Stable Neighbor Denoising for Source-free Domain Adaptive Segmentation

Dong Zhao[1], Shuang Wang[1] ✉, Qi Zang[1], Licheng Jiao[1], Nicu Sebe[2], Zhun Zhong[3,4] ✉
[1] School of Artificial Intelligence, Xidian University, Shaanxi, China
[2] Department of Information Engineering and Computer Science, University of Trento, Italy
[3] School of Computer Science and Information Engineering, Hefei University of Technology, China
[4] School of Computer Science, University of Nottingham, NG8 1BB Nottingham, UK

Abstract

We study source-free unsupervised domain adaptation (SFUDA) for semantic segmentation, which aims to adapt a source-trained model to the target domain without accessing the source data. Many works have been proposed to address this challenging problem, among which uncertainty-based self-training is a predominant approach. However, without comprehensive denoising mechanisms, they still largely fall into biased estimates when dealing with different domains and confirmation bias. In this paper, we observe that pseudo-label noise is mainly contained in unstable samples in which the predictions of most pixels undergo significant variations during self-training. Inspired by this, we propose a novel mechanism to denoise unstable samples with stable ones. Specifically, we introduce the Stable Neighbor Denoising (SND) approach, which effectively discovers highly correlated stable and unstable samples by nearest neighbor retrieval and guides the reliable optimization of unstable samples by bi-level learning. Moreover, we compensate for the stable set by object-level object paste, which can further eliminate the bias caused by less learned classes. Our SND enjoys two advantages. First, SND does not require a specific segmentor structure, endowing its universality. Second, SND simultaneously addresses the issues of class, domain, and confirmation biases during adaptation, ensuring its effectiveness. Extensive experiments show that SND consistently outperforms state-of-the-art methods in various SFUDA semantic segmentation settings. In addition, SND can be easily integrated with other approaches, obtaining further improvements. The source code is available at https://github.com/DZhaoXd/SND.

This work was supported in part by the National Natural Science Foundation of China No. 62271377, the Key Research and Development Program of Shannxi Program No. 2021ZDLGY0106 and No. 2022ZDLGY0112, the National Key R&D Program of China under Grant 2021ZD0110400 and 2021ZD0110404 the Key Scientific Technological Innovation Research Project by Ministry of Education, the Joint Funds of the National Natural Science Foundation of China (U22B2054).
✉ Co-corresponding author.

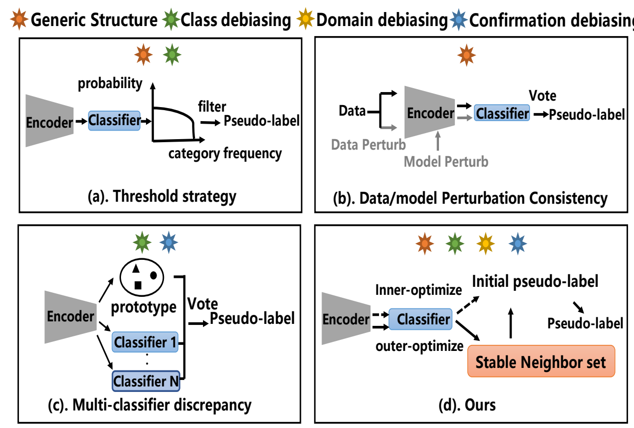

Figure 1. Comparison of advantages of different uncertain estimation strategies in self-training method for SFUDA.

1. Introduction

Unsupervised domain adaptation (UDA) [47] transfers knowledge of the labeled-rich source domains to the unlabeled target domains, providing an effective solution for semantic segmentation towards lower annotation burden [20] and strong generalization in an open environment [50, 71]. However, the access requirements of source domain data make it impossible to handle scenarios involving privacy, property rights protection, and confidentiality [15]. Thus, this work focuses on a more practical task in semantic segmentation, source-free unsupervised domain adaptation (SFUDA) [38, 75], which aims to adapt a source-trained model to a target domain without accessing the source data.

Currently, self-training is the mainstream technology to address the SFUDA problem in semantic segmentation, which strives to adapt the model with high-quality pseudo-labels produced by uncertain estimation functions, *e.g.*, probability thresholding [1, 21–23, 35, 73, 85], perturbation consistency [15,29,76], or discrepancy classifier voting [5, 29, 56, 65, 82, 89]. Generally, a strong cross-domain self-training method should comprehensively consider the following aspects. ❶ *Generic structure*: It does not require the adoption of a specific structural design (e.g., a specific segmentation head) on the source side; otherwise, it would not be conducive to target deployment [46]. ❷ *Class debiasing*: the less-learned (hard) categories in the source model should be paid more attention, as the class imbalance problem often exists in segmentation tasks [1, 21, 43, 58, 90]. ❸ *Domain debiasing*: the self-training method needs to maintain the adaptability of distinct domains, since the cross-domain segmentation tasks often face multiple or compound domains, *e.g.*, different weather conditions. ❹ *Confirmation debiasing*: since the running confidence of the model will increase with self-training [3, 86], a strong self-training method should be able to dynamically adjust the uncertain estimation function for addressing the confirmation bias. However, existing self-training methods for SFUDA, lacking a comprehensive denoising mechanism, often fall short in one or more aspects (see Fig. 1(a-c)), leading to inefficiency and under-adaptation.

Our motivation comes from experimental observations of implementing the vanilla self-training [1] on SFUDA. As shown in Fig. 2, for each target training sample at different iterations, we record 1) its mIoU scores with the ground truth (mIoU-axis); and 2) the change degrees between the segmentation results of the current step and the ones obtained by the initial step (stability-axis). Throughout the training process, we observe that samples with high stability consistently maintain high mIoU scores, while those with low stability consistently show low mIoU scores. This indicates that noise in samples with low stability is the main contributor to error accumulation during self-training. Moreover, this issue does not significantly improve as training progresses. This motivates us to identify stable samples (*i.e.*, whose pseudo-label change degree are low) and unstable samples to address the above issue.

To this end, we propose a novel denoising approach, called *Stable Neighbor Denoising* (SND), which can effectively discover highly correlated stable and unstable samples by nearest neighbor retrieval and denoise unstable samples with the guidance of the stable ones by bi-level optimization. Specifically, we resort to bi-level optimization [48, 70] to establish the connection between the stable and unstable sets and guide the evolution direction of unstable samples. The principle here is that if the model is trained on unstable samples with correct pseudo-labels, the loss on stable samples will be also minimized. Importantly, by the analysis of the gradient of bi-level optimization (detailed in Section 3.2), we find that matching unbiased and highly correlated stable samples for unstable samples is the key to achieving this principle. This is because the loss gradients of biased/low-correlation samples often have non-intersecting/different optimization directions, leading to bias in bi-level optimization for denoising. To tackle this problem, we propose retrieving stable

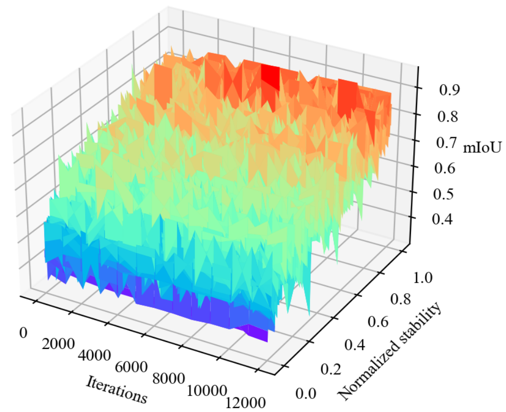

Figure 2. The plot of the mIoU scores versus stability for each target sample throughout training. Stability is calculated by the difference between the initial and the current segmentation. It shows a positive correlation between mIoU (pseudo-label quality) and stability. This observation holds during the whole training process. Experiments are from the GTA5 → Cityscapes SFUDA task.

neighbors by image style and spatial layout factors to eliminate the domain bias in one-step bi-level denoising optimization. In addition,we propose refining the stable samples into two sets: a whole-stable set and a category-stable set. Then, we compensate for stable neighbors using diverse object-level pasting to eliminate category bias. In this way, bi-level optimization can efficiently denoise unstable samples with stable ones under the principle.

