

DUDA: Online-Offline Dual Domain Adaption for Semantic Segmentation

Antao Pan
pat@zju.edu.cn

Yawei Luo✉
yaweiluo@zju.edu.cn

Yi Yang
yiyang@zju.cs.edu.cn

Jun Xiao
junx@cs.zju.edu.cn

Zhejiang University
Hangzhou, CN

Abstract

Self-training-based methods have achieved superior performance on unsupervised domain adaptive semantic segmentation task. However, these methods severely suffer from noisy pseudo label assignment. In this paper, we propose a simple yet effective dual pseudo label updating method that employs both online and offline mechanisms to dynamically update the two groups of pseudo labels. The online updating module employs a mean model to generate pseudo labels on-the-fly while the offline updating module capitalizes on the temporal consistency information to correct noisy labels. Furthermore, we present an online-offline dual regularization to further improve the noise-tolerant ability of the model. Combining the online-offline dual updating and online-offline dual regularization, we propose a novel mean-teacher based framework dubbed *Online-Offline Dual Domain Adaption* (DUDA). Experiments show the proposed DUDA brings large performance gain and achieves state-of-the-art performance on two challenging benchmarks: GTA-to-Cityscapes and SYNTHIA-to-Cityscapes (58.4% mIoU and 59.7% mIoU respectively). Our code will be available at [DUDA-Semantic-Segmentation-UDA](https://github.com/taoan/DUDA-Semantic-Segmentation-UDA).

1 Introduction

Deep convolution neural networks (DCNNs) have made great progress in semantic segmentation task over these years [1, 8, 14]. Training a DCNN model usually requires a large amount of data while labelling is expensive and infeasible at large scales for semantic segmentation. Therefore, computer graphics technology is employed to simulate images, which can automatically generate accurate pixel-level semantic labels [14, 15]. However, due to the significant appearance gap between synthetic and natural styles, the model trained by simulated images usually suffers a huge performance drop when deployed to the real scene. To address this problem, the domain adaptation technique is proposed with the goal of transferring the knowledge from simulated images (source domain) to real scenes (target domain)

thus reducing the performance drop caused by the domain gap [16, 17, 18, 19]. In this campaign, the unsupervised domain adaptation (UDA) is one of the most popular and realistic settings, where the labels of target domain are totally unavailable.

This paper focuses on the UDA semantic segmentation problem. Currently, adversarial training [16, 17, 19, 30, 31, 35, 44] and self-training [9, 13, 15, 20, 40, 41, 46, 47] are the two most commonly used methods for UDA. Adversarial training follows the idea of Generative Adversarial Networks (GAN) [5] and employs a discriminator to align the distribution of source and target domains in feature space, while self-training employs the pseudo labels on target domain to train the model directly. Most recent domain adaptation pipelines [15, 20, 40, 41] utilize adversarial training to acquire an initial model firstly and then apply self-training to further improve performance. However, most self-training methods employ the offline pseudo label updating strategies and suffer from an inevitable problem, *i.e.* the noisy labels assignment, which would lead the model to overfit to the incorrect pixel labels and hinder further performance improvement. To tackle this issue, recent works begin to employ online pseudo label strategy [40, 41] due to its self-correcting property, which uses a mean-teacher model to generate pseudo labels on-the-fly as supervision for student model. Particularly, the current state-of-the-art method MFA [40] combines both online and offline pseudo labels, where online pseudo labels are provided by another model in co-learning [6] framework.

We find that self-correcting property in aforementioned methods brings significant performance gain, but the offline labels are always ignored which are simply kept fixed. Based on this thought, we argue that the offline pseudo labels should also be dynamically updated to achieve self-correcting. To this end, this paper presents a novel Online-Offline Dual Domain Adaption (DUDA) method to reduce the effect of noisy labels in the self-training stage. DUDA includes an online-offline dual update module and an online-offline dual regularization module to achieve better denoising performance. Inspired by SWLF [27] which provides a new standpoint for denoising and proves that the temporal consistency information of model output during training can be used to filter out the noisy labels, we assume that the predicted result of noisy pseudo labels may be jittered and temporal inconsistent. Basing on this assumption, we propose a novel algorithm named *Temporal-aware Offline Pseudo Label Update* (TOPLU) that can update the offline pseudo labels via utilizing the model improvement and the temporal consistency information during training. TOPLU together with the online updating strategy form our online-offline dual updating module.

Another work that is inspiring to ours is HCL [10]. Both HCL and our DUDA employ the historical consistency information but some differences exists. In our setting, the DUDA uses the historical information on label index map directly to update the offline pseudo label under self-training framework. While HCL utilizes the information in feature level and employs contrastive learning method. In HCL the historical consistency information is used to prevent the model forgetting source information and reweight the contrastive loss.

