Informative Data Mining for One-shot Cross-Domain Semantic Segmentation

Yuxi Wang[1,2] Jian Liang[2,3,4] Jun Xiao[3] Shuqi Mei[5] Yuran Yang[5] Zhaoxiang Zhang[1,2,3,4,*]

[1]Centre for Artificial Intelligence and Robotics, HKISI-CAS

[2]Institute of Automation, Chinese Academy of Sciences [3]University of Chinese Academy of Sciences

[4]State Key Laboratory of Multimodal Artificial Intelligence Systems [5]Tencent

{yuxiwang93, liangjian92}@gmail.com zhaoxiang.zhang@ia.ac.cn

Abstract

Contemporary domain adaptation offers a practical solution for achieving cross-domain transfer of semantic segmentation between labelled source data and unlabeled target data. These solutions have gained significant popularity; however, they require the model to be retrained when the test environment changes. This can result in unbearable costs in certain applications due to the time-consuming training process and concerns regarding data privacy. One-shot domain adaptation methods attempt to overcome these challenges by transferring the pre-trained source model to the target domain using only one target data. Despite this, the referring style transfer module still faces issues with computation cost and over-fitting problems. To address this problem, we propose a novel framework called Informative Data Mining (IDM) that enables efficient one-shot domain adaptation for semantic segmentation. Specifically, IDM provides an uncertainty-based selection criterion to identify the most informative samples, which facilitates quick adaptation and reduces redundant training. We then perform a model adaptation method using these selected samples, which includes patch-wise mixing and prototype-based information maximization to update the model. This approach effectively enhances adaptation and mitigates the overfitting problem. In general, we provide empirical evidence of the effectiveness and efficiency of IDM. Our approach outperforms existing methods and achieves a new state-of-the-art one-shot performance of 56.7%/55.4% on the GTA5/SYNTHIA to Cityscapes adaptation tasks, respectively. The code will be released at https://github.com/yxiwang/IDM.

1. Introduction

Semantic segmentation is a fundamental computer vision task that has made remarkable progress with the help of

---

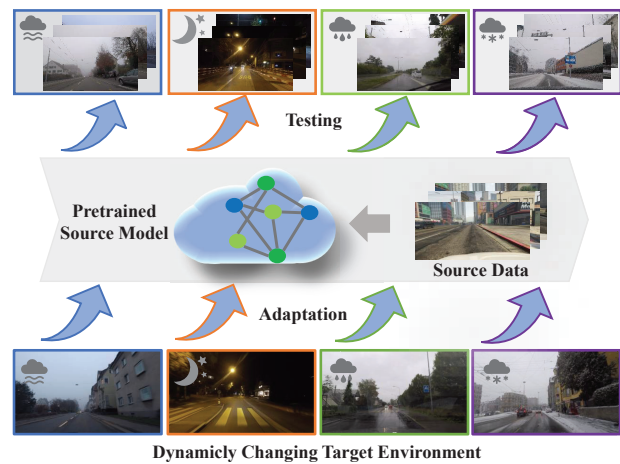*[*]Zhaoxiang Zhang is the corresponding author.

Figure 1. We consider the one-shot domain adaptation scenario, where only one single target image is used to fit the trained source model. It is realistic for the adaptation model to tackle dynamically changing target environments, such as suddenly occurrent weather (*e.g.* foggy, rainy, night, snow). Our method aims to achieve quick adaptation for one-shot domain adaptation.

vast amounts of pixel-level annotated training data. However, collecting such large-scale datasets requires tremendous labeling efforts in terms of time and cost. For instance, annotating a single image in the Cityscapes dataset can take about 90 minutes and cost 1.5 dollars [7]. To reduce this burden of labeling, previous cross-domain semantic segmentation approaches [24, 39, 26, 41, 53, 21, 56, 38] have been developed. These methods transfer the knowledge from label-rich synthetic data (source) [32, 33] to the unlabeled real-world data (target) [7], referred to as domain adaptive semantic segmentation (DASS).

Despite significant efforts in developing DASS methods, most of them still suffer from the following limitations. First, existing approaches train an adaptation model from a specific source domain to a specific target domain. Therefore, they require retraining the model every time when the

test environment changes, which is inflexible and inefficient in handling a dynamic domain shift scenario. Second, these methods need access to the entire target dataset to achieve model adaptation, which is impractical in some realistic scenarios due to privacy or storage concerns. As depicted in Figure 1, an autonomous driving model often faces sudden weather or illumination changes, resulting in a scarcity of images at the beginning of environmental shifts. Therefore, it is crucial for the model to rapidly adapt to the new conditions with a limited amount of target data. Although previous works [25, 49] have attempted to address the one-shot domain adaptation problem, the introduced style transfer module usually requires additional style images by an extra optimizing model. This makes it inefficient to adapt quickly to different scenarios.

In contrast to pioneering works, our method provides a new direction for OSDA by exploring the abundant information hidden in the source domain due to the rare target data accessible. To achieve this goal, two key ideas exist to address this problem. First, we select the most informative samples for training, which can reflect the target distribution and reduce the redundant back-propagation process. Second, we diversify the target distribution to alleviate the over-fitting problem caused by the limited availability of one-shot target data. To this end, we propose the Informative Data Mining (**IDM**) approach, which includes a sample selection strategy and an efficient model adaptation technique. Specifically, we first introduce an uncertainty-based selection criterion to identify the most informative training samples from the source data. These samples can reflect the target distribution and contribute significantly to model adaptation. Therefore, the sample selection criteria aim to filter the most informative images with 1) higher prediction uncertainty values and 2) higher diversity. Unlike existing works choosing low-uncertainty target samples to refine the target model, our method filters target-style-like source images for training. On the other hand, we devise a model adaptation method seeking to alleviate the over-fitting problem by diversifying the distribution of the target domain. Concretely, we first diversify the semantic content of target data by a patch-wise mixing method and then use prototype-based information maximization to ensure the output of diversity for the selected source data, which can significantly alleviate the over-fitting problem. With the proposed IDM method, the model can be efficiently adapted to the target data without over-fitting.