Through the above techniques, the proposed SND shows superiority over previous works as follows. ❶ *Generic structure*: SND does not require a specific network structure on both the source and target sides, endowing its universality. ❷ *Class debiasing*: SND constructs stable neighbors with balanced categories and uses them to guide the adaptation of less-learned categories. ❸ *Domain debiasing*: SND denoises distinct domains by retrieving neighbors with similar domain factors and can handle complex compound domain scenarios (See Table. 3). ❹ *Confirmation debiasing*: SND designs a stability-oriented denoising mechanism that filters out noisy pseudo-labels without the need for probability or prototype distance, thereby mitigating confirmation bias (See Fig. 6). Extensive experiments show that SND consistently outperforms state-of-the-art methods in various SFUDA semantic segmentation settings.

2. Related Work

Unsupervised Domain Adaptation. Domain adaptive semantic segmentation has achieved significant adaptability improvements in multiple adaptation scenarios such as synthesis-to-reality [10, 21–23, 56], cross-weather [31, 53, 54], multi-source [18, 69], etc. Overall, the current work mainly realizes cross-domain transfer from the following aspects. 1) Use the source domain data to learn a generalized representation [8]. This line of work employs data augmentation (e.g. copy-paste [58], random masking [23]) and domain randomization (e.g. expanding image styles [12, 24, 30, 56, 81, 88]) to expand the representation space of source domain models. 2) Align the source and target domains. The following works adopt a variety of domain alignment strategies (e.g. adversarial training or statistical matching) in a variety of alignment spaces (e.g. pixels space [7, 20], Fourier space [77], features space [49, 63, 66], output space [41, 59, 61], etc.), reducing the statistical difference of both domains. 3) Target characteristic mining. Most of such works use augmentation consistency [2, 11, 44, 82] and pseudo-labels [6, 81] to further improve the model's adaptive ability, *e.g.* tail category and local distribution.

Source-Free Unsupervised Domain Adaptation (SFUDA). In classification tasks, prior SFUDA works propose implicit distribution alignment [14, 37], instance contrast [83], and model perturbation [27] to align domains without source data. In segmentation tasks, the above idea is hard to adopt due to complex semantic feature spaces. Most SFUDA works for segmentation adopt self-training, which filters and retrains the well-adapted pixels or regions through probability thresholds [85], data [25] or model discrepancy [15, 29], but they often fall into various biases.

Noisy Label Learning. SFUDA can be viewed as a noisy label learning problem [74, 78, 87]. Currently, noisy label learning is mostly discussed on classification tasks. The mainstream technologies include robust loss design [42, 67, 84], self-label correction [9, 33], prototype denoising [17], meta-learning based denoising [48, 70], etc. Most of those methods are designed for instance-level and are not suitable for pixel-level segmentation tasks. In particular, the prototype [34, 82] and meta-learning [16] based denoising methods have been applied to UDA segmentation tasks, but their dependence on source data limits on SFUDA.

3. Method

Preliminary. In the setting of source-free unsupervised domain adaptation (SFUDA), we are given a segmentation model $\mathcal{H}$ with parameter Θ pre-trained on the labeled source dataset $\mathcal{D}_s = \{(x_s^i, y_s^i)\}_i^{N_s}$, and the unlabeled target dataset $\mathcal{D}_t = \{x_t^i\}_i^{N_t}$. The goal is to adapt the network $\mathcal{H}$ and achieve low risk on the target dataset without accessing the source data $\mathcal{D}_s$. To achieve this goal, most works mainly conduct self-training to optimize the $\mathcal{H}$ as follows,

$$\Theta^{\star} = \arg\min_{\Theta} \sum_{i}^{N_t} \sum_{l}^{H\times W} \omega^{(i,l)} \mathcal{L}[\mathcal{H}(x_t^{(i,l)}|\Theta), \hat{y}_t^{(i,l)}], \quad (1)$$

where $\mathcal{L}$ is the cross-entropy loss and $\hat{y}_t$ is the pseudo-label. ω is a weighting factor calculated by uncertainty, in which the lower the uncertainty, the closer the value is to 0. This paper proposes stable neighbor denoising techniques to estimate ω in an unbiased way.

3.1. Division of Stable & Unstable Sets

Following the observation in Fig 2, we aim to divide target domain samples into stable and unstable sets by the change degree (also called evolution stability) of their segmentation results during the vanilla self-training [1]. Formally, for each target sample x_t^i, we define its evolution stability (ES) at τ-th iteration as follow,

$$\mathrm{ES}^{i,\tau} = \sum_{l}^{H\times W} \mathrm{SIM}[\mathcal{H}(x_t^{i,l}|\Theta^0), \mathcal{H}(x_t^{i,l}|\Theta^{\tau})], \quad (2)$$

where $\mathrm{SIM}[\cdot]$ is the cosine similarity. Θ^{τ} represent the parameters of the model at τ-th iteration during self-training. The larger the ES value of the sample, the more stable it is. As ES may fluctuate in different degrees of domain shift tasks, it is hard to set a suitable threshold to determine whether it is stable. Instead, we take the top-$k\%$ ranked highly stable samples to divide the $\mathcal{D}_t$ into the stable set $\mathcal{D}_{se}$ and the unstable set $\mathcal{D}_{ue}$ as follows,

$$\mathcal{D}_{se} = \{(x_t^i, \bar{y}^i)|x_t^i \in \mathcal{D}_t, \mathrm{ES}^{i,\tau} \in \top_k\}, \quad (3)$$

$$\mathcal{D}_{ue} = \{x_t^i|x_t^i \in \mathcal{D}_t, \mathrm{ES}^{i,\tau} \notin \top_k\}. \quad (4)$$

Next, the main challenge is how to utilize $\mathcal{D}_{se}$ to assist in predicting unbiased ω of $\mathcal{D}_{ue}$.

3.2. Build Relations between Stable & Unstable Sets

An intuitive way is to exploit the cross-image feature similarity [68] of samples in the two sets to build relations. However, we find that the un-adapted source-trained model fails to capture this cross-image feature similarity because the representation ability of deep networks is always insufficient under the domain shift (see Table 3). Instead, we propose to build relations between the two sets from an optimization perspective and utilize bi-level learning [48, 70] to achieve this. In a nutshell, we formulate each optimization step as the inner- and outer loops, in which the former learns uncertainty parameter ω for the latter.