We also find that regularization is effective for UDA task but ignored in previous online pseudo label assignment. In DUDA, we introduce a region regularization paradigm and extend it to dual regularization mode to further boost the performance. The “region” indicates the regularization is applied on area-level unlabelled or labelled pixels, and the regularization term includes an entropy minimization regularization [30] and a KLD-regularization [47]. Here we apply it on both online and offline pseudo labels and dubbed it “dual” region regularization. The main contributions of our work are summarized as follows:

- We proposed an online-offline dual self-training module which includes both online

and offline pseudo label updating to reduce the amount of noisy pseudo labels. To our best knowledge, we are among the pioneers to update **offline** pseudo labels in UDA semantic segmentation task.

- Region regularization is introduced in this work and we extend it into the dual regularization mode that applies the region regularization on both online and offline pseudo labels. Comparing with the traditional offline-only regularization, we show the effectiveness of our dual strategy under both online and offline settings.
- DUDA is built within a mean-teacher framework equipped with above two contributions. Extensive experiments demonstrate the effectiveness of DUDA: on DeepLabv2 backbone, DUDA achieves 58.4% mIoU on GTA5-to-Cityscapes benchmark and 59.7% mIoU on SYNTHIA-to-Cityscapes benchmark without any external steps such as multi-stage self-training or model distillation.

2 Related work

In this section, we briefly review existing methods for self-training based UDA semantic segmentation, noisy label learning techniques and regularization method, which form the main motivations of our method.

Self-training for UDA semantic segmentation. Self-training methods using pseudo labels are widely applied in the field of UDA semantic segmentation [20, 40, 41, 42, 46]. In CBST [46] and IAST [20], class balance strategies based on softmax probability are proposed for selecting pseudo labels. In CAG [42], a novel pseudo label select method using category anchors is proposed which does not depend on complex class balancing strategy. The above methods all apply the offline pseudo label strategy in which pseudo labels remain unchanged during training. In ProDA [41], an online pseudo label strategy is presented and uses a mean-teacher model to generate online pseudo labels as supervision of student model. MFA [40] combines the offline and online pseudo labels and achieves better performance. Recently, CPSL [13] further improve the ProDA via considering the class-balance issue and employing optimal transport to achieve label assignment.

Noisy label learning. Robust loss function and co-learning are two common noisy label learning methods. For robust loss function method [9, 32, 33, 43], the goal is to design a loss function that is tolerant to noisy labels. In [9], Mean Absolute Error (MAE) is theoretically proved to be robust to noisy label. Generalized cross-entropy(GCE) [43] and symmetric cross-entropy(SCE) [33] combine reverse cross-entropy together with the cross entropy to achieve more noise-tolerance performance. For the co-learning based methods, they use multi-model information to suppress noisy labels [6, 28, 34, 37]. Co-teaching [8] maintains two networks and selects the small-loss samples to train another network. Co-teaching+ [37] uses the disagreement predictions to maintain the inconsistency of the two models, leading to a more robust model against the noise. JoCoR [34] presents the so-called co-regularization and proves the consistency information of the two different models is also important.

Regularization. This paper focuses on consistency regularization and entropy regularization. Consistency regularization has been used in UDA task [59, 40, 41] and it is first proposed by [10] and is widely accepted in mean-teacher framework [29]. FixMatch [27] and FlexMatch [68] employ the consistency regularization loss, which enforces the model output similar between the weak augmentation and strong augmentation input. In UDA task,

entropy regularization is commonly used. Advent [60] adapts the entropy minimum regularization for adversarial training and CRST [47] proposes KLD regularization to avoid the model overfitting to overconfident wrong label. IAST [20] combines the two regularizations into different region of pseudo labels.

Our work is mostly inspired by MFA [40] where we employ a mean-teacher framework and online pseudo label assignment strategy. Different from MFA, we additionally employ the temporal consistency information to update the offline pseudo label. Besides, the traditional region regularization [20] applied in the offline setting is also extended into online mode to further improve the performance.

3 Method

3.1 Preliminaries

In UDA semantic segmentation task, datasets are divided into the source and target domains. The two domains share the same K classes. For source dataset $D^s = \{(X_i^s, Y_i^s)\}$, where $X^s \in \mathbb{R}^{H \times W \times 3}$ are RGB images and $Y^s \in \mathbb{R}^{H \times W \times K}$ are the corresponding labels in one-hot format. Target dataset $D^t = \{(X_i^t, Y_i^t)\}$ is similar to source dataset in format while X_i^t is available only. The segmentation model can be written as $h = d \circ f$ which is composited by a feature extractor f and a classifier d . We denote $P = h(X|\theta)$ as a forward calculation of a model with parameter θ , and the input and output are X and P respectively. Cross-entropy is commonly used as a training loss.

$$\mathcal{L}_{ce}(Y, P) = - \sum_{i=1}^{H \times W} \sum_{k=1}^K y^{(i,k)} \cdot \log(p^{(i,k)}) \quad (1)$$

where Y is the ground truth in one-hot format and P is the softmax output of the model. The $y^{(i,k)}$ and $p^{(i,k)}$ are the pixel level representation. In addition, we denote I_{igr} as the ignore index for cross-entropy loss.

