson, L. 2021. DACS: Domain adaptation via cross-domain mixed sampling. In *WACV*, 1379–1389.

Tsai, Y.-H.; Hung, W.-C.; Schulter, S.; Sohn, K.; Yang, M.-H.; and Chandraker, M. 2018. Learning to adapt structured output space for semantic segmentation. In *CVPR*, 7472–7481.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial discriminative domain adaptation. In *CVPR*, 7167–7176.

Valanarasu, J. M. J.; Yasarla, R.; and Patel, V. M. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, 2353–2363.

Varma, G.; Subramanian, A.; Namboodiri, A.; Chandraker, M.; and Jawahar, C. 2019. IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments. In *WACV*, 1743–1751.

Vu, T.-H.; Jain, H.; Bucher, M.; Cord, M.; and Pérez, P. 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *CVPR*, 2517–2526.

Wang, M.; and Deng, W. 2018. Deep visual domain adaptation: A survey. *Neurocomputing*, 312: 135–153.

Wang, Q.; Dai, D.; Hoyer, L.; Van Gool, L.; and Fink, O. 2021a. Domain adaptive semantic segmentation with self-supervised depth estimation. In *ICCV*, 8515–8525.

Wang, W.; Cao, Y.; Zhang, J.; He, F.; Zha, Z.; Wen, Y.; and Tao, D. 2021b. Exploring Sequence Feature Alignment for Domain Adaptive Detection Transformers. In *ACM MM*, 1730–1738.

Wei, H.; Xie, R.; Cheng, H.; Feng, L.; An, B.; and Li, Y. 2022. Mitigating Neural Network Overconfidence with Logit Normalization. *arXiv preprint arXiv:2205.09310*.

Wu, X.; Wu, Z.; Ju, L.; and Wang, S. 2021. A One-Stage Domain Adaptation Network with Image Alignment for Unsupervised Nighttime Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1–1.

Wulfmeier, M.; Bewley, A.; and Posner, I. 2018. Incremental adversarial domain adaptation for continually changing environments. In *ICRA*, 4489–4495.

Xie, B.; Li, S.; Li, M.; Liu, C. H.; Huang, G.; and Wang, G. 2022. SePiCo: Semantic-Guided Pixel Contrast for Domain Adaptive Semantic Segmentation. *arXiv preprint arXiv:2204.08808*.

Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J. M.; and Luo, P. 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *NeurIPS*.

Yang, Y.; and Soatto, S. 2020. FDA: Fourier Domain Adaptation for Semantic Segmentation. In *CVPR*, 4085–4095.

Zhang, P.; Zhang, B.; Zhang, T.; Chen, D.; Wang, Y.; and Wen, F. 2021a. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *CVPR*, 12414–12424.

Zhang, X.; Shen, P.; Luo, L.; Zhang, L.; and Song, J. 2012. Enhancement and noise reduction of very low light level images. In *ICPR*, 2034–2037.

Zhang, Y.; Carballo, A.; Yang, H.; and Takeda, K. 2021b. Autonomous Driving in Adverse Weather Conditions: A Survey. *arXiv preprint arXiv:2112.08936*.

Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; and Dai, J. 2021. Deformable DETR: Deformable Transformers for End-to-End Object Detection. In *ICLR*.

Zou, Y.; Yu, Z.; Kumar, B.; and Wang, J. 2018. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *ECCV*, 289–305.

Zou, Y.; Yu, Z.; Liu, X.; Kumar, B.; and Wang, J. 2019. Confidence regularized self-training. In *ICCV*, 5982–5991.