Specifically, for m-th optimization step, in the inner loop, the optimizable ω is first initialized as ω_0 and the model with parameters Θ^m is optimized on unstable samples by Eq.1,

$$\Theta_{inner}^{m+1} = \arg\min_{\Theta} \sum_{x_t \in \mathcal{D}_{ue}} \sum_{l}^{H\times W} \omega_0^i \mathcal{L}[\mathcal{H}(x_t^{(i,l)}|\Theta^m), \hat{y}_t^{(i,l)}]. \quad (5)$$

ω_0^i is simply set to be an all-one matrix. For optimization, Eq. 5 can be approximated by one- or multi-step gradient calculation, *i.e.*, $\Theta_{inner}^{m+1} = \Theta^m - \alpha \sum_{x_t \in \mathcal{D}_{ue}} \omega_0^i \frac{\partial \mathcal{L}(x_t^i|\Theta^m)}{\partial \Theta}|_{\Theta^m}$, where α is the inner-loop learning rate. It can be seen that ω is differentiable relative to Θ_{inner}^{m+1}. Subsequently, we optimize the ω by enforcing the optimized model with parameters Θ_{inner}^{m+1} also achieves low risk on the stable samples in $\mathcal{D}_{se}$,

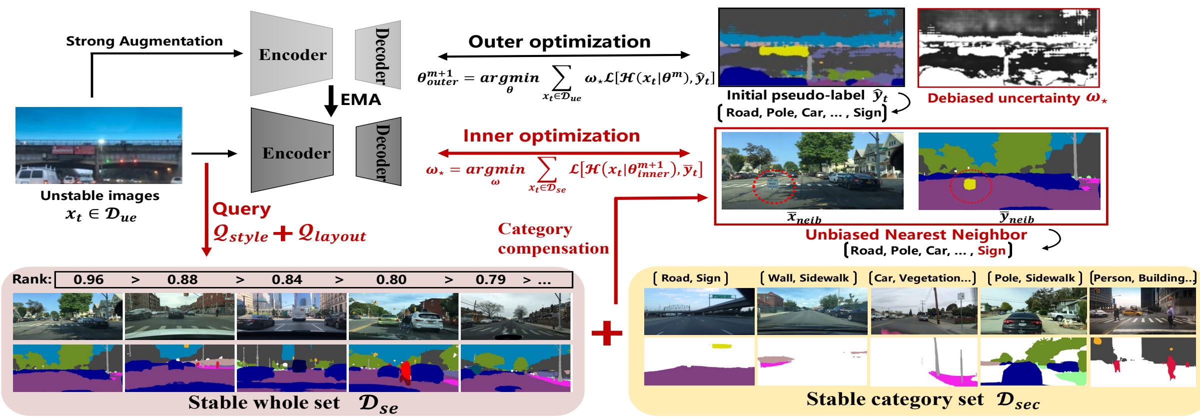

Figure 3. The pipeline of the proposed *Stable Neighbor Denoising* (SND). It is formed by the student-teacher model [57]. In each optimization, SND performs the inner and outer optimizations sequentially. In the inner loop (red line), SND utilizes the style $\mathcal{Q}_{style}$ and the layout $\mathcal{Q}_{layout}$ factors to retrieve stable neighbors for unstable samples and then performs category compensation to reduce category bias. Thereafter, SND executes Eq. 8 using the teacher model to obtain the unbiased uncertainty map $\omega_\star$ and initial pseudo-label $\hat{y}_t$. In the outer loop (black line), SND performs Eq. 7 to optimize the student model. EMA denotes the Exponential Moving Average.

$$\omega_\star = \arg\min_{\omega} \sum_{(x_t, \bar{y}_t) \in \mathcal{D}_{se}} \sum_{l}^{H \times W} \mathcal{L}[\mathcal{H}(x_t^{(i,l)}|\Theta_{inner}^{m+1}), \bar{y}_t^{(i,l)}]. \tag{6}$$

This object is driven by the following *principle*: ω is optimized to increase/decrease the weight of these regions so that the model optimized on the unstable set samples has a small/large loss on the stable set. In this way, we establish an implicit relation between stable and unstable sets, because optimizing noisy regions of unstable samples has misleading optimization goals. Finally, in the outer loop, the model can be optimized by the learned $\omega_\star$,

$$\Theta_{outer}^{m+1} = \arg\min_{\Theta} \sum_{x_t \in \mathcal{D}_{ue}} \sum_{l}^{H \times W} \omega_\star \mathcal{L}[\mathcal{H}(x_t^{(i,l)}|\Theta^m), \hat{y}_t^{(i,l)}]. \tag{7}$$

Problem Discussion. One important question in the above learning strategy is, does this *principle* work by taking arbitrary samples from the stable and unstable set to optimize? We present further analysis of the optimization of ω to answer this question. For optimization, Eq. 6 is performed by gradient descent as follows,

$$\begin{aligned}
\omega_\star &= \omega_0 - \beta \sum_{x_t \in \mathcal{D}_{se}} \frac{\partial \mathcal{L}(x_t^i|\Theta_{inner}^{m+1})}{\partial \omega}|_{\omega_0} \\
&= \omega_0 - \beta \sum_{x_t \in \mathcal{D}_{se}} \frac{\partial \mathcal{L}(x_t^i|\Theta_{inner}^{m+1})}{\partial \Theta}|_{\Theta_{inner}^{m+1}} \cdot \frac{\partial \Theta_{inner}^{m+1}}{\partial \omega}|_{\omega_0} \\
&= \omega_0 + \\
&\beta \underbrace{\sum_{x_t \in \mathcal{D}_{se}} \frac{\partial \mathcal{L}(x_t^i|\Theta_{inner}^{m+1})}{\partial \Theta}|_{\Theta_{inner}^{m+1}}}_{\text{Loss gradient of stable set}} \cdot \alpha \underbrace{\sum_{x_t \in \mathcal{D}_{ue}} \frac{\partial \mathcal{L}(x_t^i|\Theta^m)}{\partial \Theta}|_{\Theta^m}}_{\text{Loss gradient of unstable set}} .
\end{aligned} \tag{8}$$

Thus, when the continuous loss gradient on the stable samples in $\mathcal{D}_{se}$ has a similar direction to that on the unstable samples in $\mathcal{D}_{ue}$, the ω value of the corresponding area of the unstable samples will increase, and vice versa. This indicates that optimizing noisy pseudo-labels would deviate from the optimization goal of stable samples and thus explains why noise in pseudo-labels can be estimated. Moreover, Eq. 8 also gives us the answer that arbitrarily selecting samples from the two sets will induce bias in the estimation on ω, especially for samples with low correlation. This is because heterogeneous data always enjoy different optimal weight spaces [26], which will mislead the estimation of ω. We provide below a detailed analysis.

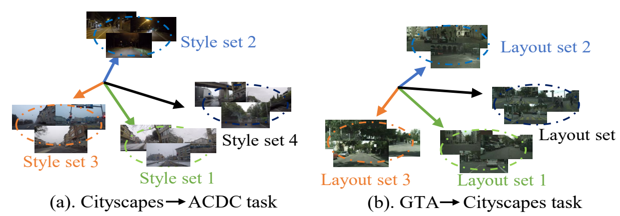

Figure 4. Loss gradient directions in bi-level optimization. It is shown by computing the average gradient direction of different sets of classifier weights (ASPP) using the true labels. The black lines are inner-loop and the other colored lines are outer-loop.

Reason & Analysis (1) We find that samples from both sets suffering large domain shifts tend to have large differences in optimization directions, which will cause the optimization of ω in Eq. 8 to be weakened by domain differences rather than the pseudo-label noise. As shown in Fig. 4, despite using real labels in bi-level optimization, domain factors (*e.g.* the spatial layout or image style) will still bias the optimization direction. (2) We find that samples from both sets with inconsistent category distribution cannot produce intersecting gradients in Eq. 8, causing the corresponding gradient of ω to be 0 as well. Moreover, in actual optimization, it is unacceptable to use massive stable samples to perform multi-step gradient descent for Eq. 6 due to huge computational and memory overhead. As a result, stable and unstable samples always suffer from mismatched class distributions, resulting in no gradient for missing classes.

In conclusion, it is reasonable to argue that matching highly correlated samples from two sets is critical for unbiased uncertain estimation.

3.3. Stable Neighbor Denoising

We propose to query the stable neighbor with similar domain properties for each unstable sample and compensate for missing categories for the neighbor, resulting in an unbiased estimate of ω. In the querying part, following the observation in Eq. 4, we specify domain factors, namely image style and spatial layout, as the proxy for querying.