DUDA is formed by a standard mean-teacher framework, which consists of a gradient updated student model and a moving average updated teacher model. Different from regular exponential moving average (EMA) update [29] for the teacher model, we follow the spirit of MFA [40] and use SWA [12] to update the teacher model:

$$\theta_t^{mean} = \theta_{t-1}^{mean} + \frac{(\theta_t - \theta_{t-1}^{mean})}{n_{avg} + 1} \quad (2)$$

where θ_t and θ_t^{mean} represent the parameters of student and teacher in the t step respectively, and n_{avg} means the number of update times. Like most previous domain adaptation methods, the loss function of DUDA can be roughly described as:

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_t + \mathcal{L}_{reg} \quad (3)$$

\mathcal{L}_s is the source-domain supervision, which is implemented as a naive cross-entropy loss:

$$\mathcal{L}_s = \mathcal{L}_{ce}(Y_s, P_s) \quad (4)$$

\mathcal{L}_t is the target-domain loss including the online loss \mathcal{L}_t^{on} and offline loss \mathcal{L}_t^{off} . \mathcal{L}_{reg} is the regularization item include consistency loss \mathcal{L}_{cst} and dual region regularization loss $\mathcal{L}_{region}^{dual}$.

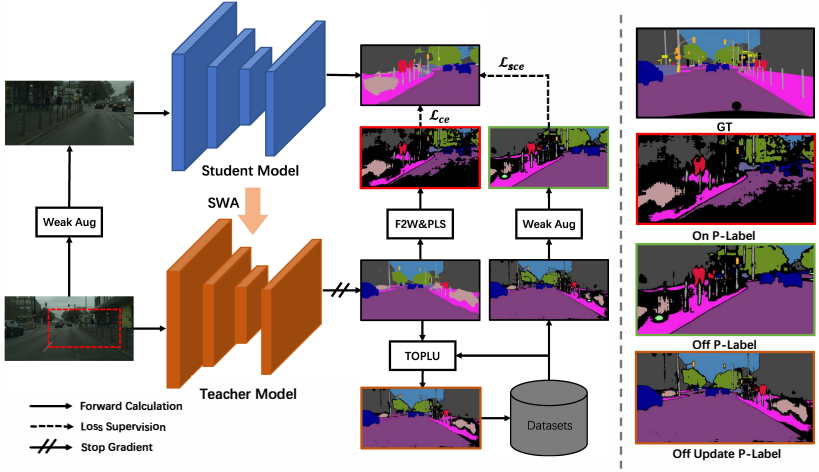


Figure 1: The simplified overview of our proposed DUDA. DUDA includes a student model (the blue one) and a teacher model (the orange one). The online pseudo labels are generated by the teacher model with the full resolution input firstly, then the Full-to-Weak (F2W) and Pseudo-Label-Selection (PLS) operations are applied. TOPLU combines the information of current predicted results and historical pseudo labels to update the offline pseudo label.

The consistency loss \mathcal{L}_{cst} can be written as follows:.

$$\mathcal{L}_{cst} = \mathbb{E}_x \left[\left\| F2W(h(x_T^{full} | \theta^{mean})) - h(x_T | \theta) \right\| \right] \quad (5)$$

in which x_T^{full} represents the full-resolution image without data augmentation. Different from MFA, we borrow the full-to-weak operation ($F2W(\cdot)$ in Equation. 5) from ProDA [41] to take advantage of the larger receptive field. The remaining losses will be described in the following section. Figure. 1 shows the simplified overview of our DUDA.

3.2 Online-Offline Dual Self-training

Online pseudo label update. The Online pseudo label update is one of the popular denoising methods due to its self-correcting property during the self-training process. We apply the online update to our framework following ProDA [41] and MFA [41]: our online pseudo label scheme follows MFA mostly, but we employ the full-to-weak operation in ProDA to generate more accurate online pseudo labels which can be written as:

$$\hat{Y}_t^{on} = PLS(F2W(P_{full})) \quad (6)$$

where P_{full} is the mean model’s softmax output of the full-resolution image data. After getting predicted results, the pseudo labels selecting strategy ($PLS(\cdot)$ in Equation. 6) is applied to the results and we employ Online-CBST presented in MFA for selecting high confidence pixels as pseudo labels. After getting \hat{Y}_t^{on} , we calculate the online pseudo loss as:

$$\mathcal{L}_t^{on} = \mathcal{L}_{ce}(\hat{Y}_t^{on}, P_t) \quad (7)$$

in which P_t refers to the student model’s softmax output for current batch input. Note that the online pseudo labels are updated on-the-fly as the parameters are updated during training. A more detailed introduction to Online-CBST is given in the supplementary material.