Our contributions can be summarized as follows: 1) We propose a new efficient one-shot adaptation framework for cross-domain semantic segmentation, which aims at quickly adapting the trained source model to the target data with only hundreds of training iterations. 2) We propose a novel sample selection scheme to filter out the most informative training samples for reducing redundant optimiza-

tion and devise an uncertainty minimization training technique for model adaptation. 3) We show the efficacy and efficiency of our method by achieving a new state-of-the-art performance on one-shot domain adaptive semantic segmentation, with 56.7% mIoU on GTA5 to Cityscapes and 55.4% mIoU on SYNTHIA to Cityscapes.

2. Related work

Unsupervised Domain Adaptive Semantic Segmentation aims to transfer the pixel-level annotations from the source domain to the target domain. Existing approaches can be roughly categorized into two groups: adversarial learning based method and self-training based method. For adversarial learning, numerous works focus on reducing the distribution misalignment in the image level [4, 5, 10, 34, 50, 53, 18], feature level [1, 11, 3, 8, 13], and output level [39, 26, 41, 40, 57]. For the self-training based method, the essential idea is to generate reliable pseudo labels. Typical approaches usually consist of two steps: 1) generate pseudo labels based on the source model [60, 59, 47, 45, 46, 48, 36] or the learnt domain-invariant model [58, 22, 35, 55], 2) refine the target model supervised by the generated pseudo labels. PLCA [17, 44] introduces a new paradigm for domain adaptive semantic segmentation via building pixel-level cycle association between source and target pixel pairs. Although these methods have achieved promising results, they usually access amounts of source and target data. In this paper, we address a challenging and practical target data-scarce setting where only a one-shot unlabeled target image is available during adaptation.

One-shot Domain Adaptation (OSDA) aims to overcome the need for larger training sets and improve the capability of transferring the trained source model to a new target domain with having access to the source data and only one target data. Recently, OSDA has achieved significant progress in dealing with face generalization [52], object detection [42], and semantic segmentation [25, 49]. For semantic segmentation, ASM [25] proposes an adversarial style mining algorithm by mutually optimizing the style-transfer module and the segmentation network via an adversarial regime. [49] integrates a style-mixing technique into the segmentor to stylize the source images without introducing any learned parameters. S4T [31, 23] is another representative method to design a regularized self-learning signal at test-time, which proposes a selective self-training scheme for semantic segmentation by regularizing pseudo labels with aligned predictive view generation. On the contrary, we aim to mine the most informative images to achieve quick adaptation and reduce the over-fitting problem.

Domain Generalized Semantic Segmentation (DG) has attracted considerable attention in recent years, which aims to learn a generalized model on the source domain and performs well on a novel domain. To improve performance
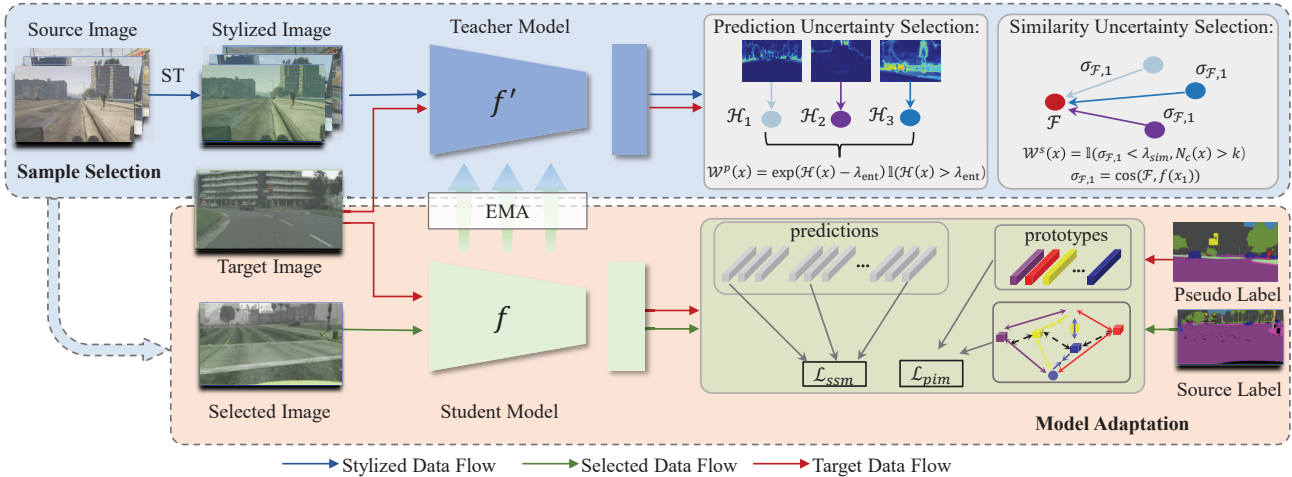
Figure 2. **An overview of our proposed IDM.** To achieve one-shot domain adaptation, our method consists of two strategies. 1) Sample selection aims to identify the most informative images to optimize the model. It first generates target-like stylized images by the style transfer (ST) technique. Then, the stylized images are fed into the teacher model to select the trained data by the proposed prediction and similarity uncertainty selection techniques; 2) The model adaptation aims to update the segmentation model on the selected and given target data. In this process, we update the model by the proposed $\mathcal{L}_{pim}$ and $\mathcal{L}_{ssm}$ to alleviate over-fitting problem. The teacher model is initialized by the pretrained source model, and it is updated in the EMA manner.

on novel domains, most existing studies focus on whitening [6], normalizing [29], and diversifying [30, 14, 54] styles to avoid over-fitting to the style of the source domain. Domain generalized semantic segmentation is close to our work, which assumes target data is inaccessible during training. The difference is that no target data is available in DG, while one target image is accessible in our work.