Specifically, *for image style*, a series of works [36, 77] pointed out that the Fourier amplitude spectrum has a strong correlation with the image style. Thus, we adopt the flattened vector of the low-frequency Fourier amplitude spectrum as the style proxy, *i.e.*, $\mathcal{Q}_{style} = Low(\mathcal{A}(\mathcal{F}(x))$. *For spatial layout*, [32] shows that the row-wise and column-wise label statistical histogram vectors can represent the spatial distribution. Thus, we use the pseudo-labels of both sets samples for computation as an alternative, *i.e.*, $\mathcal{Q}_{layout} = [\mathcal{Q}^{row}_{layout}, \mathcal{Q}^{col}_{layout}]$, where $Q^{row,c}_{layout} = \frac{\sum_{row} \arg\max[\mathcal{H}(x^l_t|\Theta^m)]=c}{\sum_{row}\sum_{column} \arg\max[\mathcal{H}(x^l_t|\Theta^m)]=c}$, $c \in [0, 1, ...C]$, and C is the number of categories. With $\mathcal{Q}_{style}$ and $\mathcal{Q}_{layout}$, we can retrieve the nearest stable neighbor $x^i_{neib} \in \mathcal{D}_{se}$ for the unstable sample $x^i \in \mathcal{D}_{ue}$ via the similarity between the proxy vectors.

The retrieved domain-associated neighbors may still suffer from missing categories, leading to problems described in the analysis (2). Thus, we further propose to borrow the missing class objects from other stable samples and paste them to the neighbor as compensation. However, the tail categories in the stable set $\mathcal{D}_{se}$ may be scarce, making the objects to be copied far less abundant. To enrich the diversity of tail categories, we additionally maintain top-k ranked stable category sets $\mathcal{D}^c_{sec} = \{(x^i, \bar{y}^i)|x^i \in \mathcal{D}_t, \mathrm{ES}^{i,m,c} \in \top_k\}$ for each category c as the supplement. We name the $\mathcal{D}_{se}$ as a stable whole set and the $\mathcal{D}^c_{sec}$ as a stable category set for distinction. Subsequently, the nearest stable neighbor $x^i_{neib} \in \mathcal{D}_{se}$ can be compensated as follows,

$$\bar{x}^i_{neib}, \bar{y}^i_{neib} = \mathrm{CP}[(x^i_{neib}, \bar{y}^i_{neib}), (x_{mix}, \bar{y}_{mix}) \in \mathcal{D}^c_{sec}], \tag{9}$$

where $\mathrm{CP}[\cdot, \cdot]$ is the copy-paste operation [58] that copies objects from the latter to the former.

Through these efforts, we can use unbiased stable neighbors $\bar{x}^i_{neib}$ for Eq. 6 to denoise unstable samples in an unbiased way. Fig. 3 shows the pipeline of the proposed SND and the corresponding is shown in Algorithm 1. Algorithm 1 implements our SND within the student-teacher framework [57] to align with existing methods. See detailed explanations and ablation in the Appendix.

Algorithm 1 The training step of SND.

Input: Target data $\mathcal{D}_t$, source-trained segmentation network $\mathcal{H}$ with parameter Θ, max iteration M.
Output: Well-adapted segmentation model $\mathcal{H}$.

```
Split D_t into D_se and D_ue via vanilla self-training by Eq. 2, Eq. 3 and Eq. 4
Initialize the teacher student model as Θ_tea and Θ_stu
for m ← 0 to M − 1 do
    Sample batch samples {x_t^i}_{i=1}^B ∈ D_ue
    Inner loop Optimization:
    Retrieve nearest stable neighbor (x_neib^i, y_neib^i) ∈ D_se for each unstable sample x_t^i by Q_style and Q_layout
    Compensate (x_neib^i, y_neib^i) by Eq. 9
    Optimize Θ_tea by Eq. 8 to obtain the uncertainty ω_⋆ and pseudo-label ŷ_t
    Outer loop Optimization:
    Update Θ_stu by Eq. 7 with ω_⋆ and ŷ_t.
    Update Θ_tea by exponential moving average
end for
Return Adapted model H with Θ_tea.
```

4. Experiments

4.1. Datasets & Setup

Datasets. GTA5 [51] dataset provides 24,999 game-rendering urban scene images with a resolution of 1914×1024. SYNTHIA [52] dataset includes 9,400 virtual images with a resolution of 1280×760. We use 19 and 16 common categories in these two datasets respectively as source data. Cityscapes [13] dataset contains 3,975 real urban scene images from 50 different cities in primarily Germany, with a resolution of 2048×1024. ACDC [55] is the real-world dataset on adverse visual conditions, which comprises diverse weather scene images with a resolution of 1280×720. BDD100K [80] is a compound real-world dataset consists of 8000 training images and 1000 verification images with a resolution of 1280×720.

Setup. We conduct experiments using the Convolutional and Transformer structures respectively. For Convolutional structures, we adopt the Deeplab-v2 [4] with ResNet-101 [19] as the segmentation model. SGD optimizers are used for both inner and outer optimization, where the outer optimizer is set an initial learning rate (β) to 2.5×10^{-4} with weight decay 0.01 while the inner optimizer is set a fixed learning rate (α) of 0.01. The batch size is set as 4, and the framework is trained for 20,000 iterations on all SFUDA tasks. For Transformer structures, we adopt the SegFormer