Offline pseudo label update. Previous self-training methods in UDA semantic segmentation update the offline pseudo labels by stage [20, 42, 46]. Inspired by online pseudo labels update and SWLF [21], we propose a *Temporal-aware Offline Pseudo Label Update* (TOPLU) algorithm to update offline pseudo labels and reduce the negative impact of noisy label via updating the offline labels by iteration. The proposed TOPLU algorithm consists of three steps. The first step is high-confidence acceptance and we use the Online-CBST again in this step to filter out the full-resolution image prediction results \hat{Y}_{curr} to get rough pseudo labels \hat{Y}'_{curr} . Note that our selection strategy is general and can be replaced by any other strategies such as fixed threshold. We employ Online-CBST for the sake of class balance. Then the two steps follows: 1) inconsistent discard and 2) consistent acceptance.

For the inconsistent discard step, the pixel labels that are inconsistent with previous pseudo labels are dropped even though they are regarded as “high confidence” according to the high-confidence acceptance step. This step is based on the assumption that the model may predict the inconsistent result for noisy labels during training. It can be written as:

$$\mathbb{I}_{[\hat{Y}_{prev} \neq \hat{Y}_{curr}]} \mathbb{I}_{[\hat{Y}_{prev} \neq I_{igr}]} (\hat{Y}'_{curr}) = I_{igr} \quad (8)$$

in which \mathbb{I} is the indicator function and \hat{Y}_{prev} is the previous offline pseudo labels.

For the consistent acceptance step, we design this step to add the pixel labels that are dropped by the high-confidence acceptance step but have consistency with historical pseudo labels. These pixels are believed to be correctly predicted according to our assumption. It can be written as:

$$\mathbb{I}_{[\hat{Y}_{prev} = \hat{Y}_{curr}]} \mathbb{I}_{[\hat{Y}_{prev} \neq I_{igr}]} (\hat{Y}'_{curr}) = \hat{Y}_{prev} \quad (9)$$

After these three steps, the updated offline pseudo label \hat{Y}_t^{off} is gotten and saved back to the dataset. We adopt symmetric cross-entropy(SCE) [43] loss on offline pseudo label to enhance the denoise ability.

$$\mathcal{L}_t^{off} = \alpha \mathcal{L}_{ce}(P_t, \hat{Y}_t^{off}) + \beta \mathcal{L}_{ce}(\hat{Y}_t^{off}, P_t) \quad (10)$$

where α and β are balancing coefficients and set to 0.1 and 1 respectively. Combining the online and offline supervision, the loss function of the target domain can be summarized as:

$$\mathcal{L}_t = \mathcal{L}_t^{off} + \lambda^{on} \mathcal{L}_t^{on} \quad (11)$$

the λ^{on} is a trade-off coefficient for online supervision. We employ sce loss for offline pseudo label while cross-entropy loss for online pseudo label. Such configuration is because that sce loss will make the model under-fit while cross-entropy loss is not robust to noise and the combination of the two loss is a better choice.

3.3 Online-Offline Dual Region Regularization

Previous arts [20, 31, 47] have proven the effectiveness of using regularization strategy. To begin with, we simply review the region regularization used in IAST [20]. Firstly, the KLD-regularization is employed to prevent the model over-fitting to the confident pixel via

minimize the KL-Divergence of the softmax output. Its formulation can be written as below following:

$$\mathcal{L}_{kld} = -\frac{1}{|X_t|} \sum_{x_t \in X_t} \mathbb{I}_{[\hat{y} \neq I_{igr}]} \frac{1}{C} \sum_{c=1}^C \log(p_c) \quad (12)$$

Besides, for the unlabelled pixels, the entropy minimum regularization is applied to make the distribution of the softmax output “sharper”. It can be written as following:

$$\mathcal{L}_{ent} = -\frac{1}{|X_t|} \sum_{x_t \in X_t} \mathbb{I}_{[\hat{y} = I_{igr}]} \frac{1}{C} \sum_{c=1}^C p_c \log(p_c) \quad (13)$$

However, most previous methods use regularization in the offline situation while the regularization for online pseudo labels is ignored. In DUDA we extend the region regularization into both offline and online situation and present so called dual regularization strategy. It can be formulated as below:

$$\mathcal{L}_{kld}^{dual} = \mathcal{L}_{kld}^{off} + \lambda^{on} \mathcal{L}_{kld}^{on}, \quad \mathcal{L}_{ent}^{dual} = \mathcal{L}_{ent}^{off} + \lambda^{on} \mathcal{L}_{ent}^{on} \quad (14)$$

where λ^{on} denotes the online part coefficient shared with Equation. 11. Combine the aforementioned consistency regularization, the \mathcal{L}_{reg} item in Equation. 3 is the sum of consistency regularization and dual region regularization:

$$\mathcal{L}_{region}^{dual} = \lambda^{ent} \mathcal{L}_{ent}^{dual} + \lambda^{kld} \mathcal{L}_{kld}^{dual}, \quad \mathcal{L}_{reg} = \mathcal{L}_{cst} + \mathcal{L}_{region}^{dual} \quad (15)$$

where λ^{ent} and λ^{kld} is the trade-off coefficient for entropy minimum and KL-Divergence regularization. \mathcal{L}_{cst} is the consistency regularization referred before.