3. Method

In this section, we formally introduce the proposed Informative Data Mining (**IDM**) method that aims to improve the efficiency of one-shot domain adaptive semantic segmentation. In this setting, we only access one unlabeled target image $x_t \in \mathcal{X}_t$, and n_s source data $\{x_s^i, y_s^i\}_{i=1}^{n_s} = \{\mathcal{X}_s, \mathcal{Y}_s\}$. As shown in Figure 2, the proposed **IDM** consists of two strategies, 1) sample selection step and model adaptation step. The former attempts to identify the most informative images for training, and the latter aims to achieve quick adaptation without over-fitting. We assume the informative data should have two properties: contributing more to adaptation and reducing redundant optimization. Therefore, we propose prediction and similarity uncertainty selection techniques to filter the most informative training samples. After that, model adaptation seeks to alleviate the over-fitting problem by diversifying the distribution of the target domain, following the direction of the given target data.

3.1. Sample Selection

To achieve one-shot domain adaptation quickly, detecting the most informative samples that can reflect the target distribution for backward propagation is essential. Therefore, we propose a sample selection strategy considering the following two criteria: 1) samples should be *target-style closer*, and 2) the distributions of involved samples should be *diverse*, including various distributions and categories.

Prediction Uncertainty Selection. Different from previous works selecting target samples with lower uncertainty to refine the model [27], our method aims to filter the target-like source images for adaptation. Although images with lower uncertainty reveal reliable predictions, they often refer to source-like images rather than target-like ones. Therefore, we filter uncertain samples and assign them higher training weights as they contribute more to adaptation. The motivation lies in the following two aspects. (a) *The initialized model is trained on the source data.* (b) *Fine-tuning on underperforming samples (higher uncertainty) is more valuable.* Due to (a), target-like samples generally perform higher prediction uncertainty than source-like ones, which can alleviate the confusion of higher uncertainty only belonging to "hard samples". As (b), we should select higher uncertainty samples that have ground truth labels (source labels). Besides, target-like samples have high-uncertainty predictions on the source model, but not all images with high-entropy predictions are similar to the target domain. To alleviate this confusion, we apply the style transfer tech-

nique [15, 53] to generate target-style images by treating the given target image as an "anchor style". Then the prediction uncertainty selection is performed as:

$$\mathcal{W}^p(\hat{x}_s) = \exp(\mathcal{H}(\hat{x}_s) - \lambda_{ent}) \cdot \mathbb{I}(\mathcal{H}(\hat{x}_s) > \lambda_{ent}), \qquad (1)$$

where $\mathbb{I}(\cdot)$ is an indicator function, λ_{ent} denotes a predefined threshold, and $\mathcal{H}(\hat{x}_s)$ is the mean entropy of stylized source image $\hat{x}_s$. For $\hat{x}_s$ generation, we transfer the style of the target image x_t to the source data following [15, 20]:

$$\hat{x}_s = \beta(x_t)(\frac{x_s - \mu(x_s)}{\sigma(x_s)}) + \gamma(x_t). \qquad (2)$$

$\mu(x_s)$ and $\sigma(x_s)$ are the mean and standard deviation of source images. $\beta(x_t)$ and $\gamma(x_t)$ are the reconstructed target statistic of mean and standard deviation, formulating as:

$$\begin{aligned} \gamma(x_t) &= \mu(x_t) + \delta_\mu \|\mu(x_t) - \mu(x_s))\|, \\ \beta(x_t) &= \sigma(x_t) + \delta_\sigma \|(\sigma(x_t) - \sigma(x_s))\|, \end{aligned} \qquad (3)$$

where δ_μ and δ_σ control the weights of statistic offset and they are randomly sampled from Gaussian $\mathcal{G}(0, 1)$.

Similarity Uncertainty Selection. Although Eq. (1) identifies the most informative samples with higher prediction uncertainty, the restriction may still be limited. For example, selected images may only have some frequent classes, harming rare classes' performance. Moreover, the two selected samples have similar representations, with both performing a higher prediction entropy than λ_{ent}. Therefore, it is redundant for optimization as they produce an equal contribution to back-propagation.

To address this problem, we exploit the samples with diverse representations and categories in this section. Specifically, we first ensure the involved images contain different categories and then guarantee the output of selected images is not similar. Since calculating the similarity between the current image and all filtered samples is time-consuming and computationally expensive, inspired by [27], we conduct a memory bank to store the average outputs of selected samples, denoted as $\mathcal{F}$. Then the similarity uncertainty selection is formulated as follows:

$$\mathcal{W}^s(\hat{x}_s) = \mathbb{I}(cos(f(\hat{x}_s), \mathcal{F}) < \lambda_{sim}, N_c(\hat{x}_s) > k), \qquad (4)$$

where $cos(\cdot, \cdot)$ is the cosine similarity operator, f is the segmentation network, λ_{sim}, k are pre-defined thresholds, and $N_c(\hat{x}_s)$ indicates the number of classes that $\hat{x}_s$ contains.