	road	sidewalk	Building	Wall	fence	pole	light	sign	vege.	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
Source-free Synthetic-to-Real: GTA → Cityscapes (Val.)																				
HCL (NIPS'21) [25]	92.6	54.6	82.8	33.2	26.2	39.8	38.1	31.9	84.5	38.6	85.3	61.3	30.2	85.4	33.1	41.6	14.4	27.3	44.0	49.7
SFDASEG (ICCV'21) † [29]	91.7	53.4	86.1	37.6	32.1	37.4	38.2	35.6	86.7	48.5	89.9	62.6	34.3	87.2	51.0	50.8	4.2	42.7	53.9	53.4
DTST (CVPR'23) [85]	93.5	57.6	84.7	36.5	25.2	33.4	**44.7**	36.7	86.8	42.8	81.3	62.3	37.2	88.1	48.7	**50.6**	35.5	48.3	59.1	55.4
CrossMatch(ICCV'23)† [79]	95.1	67.8	87.7	51.3	41.5	36.3	47.4	51.3	87.8	47.8	87.3	67	34.2	87.5	41	51.8	0	42.6	46.4	56.4
CROTS(IJCV'24) [40]	92	52.4	85.9	37.3	35.8	34.6	42.2	38.4	86.9	45.6	91.1	65.1	36.1	87.3	41.6	51.1	0	41.4	56.2	53.7
SND (Ours)	93.0	54.0	84.6	35.6	30.3	31.0	41.9	41.6	87.6	44.6	86.4	62.6	38.5	87.5	48.7	42.9	**36.6**	49.5	58.7	55.6
DTST + SND (Ours)	**93.9**	**60.0**	**86.7**	**38.6**	**35.9**	**37.5**	43.4	**48.3**	**87.6**	**44.6**	**90.1**	**65.3**	**39.9**	**88.5**	**54.9**	44.4	33.1	**49.9**	**60.9**	**58.1**
Source-free Synthetic-to-Real: Synthia → Cityscapes (Val.)																				
HCL (NIPS'21) [25]	86.7	38.1	82.7	10	0.6	30.3	25.4	29.7	82.8	-	85.9	61.9	24.8	84.5	-	38.9	-	22.6	37.9	46.4
SFDASEG (ICCV'21) † [29]	**90.5**	**50.0**	81.6	13.3	2.8	34.7	25.7	33.1	83.8	-	**89.2**	**66.0**	**34.9**	85.3	-	53.4	-	46.1	46.6	52.3
DTST (CVPR'23) [85]	88.9	45.8	83.3	13.7	0.8	32.7	31.6	20.8	85.7	-	82.5	64.4	27.8	**88.1**	-	50.9	-	37.6	**57.3**	50.7
CrossMatch(ICCV'23)† [79]	91.5	55.5	**85.4**	**34.4**	8.3	**40.8**	**40.0**	44.4	86.6	-	84.3	62.4	22.0	**88.3**	-	**60.0**	-	40.6	45.6	55.6
CROTS(IJCV'24) [40]	89.4	41.6	82.7	15.1	1.2	34.7	33.7	25.7	83.7	-	87.9	66.6	34.6	85.4	-	45.9	-	43.5	49.6	51.3
SND (Ours)	88.1	47.4	80.1	28.1	**32.2**	34.9	33.6	41.3	83.3	-	86.7	59.9	27.2	86.7	-	48.1	-	36.2	52.5	54.1
DTST + SND (Ours)	88.7	43.7	83.6	32.1	26.0	32.4	38.0	**44.7**	**87.2**	-	87.9	62.2	**35.5**	87.4	-	40.4	-	**46.9**	57.2	**55.9**
Source-free clean-Adverse-Weather: Cityscapes → ACDC (Test)																				
URMDA (CVPR2021) † [15]	74.1	25.6	43.5	19.5	24.3	25.9	48.6	43.2	66.2	27.5	79.8	45.0	20.9	70.7	36.8	35.9	32.7	27.7	29.9	40.9
HCL (NIPS2021) [25]	72.6	24.7	68.5	21.4	19.4	31.0	51.8	46.5	71.8	28.5	81.2	43.8	21.5	76.8	42.8	44.4	36.1	30.1	24.0	44.0
SFDASEG (ICCV21) † [29]	73.5	29.0	70.7	19.7	21.9	36.1	53.2	51.4	72.0	31.1	85.3	41.5	26.2	76.4	41.9	46.1	39.8	33.3	32.0	46.4
DTST (CVPR2023) [85]	73.4	28.5	69.7	16.7	20.5	32.8	50.2	51.1	71.5	30.8	85.0	50.8	22.3	71.6	40.3	41.7	35.6	31.9	38.2	45.4
SND (Ours)	**74.8**	27.6	69.3	**23.6**	**26.3**	36.5	52.8	54.6	75.5	35.0	84.8	52.5	25.5	78.8	44.9	48.1	37.6	31.1	34.9	48.1
DTST + SND (Ours)	73.4	**29.4**	**70.9**	22.0	25.1	**38.5**	**54.6**	**55.5**	**77.7**	**35.2**	**86.6**	**53.4**	**27.4**	**80.9**	**45.6**	**49.0**	**41.0**	**36.7**	**40.4**	**49.6**
Source-free Open-compound: GTA → BDD100k (Test)																				
URMDA (CVPR'21) † [15]	83.9	38.3	78.7	9.6	7.3	29.1	11.1	4.9	70.7	-	74.2	53.8	15.0	81.2	-	35.0	-	22.8	30.5	40.4
HCL (NIPS'21) [25]	88.6	39.2	81.0	8.2	7.9	28.4	11.4	5.7	71.0	-	77.2	54.2	16.0	81.8	-	41.4	-	22.6	31.4	41.6
SFDASEG (ICCV'21) † [29]	**87.9**	40.2	**80.6**	13.1	8.2	30.2	22.8	17.1	71.1	-	78.1	51.4	27.9	80.2	-	**43.7**	-	30.3	42.3	45.3
DTST (CVPR'23) [85]	83.1	39.9	64.9	8.9	14.5	29.5	27.0	27.1	**71.9**	-	83.2	52.9	31.3	74.7	-	41.1	-	30.0	42.1	45.2
SND (Ours)	84.1	42.6	74.1	15.2	21.2	31.1	31.0	25.5	70.4	-	83.9	52.8	33.9	79.9	-	39.1	-	37.5	41.9	47.8
SND + DTST (Ours)	86.5	**44.4**	77.3	**21.3**	**22.9**	**32.4**	**33.0**	**27.4**	69.6	-	**86.7**	**54.3**	**34.3**	**82.1**	-	39.4	-	**38.3**	**42.4**	**49.5**

Table 1. Comparison of SND with state-of-the-art works on the tasks of source-free domain adaptation in semantic segmentation. The model is deeplab-v2 with ResNet101. The report metric is IoU(%). †denotes using the specific network on the source side, *e.g.*, SFDASEG using multiple heads, CrossMatch using two segmentation models with depth estimation. DTST+SND means using the minority class resampling strategy in DTST [85], as minority class adaptation is very challenging in source-free UDA.

[72] with MiT-B5 [72] . AdamW [39] and Adam [28] optimizer are used for inner and outer optimization respectively, where the outer optimizer is set an initial learning rate (β) 6×10^{-5} with weight decay 0.01, while the inner optimizer is set a fixed learning rate (α) of 0.01. The batch size is set as 2 and the model is trained for 20,000 iterations on all tasks. For set dividing, we use the stability (ES) evaluated at 10k (τ) iterations and regard the top 5% ($k\%$) ranked stable samples as the stable set for all adaptation tasks.

Continual source-free adaptation: Cityscapes→ACDC					
Time	$t \longrightarrow$				
	Fog	Night	Rain	Snow	mIoU
Source model [72]	69.1	40.3	59.7	57.8	56.7
TENT * [62]	68.5	36.3	59.9	54.7	54.9
CoTTA * [64]	70.4	41.6	63.9	60.8	59.2
HCL [25]	70.0	39.9	63.7	61.2	58.7
SFDASEG † [29]	70.1	42.1	62.4	61.8	59.1
SND	**72.1**	**43.1**	**66.3**	**65.6**	**61.8**

Table 2. Comparison on the tasks of continual source-free domain adaptation semantic segmentation. The segmentation model is SegFormer with MiT-B5 as the backbone. *means we use the target domain data for multiple round adaptations rather than one round in the original paper.

4.2. Comparison with State-of-the-art Alternatives

Performance Comparison on SFUDA. We compare our methods with the state-of-the-art approaches on source-free unsupervised domain adaptive semantic segmentation (SFUDA), including adapting to single and compound target domains. Table 1 shows that the proposed SND achieves the best performance on all SFUDA tasks. Moreover, SND can be combined with other method (DTST [85]) and achieves further significant improvements. In the two synthetic-to-real SFUDA adaptation tasks, our approach (SND+DTST) surpasses the second-highest methods by 2.7% and 3.1%, respectively. On real-to-real adaptation tasks, *i.e.*, ACDC with various weather conditions, our approach (SND+DTST) still maintains excellent performance, improving the performance of the current best method (SFDASEG [29]) by 3.2%. When adapted to the more challenging open-compound domain BDD100k, the proposed method (SND+DTST) achieves a large performance gain, surpassing the second-highest method by 4.2%. The performance under various conditions proves that our method can effectively cope with the pseudo-label noise in complex environments, and alleviate the bias problem in SFUDA tasks.

	Unweighted	PE. [62]	MCC. [89]	DAC. [29]	MPC. [15]	PD. [82]	CIA. [68]	SND
GTA → Cityscapes	49.6	50.5 (+0.9)	52.1 (+2.5)	51.7 (+2.1)	51.6 (+2.0)	50.9 (+1.3)	49.8 (+0.2)	**55.6 (+6.0)**
Synthia → Cityscapes	48.2	49.5 (+1.3)	52.9 (+2.7)	49.4 (+1.2)	51.5 (+3.3)	49.7 (+1.5)	47.2 (-1.0)	**54.1 (+5.9)**
Cityscapes → ACDC	43.1	41.3 (-1.8)	45.1 (+2.0)	42.9 (-0.3)	43.0 (-0.1)	39.1 (-4.0)	39.2 (-3.9)	**48.1 (+5.0)**
GTA → BDD100k	42.1	41.5 (-0.6)	44.8 (+2.7)	41.2 (-0.9)	42.9 (+0.8)	40.5 (-1.6)	37.1 (-5.0)	**47.8 (+5.7)**

Table 3. Ablation experiments on uncertainty estimation for source-free cross-domain segmentation tasks. We compared the following uncertainty estimation methods, probability entropy (PE.) [62], multi-classifier consistency (MCC.) [89], data augmentation consistency (DAC.) [29], model perturbation consistency (MPC.) [15], prototype distance (PD.) [82], cross-image association [68](CIA.). 'Unweighted' means original pseudo labels.