It is worth noting that the offline and online pseudo labels are both dynamic updated and the labelled pixel get increased as training going on. Therefore, the entropy minimum regularization applied on unlabelled pixel mainly works in early stage while the KLD-regularization applied on labelled pixels becomes dominating in later stage as the labelled pixels become abundant. The KLD-regularization prevents the model overfitting to the confident pixel label.

4 Experiment

4.1 Experimental Setup

Implementation Details. We employ DeepLabv2 [8] as the segmentation model with ResNet 101 [2] backbone pre-trained on ImageNet-1k. FDA [36] is employed to warmup the model before self-training stage and CBST [46] is utilized to obtain the initial pseudo labels. Random resize and crop is applied as the data augmentation. We use the SGD optimizer with an initial learning rate of 0.002 and the poly learning rate adjustment strategy is used. We train the model for 80,000 iterations totally and the batch size is set as 4. Furthermore, the trade-off coefficient λ^{off} , λ^{ent} and λ^{kld} are set as 0.5, 2.0 and 0.1 respectively. The offline pseudo label update begins at 15,000th iteration. All experiments are conducted using two 2080Ti GPUs and implemented by Pytorch.

Dataset and Evaluation Protocol. To evaluate our proposed DUDA, we use synthetic datasets GTA5 [24] and SYNTHIA [25] that are widely used in cross-domain semantic segmentation tasks as the source-domain datasets. The real scene dataset Cityscapes [3] is

employed as the target-domain dataset. GTA5 contains 24,966 training images with a resolution of 1914×1052 and we use its 19 categories shared with Cityscapes. The SYNTHIA dataset includes 9,400 images with a resolution of 1280×760 . Following [40], we use its 13 categories shared with Cityscapes. Cityscapes dataset includes 2,975 training images and 500 validation images and all images have a resolution of 2048×1024 . In terms of the evaluation metrics, we exploit Insertion over Union (IoU) to measure the performance of the compared methods.

4.2 Comparative Studies

We compare DUDA with several state-of-the-art self-training-based UDA methods. Table. 1 shows the comparison results on GTA5-to-Cityscapes benchmark. Our DUDA achieves the state-of-the-art 58.4 mIoU score. Among the 19 categories, we achieve the best scores on 14 categories. Furthermore, we observe that DUDA achieves excellent performance on the head categories such as road, building and sky. We own such superior performance to our dual update strategy that brings more correct pseudo labels to head categories. While for some hard and tail categories, such as pole and terrain, our framework can be also on par with or outperform other rivals, *e.g.*, outperforming ProDA and MFA by 4.7 mIoU and 2.7 mIoU respectively.

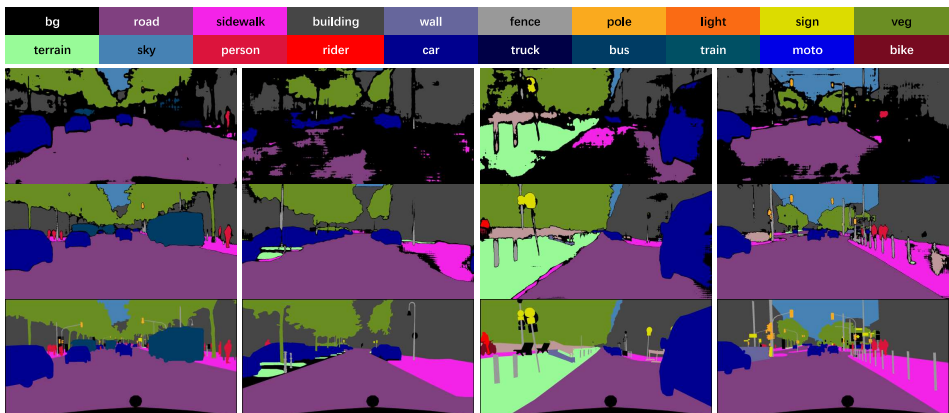


Figure 2: Visualization of the pseudo labels. First row shows the initial pseudo labels, and the second and third rows are the finally updated pseudo labels and GT labels.

Table. 2 shows the result on SYNTHIA-to-Cityscapes benchmark and exhibits the consistent performance as on GTA5-to-Cityscapes. Although the adaptation from SYNTHIA is more challenging than that from GTA5, our proposed DUDA also brings huge performance gains, reporting 59.7 mIoU on this benchmark.