Based on *prediction uncertainty selection* $\mathcal{W}^p(\hat{x}_s)$ and *similarity uncertainty selection* $\mathcal{W}^s(\hat{x}_s)$, we can obtain an overall sample-selection weight as:

$$\mathcal{W}(\hat{x}_s) = \mathcal{W}^p(\hat{x}_s) \cdot \mathcal{W}^s(\hat{x}_s). \qquad (5)$$

The selected samples are used to optimize the segmentation model by minimizing the following objective:

$$\mathcal{L}_{ssm}(\hat{x}_s) = -\mathcal{W}(\hat{x}_s) \sum_{i=1}^{N} \sum_{j=1}^{H \times W} y_s^{(i,j)} \log[p(\hat{x}_s^{(i,j)})], \qquad (6)$$

where N is the number of selected images to fine-tune the adaptation model. H and W denote the height and width of
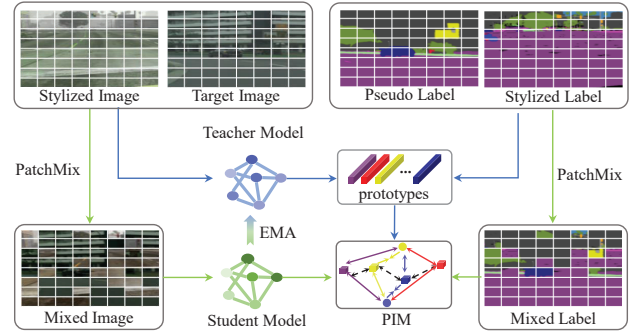

Figure 3. Illustration of the model adaptation process. PIM indicates the module of prototype-based information maximization.

the image. $p(\hat{x}_s^{(i,j)})$ indicates the probabilistic output of the j-th pixel for the i-th image $\hat{x}_s$. $\{\hat{x}_s, y_s\}$ is the training pair with a stylized source image and ground truth label.

Benefiting from the sample selection, we can select the most informative samples for training and reduce redundant optimization, helping to achieve quick adaptation for OSDA. Note that the sample-selected process does not involve any gradient back-propagation. Therefore, it is impressive to save computing resources.

3.2. Model Adaptation

Considering one-shot target images available, it only provides biased style and content information for the target domain. Despite previous works [25, 49] using style transfer to estimate the target distribution, the generated images still have a bias to the real dataset. Besides, the sole target cannot correctly reflect the target distribution, and it is easy to overfit existing categories. To alleviate this problem, we propose a novel data augmentation technique to diversify the target data. Specifically, we devise a patch-wise mixing between the selected stylized image and the given target data, which seeks to explore the content and style diversity for the training images. Then we use these mixed images to conduct a prototype-based information maximization to ensure the diversity of predictions.

Patch-wise Mixing. Class mixing [28, 37] is an excellent technique to improve adaptation performance. However, in the one-shot scenario, it is inapplicable because most categories are missing in the target data. Therefore, we introduce a patch-wise mixing method by splitting the target and stylized source images into P patches. We obtain a new mixed image by randomly replacing the source patch with the target. The corresponding labels are obtained in the same way:

$$\tilde{x}_t = \textbf{PatchMix}(\hat{x}_s, x_t), \tilde{y}_t = \textbf{PatchMix}(y_s, y'_t), \qquad (7)$$

where $\textbf{PatchMix}(\cdot, \cdot)$ is the mixing operation as shown in Figure 3. y'_t is the pseudo label of the given target image x_t.

Note that the target pseudo label is generated from a Mean-Teacher framework, where the teacher model is the exponential moving average of the student. Since our method attaches source patches to the target image, it can effectively diversify the content of training images.

Prototype-based Information Maximization. Since the original semantic structure information has been destroyed in the mixed image pair $(\widetilde{x}_t, \widetilde{y}_t)$, we adopt a metric learning method to enhance the feature representation [51, 16, 55]. To be specific, we use supervised contrastive learning to explore the semantic consistency between intra-class and inter-class. The proposed prototype-based supervised contrastive loss is as follows:

$$\mathcal{L}_{scl}(\widetilde{x}_t) = -\sum_{c=1}^{C}\sum_{i=1}^{H \times W} \widetilde{y}_t \log \frac{\exp(\mathrm{p}_c \cdot F_{\widetilde{x}_t}^{(c,i)}/\tau)}{\sum_{c=1}^{C}\exp(\mathrm{p}_c \cdot F_{\widetilde{x}_t}^{(c,i)}/\tau)}, \quad (8)$$

where τ is the temperature. $F_{\widetilde{x}_t}^{(c,i)}$ is the representation feature of mixed image $\widetilde{x}_t$ in pixel i belonging to the category c. p_c is the prototype of category c, and it is calculated on the stylized image $(\hat{x}_s, y_s)$:

$$\mathrm{p}_c = \frac{\sum_{n=1}^{N}\sum_{i=1}^{H \times W} F_{\hat{x}_s}^{(n,i)} \mathbb{I}[y_s^{(n,i)} = c]}{\sum_{n=1}^{N}\sum_{i=1}^{H \times W} \mathbb{I}[y_s^{(n,i)} = c]}. \quad (9)$$

Note that our prototype is computed based on the feature $F_{\hat{x}_s}^{(n,i)}$ of stylized images, identified by Sec. (3.1), due to the label y_s being the ground truth, which can remove the harmful label noise. To diversify the output of the adaptation model, we also provide an information maximization loss that is formulated based on the prototypes. Details are as follows.

$$\mathcal{L}_{im}(\widetilde{x}_t) = \sum_{c=1}^{C} \hat{\mathrm{p}}_c \log \mathrm{p}_c^{\widetilde{x}_t}, \quad (10)$$

where $\hat{\mathrm{p}}_c$ is the mean prototype embedding of the whole selected source image, and $\mathrm{p}_c^{\widetilde{x}_t}$ is the prototype of the mixed target image. Then we maximize the following objective for the mixed target data:

$$\mathcal{L}_{pim}(\widetilde{x}_t) = \mathcal{L}_{im}(\widetilde{x}_t) - \mathcal{L}_{scl}(\widetilde{x}_t). \quad (11)$$