Division metric	GTA → Cityscapes	Cityscapes → ACDC
Image-level entropy [45]	52.9	45.1
Patch-level entropy [59]	51.7	43.6
Image-level loss [74]	52.7	44.1
Domain distance [60]	51.6	42.4
Stability	**55.6**	**48.1**

Table 4. Comparison of the subset division metrics.

Query	GTA → Cityscapes	Cityscapes → ACDC
Random	54.2	44.7
$\mathcal{Q}_{layout}$	55.3	44.9
$\mathcal{Q}_{style}$	55.1	47.5
$\mathcal{Q}_{layout} + \mathcal{Q}_{style}$	**55.6**	**48.1**

Table 5. Ablation study of querying stable neighbor strategies. Random means randomly selecting stable samples from $\mathcal{D}_{se}$.

Performance Comparison on Continual SFUDA. As real-world machine systems always operate in non-stationary environments, we also verify the effectiveness of our approach in the continual adaptation setting. In Table 2, our method maintains good performance, with an average adaptability increase of 5.1% in multiple domains under continual settings, showing obvious advantages compared with other methods. It is verified that our method can better adapt to multi-domain environments and alleviate model degradation caused by domain bias and confirmation bias.

4.3. Ablation Study

Effectiveness of denoising module. In Table 3, we ablate the denoising part, *i.e.*, SND, and compare it with alternatives. In two synthetic-to-real SFUDA experiments, SND achieves a performance improvement of nearly 6%, showing optimal competitiveness than other alternatives. In multiple and compound adaptation tasks, SND produces a performance improvement of nearly 5%. On difficult transfer tasks, SND exhibits stable cross-domain adaptability, whereas other alternatives suffer from severe performance degradation. The effectiveness of SND is mainly contributed to its accurate estimation of uncertain and comprehensive denoising capabilities.

Effectiveness of the Stability Metric. In Table 4, we compare the proposed stability with other adaptability metrics for subset division, including regional entropy [45, 59], image-level loss [74], and domain distance [60]. Denoising using stability metrics shows obvious advantages on two adaptive tasks, with gains increased by 2.7% and 3.0% compared to the second competitor. We think it contributes to the ability to screen more reliable samples to provide effective support for denoising.

Category compensation	GTA → Cityscapes	Cityscapes → ACDC
None	53.0	46.3
$\mathcal{D}_{se}$	54.2	47.1
$\mathcal{D}_{sec}$	**55.6**	**48.1**

Table 6. Variants of category compensation methods. 'None' means not applying compensation. '$\mathcal{D}_{se}$' and '$\mathcal{D}_{sec}$' means we select samples from $\mathcal{D}_{se}$ and $\mathcal{D}_{sec}$ for compensation.

Effectiveness of Query Methods. In Table 5, we verify the impact of different retrieval factors on adaptability. It shows that gradually adding domain-related factors for retrieving stable neighbors can improve adaptability. In special, on complex compound domain adaptation task, Cityscapes → ACDC, retrieving by domain factors shows greater improvements. This further verifies matching highly correlated neighbors is crucial for bi-level optimization.

Effectiveness of Category Compensation. In Table 6, we verify the impact of category compensation on adaptability. Using the stable set $\mathcal{D}_{se}$ as the copy object for compensation improves performance by 1.2% and 0.8% on two tasks respectively, verifying the impact of missing categories on denoising. Furthermore, adding category stable sets as replication objects can improve the performance by 1.4% and 1.0%, showing that diversity compensation is more conducive to reducing the bias of bi-level optimization.

4.4. Qualitative assessment

Visualization of the Learned Uncertainty Map . Fig. 5 presents the learned uncertain map ω on two SFUDA tasks. It can be seen that the uncertainty maps from probability entropy [61] and multi-classifier voting [56,82] contain a large amount of noise. Moreover, in the harder open-compound task Cityscapes → ACDC, the estimation bias of their methods is more obvious in difficult domain data, which will be a predisposing factor for error accumulation. In contrast, SND alleviates bias estimation and presents more reasonable uncertainty maps for different categories and domains. Moreover, our SND even can give accurate estimation against noise at the segmentation edges.

Denoising Effect during Training. In Fig. 6, we plot the mIoU scores on the validation during training to show

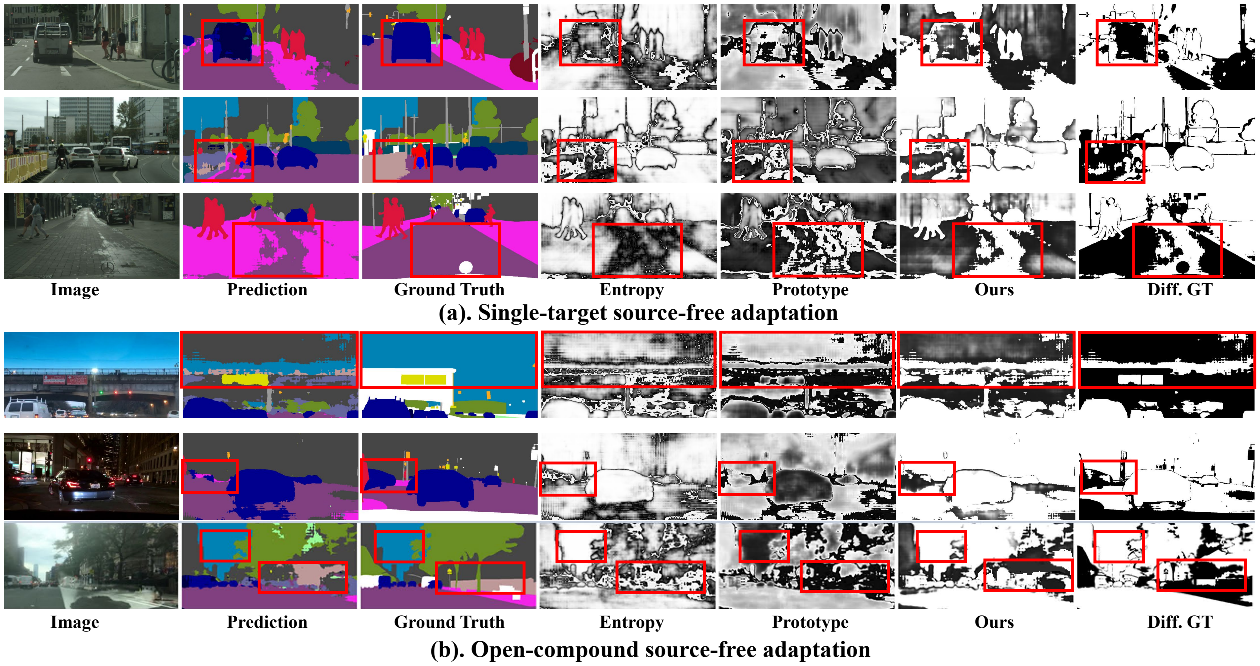

Figure 5. Visualization of different uncertainty estimation results on both GTA → Cityscapes and GTA → BDD100k tasks. Diff.GT denotes the ground truth estimation mask. Entropy map is shown by probability entropy [61]. Prototype map is shown by the difference between the Aspp classifier and prototype classifier [56, 82].