4.3 Ablation Studies

Our baseline (ST-MT) employs a standard mean-teacher framework with consistency loss and naive self-training. In this baseline, naive self-training with sce loss and consistency loss with full-to-weak operation are employed which achieve 55.0 mIoU finally. To make our analysis more clear, we divide DUDA into a dual update module (DUPT) and a dual

method	road	sdwk	blndg	wall	fence	pole	light	sign	veg	trm	sky	psn	rider	car	trunk	bus	train	moto	bike	mIoU
CBST [14]	91.8	53.5	80.5	32.7	21.0	34.0	28.9	20.4	83.9	34.2	80.9	53.1	24.0	82.7	30.3	35.9	15.0	25.9	42.8	45.9
IntraDA [14]	90.6	36.1	82.6	29.5	21.3	27.6	31.4	23.1	85.2	39.3	80.2	59.3	29.4	86.4	33.6	53.9	0.0	32.7	37.6	46.3
WSDA [14]	91.6	47.4	84.0	30.4	28.3	31.4	37.4	35.4	83.9	38.3	83.9	61.2	28.2	83.7	28.8	41.3	8.8	24.7	46.4	48.2
SUDA [14]	91.1	52.3	82.9	30.1	25.7	38.0	44.9	38.2	83.9	39.1	79.2	58.4	26.4	84.5	37.7	45.6	10.1	23.1	36.0	48.8
CaCo [14]	91.9	54.3	82.7	31.7	25.0	38.1	46.7	39.2	82.6	39.7	76.2	63.5	23.6	85.1	38.6	47.8	10.3	23.4	35.1	49.2
IAST [14]	94.1	58.8	85.4	39.7	29.2	25.1	43.1	34.2	84.8	34.6	88.7	62.7	30.3	87.6	42.3	50.3	24.7	35.2	40.2	52.2
FDA [14]	92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	63.1	16.9	27.7	46.4	50.5
Seg-U [14]	90.4	31.2	85.1	36.9	25.6	37.5	48.8	48.5	85.3	34.8	81.1	64.4	36.8	86.3	34.9	52.2	1.7	29.0	44.6	50.3
TPLD [14]	94.2	60.5	82.8	36.6	16.6	39.3	29.0	25.5	85.6	44.9	84.4	60.6	27.4	84.1	37.0	47.0	31.2	36.1	50.3	51.2
ProDA [14]	91.5	52.4	82.9	42.0	35.7	40.0	44.4	43.3	87.0	43.8	79.5	66.5	31.4	86.7	41.1	52.5	0.0	45.5	53.8	53.7
MFA [14]	94.5	61.1	87.6	41.4	35.4	41.2	47.1	45.7	86.6	36.6	87.0	70.1	38.3	87.2	39.5	54.7	0.3	45.4	57.7	55.7
DUDA(ours)	94.6	66.4	87.0	41.5	41.2	48.7	47.6	47.6	87.8	46.8	87.2	72.3	38.5	89.1	38.9	61.4	0.0	51.5	61.1	58.4

Table 1: Results on GTA5-to-Cityscapes. Our proposed DUDA achieves competitive performance compared with other state-of-the-art methods. For a fair comparison, we present the self-training stage in ProDA and MFA.

region regularization module (DREG). Each module can be further divided into the online part and the offline part. Tabel. 3 shows the result of our ablation study.

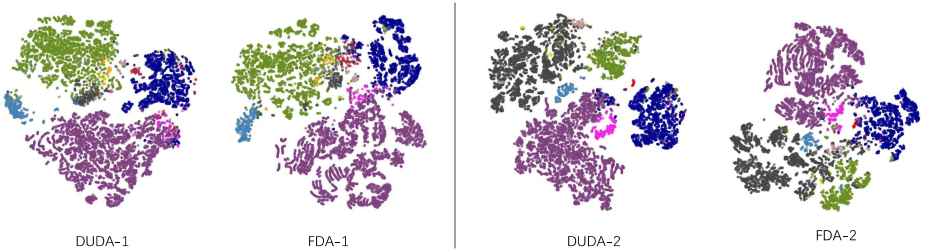


Figure 3: The feature space visualization for DUDA and FDA results via t-SNE.

The Effectiveness of Online-Offline Dual Update. The ablation study result of online-offline dual update can be seen in Table 3. The model achieves 56.9 mIoU after introducing the whole dual update module, bringing about 1.9 mIoU gain comparing with the baseline. Our ablation study further shows the effectiveness of online and offline update which brings 0.6 mIoU and 1.5 mIoU improvement respectively. Based on the observation, we can conclude that both online and offline pseudo label update strategy have their own contribution to the whole adaptation performance. Figure. 2 gives a more intuitive show of the pseudo label update result for offline part. We observe that the amount of labelled pixels are increased and some noisy labels are corrected comparing with those initial pseudo labels. These improvement further help self-training in later iteration. Furthermore, visualize the feature space of the model using t-SNE in Figure. 3 and we can find that the features can be better separated after DUDA and also these features are better gathered.