Finally, we achieve efficient one-shot domain adaptation by jointly training the sample-selected minimization and the prototype-based information maximization as follows:

$$\mathcal{L}(\hat{x}_s, \widetilde{x}_t) = \mathcal{L}_{ssm}(\hat{x}_s) + \mathcal{L}_{pim}(\widetilde{x}_t). \quad (12)$$

4. Experiments

4.1. Datasets

Cityscapes, treated as target data, is a real-world dataset collected from several German cities. It has 2,975 training images with a resolution of 2048×1024. In our experiments, we use only one unlabeled image during training. We use the full validation set with 500 images to test our network. **GTA5**, consisting of 24,966 images, is collected from the homonymous computer game, and the original image size is 1914×1052. It has 19 common categories with Cityscapes, and the ground truth is generated by the game render itself. **SYNTHIA** is another synthetic dataset that contains 9,400 fully annotated images with the original resolution of 1280×760. We only evaluate a subset of 13 and 16 classes common with Cityscapes.

4.2. Implementation Details

We conduct all experiments by using **PyTorch** trained on the **GeForce RTX 3090Ti** GPU. Following previous works [12], we adopt the transformer-based architecture as a strong baseline. For a fair comparison, we also perform all experiments on the DeepLab-v2 with ResNet101 as the backbone. We train the model with an AdamW optimizer, a learning rate of 6×10^{-5} for the encoder and 6×10^{-4} for the decoder, a weight decay of 0.01, linear learning rate warmup with 500 iterations and linear decay afterwards. Similar to [12], rare class sampling is also applied. We first train the network with a batch of two 640×640 random crops for total 40k iterations to obtain a high-quality source model. Then it is considered the initialization model. During one-shot adaptation, we follow previous works [25, 49] using only one target image to achieve quick adaptation. Concerning the hyper-parameters, we utilize $\lambda_{ent} = 0.015$, $\lambda_{sim} = 0.5$, $k = 13$, and $\tau = 100$ for all experiments. Following [25], we run our methods 5 times with different random seeds to get an average result, where each time we randomly select one target image for training.

4.3. Comparison with State-of-the-art Methods

To testify the effectiveness of our method, we compare the proposed IDM with two different cross-domain semantic segmentation scenarios, including conventional unsupervised domain adaptation (UDA) and one-shot unsupervised domain adaptation (One-shot UDA). Besides, we verify the efficiency and inference-time performance of the proposed IDM.

Unsupervised Domain Adaptation (UDA). We first compare our method with the conventional UDA approaches in that all target data is available during training. We compare representative state-of-the-art approaches using ResNet-101 [9] backbone, *e.g.* CBST [60], DACS [37], UPLR [47], ProDA [55], and CPSL [19], and Transformer-based architecture, *e.g.* DAFormer [12]. The detailed results are shown in Table 1 for GTA5 to Cityscapes and Table 2 for SYNTHIA to Cityscapes. From the results, we can observe that the proposed IDM outperforms existing traditional unsupervised domain adaptation methods. Specifically, it achieves the performance of 69.5%/67.9% mIoU compared to DAFormer 68.3%/67.4% mIoU on GTA5/SYNTHIA to Cityscapes, respectively. We also provide the comparison results based on DeepLab-v2 using ResNet-101 as the back-

Table 1. Adaptation from GTA5 to Cityscapes. # TS denotes the number of target samples used in training. The best results of one-shot domain adaptation are presented in **bold**.

Method	#TS	road	side.	build.	wall	fence	pole	light	sign	vege.	terr.	sky	person	rider	car	truck	bus	train	motor.	bike	**mIoU**
UDA																					
CBST [60]	All	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	16.0	25.9	42.8	45.9
DACS [37]	All	89.9	39.7	87.9	30.7	39.5	38.5	46.4	52.8	88.0	44.0	88.8	67.2	35.8	84.5	45.7	50.2	0.0	27.3	34.0	52.1
UPLR [47]	All	90.5	38.7	86.5	41.1	32.9	40.5	48.2	42.1	86.5	36.8	84.2	64.5	38.1	87.2	34.8	50.4	0.2	41.8	54.6	52.6
ProDA [55]	All	87.8	56.0	79.7	46.3	44.8	45.6	53.5	53.5	88.6	45.2	82.1	70.7	39.2	88.8	45.5	59.4	1.0	48.9	56.4	57.5
CPSL [19]	All	92.3	59.9	84.9	45.7	29.7	52.8	**61.5**	59.5	87.9	41.5	85.0	73.0	35.5	90.4	48.7	73.9	26.3	53.8	53.9	60.8
DAFormer [12]	All	95.7	70.2	89.4	**53.5**	48.1	49.6	55.8	59.4	**89.9**	47.9	**92.5**	72.2	44.7	92.3	74.5	**78.2**	65.1	55.9	61.8	68.3
IDM (Ours)	All	**97.2**	**77.1**	**89.8**	51.7	**51.7**	54.5	59.7	**64.7**	89.2	45.3	90.5	**74.2**	46.6	92.3	76.9	59.6	**81.2**	57.3	62.4	69.5
One-shot UDA																					
CBST [60]	One	76.1	22.2	73.5	13.8	18.8	19.1	20.7	18.6	79.5	41.3	74.8	57.4	19.9	78.7	21.3	28.5	0.0	28.0	13.2	37.1
ProDA [55]	One	80.9	32.2	68.9	24.7	21.0	24.6	29.6	14.8	71.7	28.6	66.4	55.8	17.5	81.6	21.2	24.2	20.0	25.0	13.9	38.0
ASM[25]	One	**89.5**	31.2	81.3	27.8	22.8	30.6	32.8	25.1	82.6	**35.0**	76.7	59.2	26.6	82.3	27.7	34.1	0.9	25.6	29.6	43.2
SM-PPM [49]	One	85.0	23.2	80.4	21.3	24.5	30.0	32.0	26.7	83.2	34.8	74.0	57.3	29.0	77.7	27.3	36.5	5.0	28.2	39.4	42.8
DAFormer [12]	One	88.7	**34.4**	84.9	29.1	28.5	36.9	43.9	29.7	83.4	29.6	84.1	66.0	38.0	86.8	54.9	47.3	32.8	24.6	37.8	50.6
IDM (Ours)	One	88.5	30.0	**86.7**	**35.0**	**33.6**	**45.0**	**49.9**	**50.7**	**86.9**	32.8	**86.1**	**68.1**	**40.0**	**89.1**	**66.4**	**50.6**	**45.6**	**39.3**	**52.1**	**56.7**