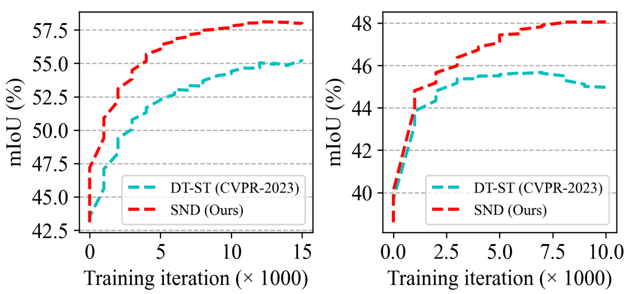

Figure 6. Comparison of the mIoU score (%) curve on the validation set during training on two transfer tasks.

the denoising effect. In single-target adaptation (a), compared with DTST [85], SND shows stronger denoising ability, allowing the model to achieve faster convergence speed during training and to obtain better performance. In multi-target adaptation (b), DTST degrades in the later stages of training, which shows that more difficult adaptation tasks place higher requirements on denoising. In contrast, SND shows a better denoising effect and can effectively alleviate model degradation.

4.5. Hyperparameter Sensitivity

In Table 7, we analyze the sensitivity of the hyper-parameter top-k% and τ in Eq. 2 on the two SFUDA tasks across three runs. For k, we set the range from 1 to 10, because too small k cannot select enough valuable samples, while too large k will inject massive noise into the stable set. On the two tasks, the fluctuation range of mIoU is within 0.5% and 0.8%, showing that SND is not sensitive to k. For τ, we set the range from $2,000$ to $12,000$ to verify the impact of the stability evaluation iteration on SND for selecting stable samples. Results show that SND has very small fluctuations in performance, within 0.3%. This indicates that SND is not sensitive to τ and also verifies the observation in Fig. 2 from the side.

k (%)	1.0	2.5	5.0	7.5	10.0
GTA → Cityscapes	57.8	57.9	**58.1**	57.8	57.8
Cityscapes → ACDC	49.9	50.2	**50.7**	50.4	49.8
τ ($\times 1000$)	2	4	8	10	12
GTA → Cityscapes	57.9	58.0	**58.1**	58.1	58.1
Cityscapes → ACDC	50.5	50.6	**50.7**	50.7	50.7

Table 7. Sensitivity study of the hyper-parameter k and τ.

5. Conclusion

In this work, we propose *Stable Neighbor Denoising* to perform unbiased denoising for the SFUDA semantic segmentation tasks. SND detects and suppresses noise in unstable samples by establishing the connection between the stable and unstable samples through bi-level optimization. The proposed retrieval nearest neighbor strategy and category compensation strategy further reduce the bias of bi-level optimization, thereby achieving effective denoising. Extensive experiments on different source-free adaptation scenarios, backbones, and ablations show that SND effectively estimates the noise of pseudo-labels and achieves state-of-the-art performance on all benchmarks.

References

[1] Nikita Araslanov and Stefan Roth. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15384–15394, 2021. 1, 2, 3

[2] Nikita Araslanov and Stefan Roth. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15384–15394, June 2021. 3

[3] Baixu Chen, Junguang Jiang, Ximei Wang, Pengfei Wan, Jianmin Wang, and Mingsheng Long. Debiased self-training for semi-supervised learning. *Advances in Neural Information Processing Systems*, 35:32424–32437, 2022. 2

[4] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018. 5

[5] Lin Chen, Zhixiang Wei, Xin Jin, Huaian Chen, Miao Zheng, Kai Chen, and Yi Jin. Deliberated domain bridging for domain adaptive semantic segmentation. *Advances in Neural Information Processing Systems*, 35:15105–15118, 2022. 1

[6] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 3

[7] Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1791–1800, 2019. 3

[8] Ziheng Chen, Yue Song, Yunmei Liu, and Nicu Sebe. A Lie group approach to Riemannian batch normalization. In *The Twelfth International Conference on Learning Representations*, 2024. 3

[9] Hao Cheng, Zhaowei Zhu, Xingyu Li, Yifei Gong, Xing Sun, and Yang Liu. Learning with instance-dependent label noise: A sample sieve approach. *arXiv preprint arXiv:2010.02347*, 2020. 3

[10] Yiting Cheng, Fangyun Wei, Jianmin Bao, Dong Chen, and Wenqiang Zhang. Adpl: Adaptive dual path learning for domain adaptation of semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2

[11] Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6830–6840, 2019. 3

[12] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11580–11590, 2021. 3

[13] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 5

[14] Ning Ding, Yixing Xu, Yehui Tang, Chao Xu, Yunhe Wang, and Dacheng Tao. Source-free domain adaptation via distribution estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7212–7222, 2022. 3

[15] Francois Fleuret et al. Uncertainty reduction for model adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9613–9623, 2021. 1, 3, 6, 7

[16] Xiaoqing Guo, Chen Yang, Baopu Li, and Yixuan Yuan. Metacorrection: Domain-aware meta loss correction for unsupervised domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3927–3936, June 2021. 3

[17] Jiangfan Han, Ping Luo, and Xiaogang Wang. Deep self-learning from noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5138–5147, 2019. 3

[18] Jianzhong He, Xu Jia, Shuaijun Chen, and Jianzhuang Liu. Multi-source domain adaptation with collaborative learning for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11008–11017, 2021. 2

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 5

[20] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018. 1, 3

[21] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9924–9935, 2022. 1, 2

[22] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Hrda: Context-aware high-resolution domain-adaptive semantic segmentation. In *European Conference on Computer Vision*, pages 372–391. Springer, 2022. 1, 2

[23] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11721–11732, 2023. 1, 2, 3

[24] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Fsdr: Frequency space domain randomization for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6891–6902, 2021. 3

[25] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, 34:3635–3649, 2021. 3, 6

[26] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407*, 2018. 4

[27] Mengmeng Jing, Xiantong Zhen, Jingjing Li, and Cees Snoek. Variational model perturbation for source-free domain adaptation. *Advances in Neural Information Processing Systems*, 35:17173–17187, 2022. 3

[28] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[29] Jogendra Nath Kundu, Akshay Kulkarni, Amit Singh, Varun Jampani, and R. Venkatesh Babu. Generalize then adapt: Source-free domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7046–7056, October 2021. 1, 3, 6, 7

[30] Jogendra Nath Kundu, Akshay Kulkarni, Amit Singh, Varun Jampani, and R Venkatesh Babu. Generalize then adapt: Source-free domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7046–7056, 2021. 3

[31] Sohyun Lee, Taeyoung Son, and Suha Kwak. Fifo: Learning fog-invariant features for foggy scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18911–18921, 2022. 2

[32] Guangrui Li, Guoliang Kang, Wu Liu, Yunchao Wei, and Yi Yang. Content-consistent matching for domain adaptive semantic segmentation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 440–456, 2020. 5

[33] Jinlong Li, Zequn Jie, Xu Wang, Xiaolin Wei, and Lin Ma. Expansion and shrinkage of localization for weakly-supervised semantic segmentation. *Advances in Neural Information Processing Systems*, 35:16037–16051, 2022. 3

[34] Jinlong Li, Zequn Jie, Xu Wang, Yu Zhou, Xiaolin Wei, and Lin Ma. Weakly supervised semantic segmentation via progressive patch learning. *IEEE Transactions on multimedia*, 2022. 3

[35] 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*, pages 11593–11603, 2022. 1

[36] Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan, and LINGYU DUAN. Uncertainty modeling for out-of-distribution generalization. In *International Conference on Learning Representations*, 2022. 5