The Effectiveness of Dual Region Regularization. Table. 3 shows the effectiveness of the dual regularization strategy. The whole dual region regularization module brings 1.3 mIoU improvement comparing with the previous results and achieves 58.3 mIoU finally. Further ablation study shows the benefit of online and offline regularization part in detail. We observe that both online and offline region regularization can significantly benefit the segmentation performance. Particularly, our experiments prove that the region regularization applies on the online pseudo label brings a significant performance boost.












method	road	sdwk	bdng	light	sign	veg	sky	psn	rider	car	bus	moto	bike	mIoU
CBST 	68.0	29.9	76.3	22.8	29.5	77.6	78.3	60.6	28.3	81.6	23.5	18.8	39.8	48.9
IntraDA 	84.3	37.7	79.5	9.2	8.4	80.0	84.1	57.2	23.0	78.0	38.1	20.3	36.5	48.9
WSDA 	92.0	53.5	80.9	3.8	6.0	81.6	84.4	60.8	24.4	80.5	39.0	26.0	41.7	51.9
SUDA 	83.4	36.0	71.3	18.2	26.7	72.4	80.2	58.4	30.8	80.6	38.7	36.1	46.1	52.2
CaCo 	87.4	48.9	79.6	17.4	28.3	79.9	81.2	56.3	24.2	78.6	39.2	28.1	48.3	53.6
IAST 	81.9	41.5	83.3	30.9	28.8	83.4	85.0	65.5	30.8	86.5	38.2	33.1	52.7	57.0
FDA 	79.3	35.0	73.2	19.9	24.0	61.7	82.6	61.4	31.1	83.9	40.8	38.4	51.1	52.5
Seg-U 	87.6	41.9	83.1	31.3	19.9	81.6	80.6	63.0	21.8	86.2	40.7	23.6	53.1	54.9
TPLD 	80.9	44.3	82.2	20.5	30.1	77.2	80.9	60.6	25.5	84.8	41.1	24.7	43.7	53.5
ProDA 	87.1	44.0	83.2	45.8	34.2	86.7	81.3	68.4	22.1	87.7	50.0	31.4	38.6	58.5
MFA 	85.4	41.9	84.1	22.2	23.9	83.6	80.7	71.5	35.8	86.6	47.6	37.2	62.5	58.7
DUDA(ours)	84.4	43.4	80.3	29.3	28.9	75.6	88.1	69.3	33.8	88.1	60.1	47.0	57.8	59.7

Table 2: Results on SYNTHIA-to-Cityscapes benchmark. Our proposed DUDA achieves competitive performance compared with other state-of-the-art methods.

Components					mIoU	gain
ST-MT	DUPT		DREG			
	ONUPT	OFFUPT	ONREG	OFFREG		
✓					55.0	-
✓	✓				55.6	+0.6
✓		✓			56.5	+1.5
✓	✓	✓			56.9	+1.9
✓	✓		✓		56.4	+1.4
✓	✓	✓	✓		57.4	+2.4
✓	✓		✓	✓	57.0	+2.0
✓	✓	✓	✓	✓	58.3	+3.3

Table 3: The ablation study of proposed components on GTA5-to-Cityscapes benchmark. ST-MT: self-training with the mean-teacher framework. DUPT: dual update module. DREG: dual region regularization. ONUPT: online pseudo label update. OFFUPT: offline pseudo label update. ONREG: online region regularization. OFFREG: offline region regularization. Note that we discard the horizontal flip trick here and achieve 58.3 mIoU finally.

4.4 Conclusion

This paper focuses on self-training based UDA semantic segmentation, and we build a new framework dubbed DUDA which includes dual pseudo label update and dual region regularization module. The “dual” indicates both pseudo label update and region regularization are applied on the online and the offline pseudo labels. With the pseudo label update method, the noisy labels are reduced and more correct labels are added, this is the main reason why the dual update strategy brings large performance gain.

Acknowledgments. This work was supported by the National Key Research & Development Project of China (2021ZD0110700), the National Natural Science Foundation of China(U19B2043,61976185), Zhejiang Natural Science Foundation(LR19F020002), Zhejiang Innovation Foundation(2019R52002), and the Fundamental Research Funds for the Central Universities(226-2022-00051).