Table 2. Adaptation from SYNTHIA to Cityscapes. # TS denotes the number of target samples used in training. The best results of one-shot domain adaptation are presented in **bold**.

Method	#TS	road	side.	build.	wall*	fence*	pole*	light	sign	vege.	sky	person	rider	car	bus	motor.	bike	**mIoU***	**mIoU**
UDA																			
CBST [60]	All	68.0	29.9	76.3	10.8	1.4	33.9	22.8	29.5	77.6	78.3	60.6	28.3	81.6	23.5	18.8	39.8	42.6	48.9
DACS [37]	All	80.6	25.1	81.9	21.5	2.9	37.2	22.7	24.0	83.7	**90.8**	67.6	38.3	82.9	38.9	28.5	47.6	48.3	54.8
UPLR [47]	All	79.4	34.6	83.5	19.3	2.8	35.3	32.1	26.9	78.8	79.6	66.6	30.3	86.1	36.6	19.5	56.9	48.0	54.6
ProDA [55]	All	**87.8**	45.7	84.6	37.1	0.6	44.0	54.6	37.0	**88.1**	84.4	74.2	24.3	88.2	51.1	40.5	45.6	55.5	62.0
CPSL [19]	All	87.2	43.9	85.5	33.6	0.3	47.7	57.4	37.2	87.8	88.5	**79.0**	32.0	**90.6**	49.4	50.8	59.8	**65.3**	57.9
DAFormer [12]	All	84.5	40.7	**88.4**	**41.5**	**6.5**	50.0	55.0	54.6	86.0	89.8	73.2	48.2	87.2	53.2	53.9	61.7	60.9	67.4
IDM (Ours)	All	87.6	**47.6**	88.1	33.4	6.3	**52.8**	**57.8**	**56.5**	83.0	77.5	66.2	**52.1**	89.3	**55.6**	**57.1**	**64.2**	60.9	**67.9**
One-shot UDA																			
CBST [60]	One	59.6	24.1	72.9	-	-	-	5.5	13.8	72.2	69.8	55.3	21.1	57.1	17.4	13.8	18.5	-	38.5
ProDA [55]	One	81.8	38.9	60.6	7.8	0	31.6	14.6	11.5	51.5	69.9	56.2	16.4	79.2	24.4	5.9	32.3	36.4	41.8
ASM [25]	One	**85.7**	39.7	77.1	1.1	0.0	24.2	2.1	9.2	76.9	81.7	43.4	11.4	63.9	15.8	1.6	20.3	34.6	40.7
SM-PPM [49]	One	79.3	35.3	75.9	5.6	**16.6**	29.8	25.4	22.7	79.9	76.8	54.6	23.5	60.2	23.9	21.2	**36.6**	41.4	47.3
DAFormer [12]	One	65.3	26.1	79.5	**24.8**	1.9	38.3	30.7	23.8	81.4	**84.0**	**66.1**	27.6	70.8	**39.3**	23.7	33.8	44.8	50.2
IDM (Ours)	One	85.4	39.4	**83.5**	11.6	0.6	**43.9**	**45.4**	**31.7**	**86.0**	83.9	62.3	23.3	**87.4**	32.4	**25.1**	34.1	**48.5**	**55.4**

bone. More results are shown in *Supplementary Materials.*
One-shot Unsupervised Domain Adaptation (One-shot UDA). To further testify the potential of IDM on one-shot domain adaptation, we compare existing methods from two aspects, 1) conventional UDA methods under one-shot scenario (*e.g.*, CBST [60] and ProDA [55], DAFormer [12]), and 2) state-of-the-art one-shot UDA methods (*e.g.*, ASM

[25], SM-PPM [49]). From the results in Table 1 and 2, we can draw some conclusions as follows. First, the proposed IDM outperforms the above methods under the one-shot UDA setting. Specifically, our method provides a significant improvement with 6.1% higher than DAFormer on GTA5 → Cityscapes. Considering SYNTHIA→Cityscapes task in Table 2, our method achieves 48.5% and 55.4% re-
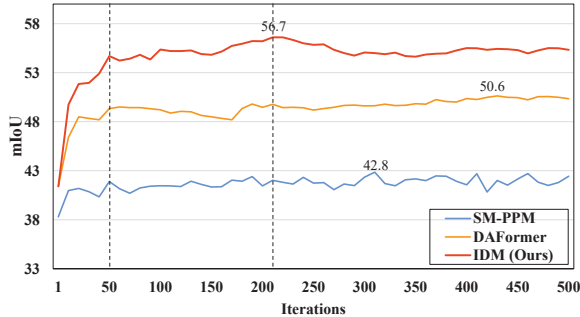
Figure 4. The convergence during training iterations.