[37] Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8602–8617, 2021. 3

[38] Yuang Liu, Wei Zhang, and Jun Wang. Source-free domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1215–1224, 2021. 1

[39] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[40] Xin Luo, Wei Chen, Zhengfa Liang, Longqi Yang, Siwei Wang, and Chen Li. Crots: Cross-domain teacher–student learning for source-free domain adaptive semantic segmentation. *International Journal of Computer Vision*, 132(1):20–39, 2024. 6

[41] Yawei Luo, L. Zheng, T. Guan, Junqing Yu, and Y. Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2502–2511, 2019. 3

[42] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. Normalized loss functions for deep learning with noisy labels. In *International conference on machine learning*, pages 6543–6553. PMLR, 2020. 3

[43] Yanbiao Ma, Licheng Jiao, Fang Liu, Yuxin Li, Shuyuan Yang, and Xu Liu. Delving into semantic scale imbalance. In *The Eleventh International Conference on Learning Representations*, 2023. 2

[44] Yanbiao Ma, Licheng Jiao, Fang Liu, Shuyuan Yang, Xu Liu, and Lingling Li. Curvature-balanced feature manifold learning for long-tailed classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15824–15835, June 2023. 3

[45] 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)*, June 2020. 7

[46] Mingjie Pan, Rongyu Zhang, Zijian Ling, Yulu Gan, Lingran Zhao, Jiaming Liu, and Shanghang Zhang. Cloud-device collaborative adaptation to continual changing environments in the real-world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12157–12166, 2023. 2

[47] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 1

[48] Hieu Pham, Zihang Dai, Qizhe Xie, and Quoc V. Le. Meta pseudo labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11557–11568, June 2021. 2, 3

[49] Nan Pu, Yu Liu, Wei Chen, Erwin M Bakker, and Michael S Lew. Meta reconciliation normalization for lifelong person re-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 541–549, 2022. 3

[50] Nan Pu, Zhun Zhong, Nicu Sebe, and Michael S. Lew. A memorizing and generalizing framework for lifelong person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):13567–13585, 2023. 1

[51] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 102–118. Springer International Publishing, 2016. 5

[52] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 5

[53] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7374–7383, 2019. 2

[54] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3139–3153, 2020. 2

[55] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. ACDC: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021. 5

[56] Fengyi Shen, Akhil Gurram, Ziyuan Liu, He Wang, and Alois Knoll. Diga: Distil to generalize and then adapt for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15866–15877, 2023. 1, 2, 3, 7, 8

[57] 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. *Advances in neural information processing systems*, 33:596–608, 2020. 4, 5

[58] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1379–1389, 2021. 2, 3, 5

[59] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 3, 7

[60] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE/CVFs Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 7

[61] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2517–2526, 2019. 3, 7, 8

[62] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020. 6, 7

[63] Haoran Wang, Tong Shen, Wei Zhang, Ling-Yu Duan, and Tao Mei. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 642–659, 2020. 3

[64] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7201–7211, 2022. 6

[65] Shuang Wang, Qi Zang, Dong Zhao, Chaowei Fang, Dou Quan, Yutong Wan, Yanhe Guo, and Licheng Jiao. Select, purify, and exchange: A multisource unsupervised domain adaptation method for building extraction. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2023. 1

[66] Shuang Wang, Dong Zhao, Chi Zhang, Yuwei Guo, Qi Zang, Yu Gu, Yi Li, and Licheng Jiao. Cluster alignment with target knowledge mining for unsupervised domain adaptation semantic segmentation. *IEEE Transactions on Image Processing*, 31:7403–7418, 2022. 3

[67] 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*, pages 322–330, 2019. 3

[68] Linshan Wu, Leyuan Fang, Xingxin He, Min He, Jiayi Ma, and Zhun Zhong. Querying labeled for unlabeled: Cross-image semantic consistency guided semi-supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8827–8844, 2023. 3, 7

[69] Xinyi Wu, Zhenyao Wu, Hao Guo, Lili Ju, and Song Wang. Dannet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15769–15778, 2021. 2

[70] Yichen Wu, Jun Shu, Qi Xie, Qian Zhao, and Deyu Meng. Learning to purify noisy labels via meta soft label corrector. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10388–10396, 2021. 2, 3

[71] Binhui Xie, Shuang Li, Mingjia Li, Chi Harold Liu, Gao Huang, and Guoren Wang. Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1

[72] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 6

[73] Feng Xue, Yicong Chang, Tianxi Wang, Yu Zhou, and Anlong Ming. Indoor obstacle discovery on reflective ground via monocular camera. *International Journal of Computer Vision (IJCV)*, pages 1573–1405, 2023. 1

[74] Jianfei Yang, Xiangyu Peng, Kai Wang, Zheng Zhu, Jiashi Feng, Lihua Xie, and Yang You. Divide to adapt: Mitigating confirmation bias for domain adaptation of black-box predictors. *arXiv preprint arXiv:2205.14467*, 2022. 3, 7

[75] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8978–8987, 2021. 1

[76] Senqiao Yang, Jiarui Wu, Jiaming Liu, Xiaoqi Li, Qizhe Zhang, Mingjie Pan, and Shanghang Zhang. Exploring sparse visual prompt for cross-domain semantic segmentation. *arXiv preprint arXiv:2303.09792*, 2023. 1

[77] 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)*, June 2020. 3, 5

[78] Li Yi, Gezheng Xu, Pengcheng Xu, Jiaqi Li, Ruizhi Pu, Charles Ling, A Ian McLeod, and Boyu Wang. When source-free domain adaptation meets learning with noisy labels. *arXiv preprint arXiv:2301.13381*, 2023. 3

[79] Yifang Yin, Wenmiao Hu, Zhenguang Liu, Guanfeng Wang, Shili Xiang, and Roger Zimmermann. Crossmatch: Source-free domain adaptive semantic segmentation via cross-modal consistency training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21786–21796, 2023. 6

[80] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 5

[81] Xiangyu Yue, Yang Zhang, Sicheng Zhao, Alberto Sangiovanni-Vincentelli, Kurt Keutzer, and Boqing Gong. Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2100–2110, 2019. 3

[82] 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/CVF Conference on Computer Vision and Pattern Recognition*, pages 12414–12424, 2021. 1, 3, 7, 8

[83] Ziyi Zhang, Weikai Chen, Hui Cheng, Zhen Li, Siyuan Li, Liang Lin, and Guanbin Li. Divide and contrast: Source-free domain adaptation via adaptive contrastive learning. *Advances in Neural Information Processing Systems*, 35:5137–5149, 2022. 3

[84] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018. 3

[85] Dong Zhao, Shuang Wang, Qi Zang, Dou Quan, Xiutiao Ye, and Licheng Jiao. Towards better stability and adaptability: Improve online self-training for model adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11733–11743, 2023. 1, 3, 6, 8

[86] Dong Zhao, Shuang Wang, Qi Zang, Dou Quan, Xiutiao Ye, Rui Yang, and Licheng Jiao. Learning pseudo-relations for cross-domain semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19191–19203, October 2023. 2

[87] Dong Zhao, Ruizhi Yang, Shuang Wang, Qi Zang, Yang Hu, Licheng Jiao, Nicu Sebe, and Zhun Zhong. Semantic connectivity-driven pseudo-labeling for cross-domain segmentation, 2023. 3

[88] Yuyang Zhao, Zhun Zhong, Na Zhao, Nicu Sebe, and Gim Hee Lee. Style-hallucinated dual consistency learning for domain generalized semantic segmentation. In *European Conference on Computer Vision*, pages 535–552. Springer, 2022. 3

[89] Zhedong Zheng and Yi Yang. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision*, 129(4):1106–1120, 2021. 1, 7

[90] Yang Zou, Zhiding Yu, BVK 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. 2