References

- [1] Philip Bachman, Ouais Alsharif, and Doina Precup. Learning with pseudo-ensembles. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.
- [2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European conference on computer vision (ECCV)*, pages 833–851, 2018.
- [3] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *European conference on computer vision (ECCV)*, pages 102–118, 2016.
- [4] Aritra Ghosh, Himanshu Kumar, Lujia Pan, and P. S. Sastry. Robust loss functions under label noise for deep neural networks. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, page 1919–1925, 2017.
- [5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [6] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [8] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution and fully connected crf. *IEEE transactions on pattern analysis and machine intelligence*, pages 834–848, 2017.
- [9] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Cross-view regularization for domain adaptive panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10133–10144, June 2021.
- [10] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. In *Advances in Neural Information Processing Systems*, volume 34, pages 3635–3649. Curran Associates, Inc., 2021.
- [11] Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1203–1214, June 2022.

- [12] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry P. Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 876–885, August 2018.
- [13] Ruihuang Li, Shuai Li, Chenhang He, Yabin Zhang, Xu Jia, and Lei Zhang. Class-balanced pixel-level self-labeling for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11593–11603, June 2022.
- [14] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [15] Yulei Lu, Yawei Luo, Li Zhang, Zheyang Li, Yi Yang, and Jun Xiao. Bidirectional self-training with multiple anisotropic prototypes for domain adaptive semantic segmentation. *arXiv preprint arXiv:2204.07730*, 2022.
- [16] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. Significance-aware information bottleneck for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6778–6787, 2019.
- [17] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2507–2516, 2019.
- [18] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. Adversarial style mining for one-shot unsupervised domain adaptation. *Advances in neural information processing systems*, 33:20612–20623, 2020.
- [19] Yawei Luo, Ping Liu, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Category-level adversarial adaptation for semantic segmentation using purified features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [20] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *European Conference on Computer Vision (ECCV)*, 2020.
- [21] Duc Tam Nguyen, Chaithanya Kumar Mummadi, Thi Phuong Nhung Ngo, Thi Hoai Phuong Nguyen, Laura Beggel, and Thomas Brox. Swlf: learning to filter noisy labels with self-ensembling. In *International Conference on Learning Representations (ICLR)*, 2020.
- [22] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [23] Sujoy Paul, Yi-Hsuan Tsai, Samuel Schuster, Amit K. Roy-Chowdhury, and Manmohan Chandraker. Domain adaptive semantic segmentation using weak labels. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 571–587, 2020.

- [24] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3213–3223, 2016.
- [25] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and M. LopezAntonio. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3234–3243, 2016.
- [26] Inkyu Shin, Sanghyun Woo, Fei Pan, and In So Kweon. Two-phase pseudo label densification for self-training based domain adaptation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 532–548, 2020.
- [27] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*, volume 33, pages 596–608. Curran Associates, Inc., 2020.
- [28] Cheng Tan, Jun Xia, Lirong Wu, and Stan Z. Li. Co-learning: Learning from noisy labels with self-supervision. In *ACM Multimedia*, October 2021.
- [29] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 1195–1204, December 2017.
- [30] Yi-Hsuan Tsai, Wei Chih Hung, Samuel Schuster, Kihyuk Sohn, Ming Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7472–7481, 2018.
- [31] Tuan-Hung Vu, Himalaya Jain, Bucher Maxime, Matthieu Cord, and Patrick Perez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2517–2526, 2019.
- [32] DengBao Wang, Yong Wen, Lujia Pan, and MinLing Zhang. Learning from noisy labels with complementary loss functions. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 10111–10119, 2021.
- [33] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.
- [34] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. Combating noisy labels by agreement: A joint training method with co-regularization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

- [35] Chang Wei-Lun, Wang Hui-Po, Peng Wen-Hsiao, and Wei-Chen Chiu. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [36] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4085–4095, 2020.
- [37] Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor W Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.
- [38] Bowen Zhang, Yidong Wang, Wenxin Hou, HAO WU, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *Advances in Neural Information Processing Systems*, volume 34, pages 18408–18419, 2021.
- [39] Jingyi Zhang, Jiaxing Huang, Zichen Tian, and Shijian Lu. Spectral unsupervised domain adaptation for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9829–9840, June 2022.
- [40] Kai Zhang, Yifan Sun, Rui Wang, Haichang Li, and Xiaohui Hu. Multiple fusion adaptation: A strong framework for unsupervised semantic segmentation adaptation. In *The British Machine Vision Conference (BMVC)*, 2021.
- [41] Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2021.
- [42] Qiming Zhang, Jing Zhang, Wei Liu, and Dacheng Tao. Category anchor-guided unsupervised domain adaptation for semantic segmentation. In *Advances in Neural Information Processing Systems*, pages 433–443, 2019.
- [43] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- [44] Zhedong Zheng and Yi Yang. Adaptive boosting for domain adaptation: Towards robust predictions in scene segmentation. *CoRR*, abs/2103.15685, 2021.
- [45] Zhedong Zheng and Yi Yang. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision (IJCV)*, 2021.
- [46] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pages 289–305, 2018.
- [47] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5982–5991, 2019.