Figure 5. Comparison results on the inference time.

sults on 13 and 16 classes, respectively, which provides a significant margin improvement over ASM, SM-PPM and DAFormer. These results demonstrate the effectiveness of the proposed informative data mining adaptation approach. Second, compared with conventional domain adaptation, the performance of one-shot UDA methods degrades significantly in such data-scarce scenarios. For example, the performance of one-shot DAFormer reduces to 50.6% mIoU on GTA5 $\rightarrow$ Cityscapes from the original 68.3% mIoU. It shows the naive model by directly reducing target data of UDA methods is infeasible for one-shot UDA due to the over-fitting to the single target image. It also reveals the essential of our method for one-shot domain adaptation. Moreover, it should notice that we have achieved an impressive performance with only 500 training iterations, which further demonstrates our efficiency.

4.4. Inference-time Adaptation Performance

To testify the efficiency of our IDM for tackling dynamically changing environment scenarios, we perform the inference-time performance comparisons in this subsection, including model convergence speed and the performance on inference-time adaptation.

Training Convergence. We first compare the convergence speed during adaptation training, which is essential for one-shot domain adaptation to adapt fast to different real-world scenarios. As ASM [25] requires 200k training iterations and an additional style transfer module optimization, we compare it with a more efficient method SM-PPM [49]. As Figure 4 shows, the proposed IDM can quickly fit into the new target domain. We observe that IDM achieves a higher and more stable adaptation result by only 50 training iterations, with a significant margin gain and quick speed for convergence. Moreover, we also provide training convergence of one-shot DAFormer. Our method outperforms DAFormer both in accuracy and speed. It reveals that our IDM is superior in both efficiency and effectiveness.

Inference-time Adaptation Results. We provide segmentation results for the referred one-shot training target data.
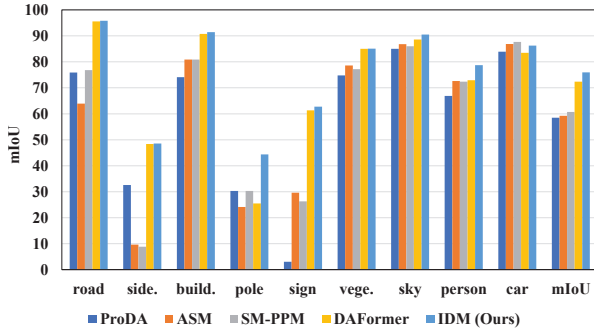
Table 3. Study on each component adopted by our IDM. SSM: sample-selected minimization, PIM: prototype-based information maximization.

Network	SSM	PatchMix	ClassMix	PIM	mIoU
1. DAFormer [12]	-	-	-	-	50.6
2. DAFormer [12]	✓	-	-	-	52.0
3. DAFormer [12]	✓	-	-	✓	53.2
4. DAFormer [12]	✓	✓	-	-	55.0
5. DAFormer [12]	-	✓	-	✓	54.8
6. DAFormer [12]	✓	-	✓	✓	55.9
7. DAFormer [12]	✓	✓	-	✓	**56.7**
8. DLv2 [2]	-	-	-	-	41.1
9. DLv2 [2]	✓	✓	-	✓	45.2

Because the target image is unlabeled, it is reasonable to evaluate this single image during inference time, as same as test time training [43], which reveals the capability of our model to fit different scenarios. We report the average results on five randomly selected target images, and the compared approaches are evaluated in the same manner. We compare with ProDA [55], ASM [25], SM-PPM [49], and DAFormer [12] in this subsection. We provide the performance of common classes among different images and the results are shown in Figure 5. We observe that the proposed IDM achieves the best results in most categories. Note that our method is computed based on the model trained for 50 iterations, while ProDA and ASM are trained for 200k iterations. It demonstrates that our IDM achieves quick adaptation for dealing with dynamic domain shift problems.

4.5. Ablation Study

Influence of Different Components. In this section, we first conduct experiments to verify the effectiveness of the proposed components, including *sample-selected minimization* (SSM), and *prototype-based information maximization* (PIM). Specifically, we also compare the influence of the proposed *PatchMix* with existing ClassMix [28] in the experiments. For a fair comparison, we provide the results of two different architectures, DAFormer [12] and DeepLab-v2 [2]. As Table 3 shows, our model provides a signifi-

Table 4. Performance on different architectures for GTA5 and SYNTHIA (SYN) to Cityscapes (CS) adaptation. Experiments are conducted on both conventional domain adaptation (UDA) and one-shot domain adaptation (OSDA).

| | GTA5 → CS | | SYN → CS | |
	OSDA	UDA	OSDA	UDA
DLv2$_{Src. Only}$	36.9	36.9	38.6	38.6
DLv2$_{IDM}$	45.2	57.3	47.2	65.9
DAFormer$_{Src. Only}$	44.5	44.5	51.3	51.3
DAFormer$_{IDM}$	56.7	69.5	55.4	67.9

cant improvement by achieving 56.7% mIoU and 45.2 % mIoU based on the DAFormer and DeepLab-v2, respectively, compared to the baseline model of 50.6% and 41.1 %. In addition, the model with SSM (model (1)) improves performance from 50.6% to 52.0%, which indicates the proposed sample selection technique can effectively transfer the target information into the trained model. Besides, the model with SSM and PIM can bring 1.2% performance gain, which indicates the proposed two parts are complementary for model adaptation. In addition, adding Patch-Mix from SSM provides a significant improvement from 52.0% (model (1)) to 55.0% (model (4)). The reason is that the proposed SSM focuses on the style of information transfer and the PIM pays more attention to context information. Moreover, comparing the PatchMix and ClassMix, model (6) and model (7), our method PatchMix outperforms 0.8% mIoU than ClassMix, owing to the rare classes' existence in the target images. Furthermore, Our method also provides a large margin improvement based on DeepLab-v2 architecture from 41.1% to 45.2%.

Performance on Different Architectures. To verify the generalization of our method, we perform the experiment on different architectures, including the traditional convolution-based architecture DeepLab-v2 [2] and advanced transformer-based architecture DAFormer [12]. Specifically, we provide the conventional domain adaptation (UDA) and one-shot domain adaptation (OSDA) results on Table 4. From the results, we can observe that the proposed method (IDM) offers a significant improvement on both GTA5 to Cityscapes (GTA5→CS) and SYNTHIA to Cityscapes (SYN→CS) adaptations. The detailed results are attached in the *Supplementary Materials*.

Influence of Different Selection Strategies. To identify the most informative samples, we have proposed two different selection strategies: prediction uncertainty selection ($\mathcal{W}^p$) and similarity uncertainty selection ($\mathcal{W}^s$). As our sample section is based on the style transferred (ST) images, we ablate both of them in this subsection. We conduct experiments on GTA5 → Cityscapes to verify the effectiveness of different strategies. The detailed results are shown in Table 5. Compared with the baseline, introducing style transfer brings performance improvement from 50.6%

Table 5. Performance on different sample selection strategies for GTA5 to Cityscapes adaptation.

Method	ST	$\mathcal{W}^p$	$\mathcal{W}^s$	mIoU	Iterations
Baseline	-	-	-	50.6	20000
(a)	✓	-	-	51.5	13000
(b)	-	✓	-	51.4	2000
(c)	✓	-	✓	51.7	4300
(d)	✓	✓	✓	52.0	300

Table 6. Study on the uncertainty threshold parameter λ_{ent}.

λ_{ent}	mIoU
0.005	50.8
0.010	51.1
0.015	51.4
0.020	50.9
0.025	51.2

Table 7. Study on the similarity threshold parameter λ_{sim}.

λ_{sim}	mIoU
0.8	51.0
0.7	51.6
0.6	51.7
0.5	52.0
0.4	51.4

Table 8. Study on the number k of categories contained in the image.

k	mIoU
10	50.5
11	50.7
12	51.5
13	51.6
14	51.6

to 51.5%, without reducing significant training iterations. Besides, adding prediction and similarity uncertainty selection techniques, the model achieves 52.0% mIoU with only 300 training iterations. This reveals our method is efficient to perform quick adaptation and verifies the effectiveness of the proposed informative sample selection strategy.

4.6. Parameters Analysis

Our framework contains several new hyper-parameters, including the prediction uncertainty selection threshold λ_{ent} in Eq. (1), the similarity uncertainty selection threshold λ_{sim} in Eq. (4), and the number k of categories contained in the selected stylized images in Eq. (4). We construct extensive experiments to analyze the influence of these hyper-parameters. The detailed results are provided in Table 6, 7, and 8. From the results, we can observe that although our method requires many manually-defined thresholds, the performance of IDM is stable and not sensitive to these hyper-parameters.

Moreover, we also provide the analysis of the number of patches P in the PatchMix. The key idea of PatchMix is to increase the diversity of training samples, so we randomly mix patches without specifying the corresponding positions replacement. This approach indeed somewhat breaks the semantic relations while remaining locally structured information. Fortunately, due to supervised information for training, the model can extract generalized features regardless of their spatial locations. This random mixing strategy promotes robustness and generalization in the model adaptation. The detailed results of patches P are shown in the Table 9.

Table 9. Study on the number of patches in the PatchMix module.

# P	16	36	48	64	96	144
IDM	55.8	56.0	56.1	56.5	56.7	56.3

5. Conclusion

This paper proposes an Informative Data Mining (**IDM**) framework, aiming at performing quick adaptation from the pre-trained source model to the target domain by only hundreds of training iterations with one-shot target data available. To achieve this goal, we first propose a novel sample selection criterion to identify the most informative samples for training reducing redundant training significantly. At the same time, we update the adaptation model by the proposed model adaptation method. Specifically, we use the prototype-based information maximization loss to enlarge the diversity of the training samples alleviating the overfitting problem. The sample-selected minimization loss enforces the pre-trained source model to fit the target data. The efficacy and efficiency of IDM have been demonstrated by achieving a new state-of-the-art performance on two standard one-shot domain adaptive semantic segmentation benchmarks.

6. Acknowledgement

References

[1] Wei-Lun Chang, Hui-Po Wang, Wen-Hsiao Peng, and Wei-Chen Chiu. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *Proc. CVPR*, pages 1900–1909, 2019. 2

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan 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, 2017. 7, 8

[3] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *Proc. ICCV*, pages 2090–2099, 2019. 2

[4] Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *Proc. CVPR*, pages 1791–1800, 2019. 2

[5] Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In *Proc. ICCV*, pages 6830–6840, 2019. 2

[6] 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 *Proc. CVPR*, pages 11580–11590, 2021. 3

[7] 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 *Proc. CVPR*, pages 3213–3223, 2016. 1

[8] Junsong Fan, Yuxi Wang, He Guan, Chunfeng Song, and Zhaoxiang Zhang. Toward few-shot domain adaptation with perturbation-invariant representation and transferable prototypes. *Frontiers of Computer Science*, 16(3):163347, 2022. 2

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016. 5

[10] 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 *Proc. ICML*, pages 1989–1998, 2018. 2

[11] Weixiang Hong, Zhenzhen Wang, Ming Yang, and Junsong Yuan. Conditional generative adversarial network for structured domain adaptation. In *Proc. CVPR*, pages 1335–1344, 2018. 2

[12] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proc. CVPR*, pages 9924–9935, 2022. 5, 6, 7, 8

[13] Dapeng Hu, Jian Liang, Qibin Hou, Hanshu Yan, and Yunpeng Chen. Adversarial domain adaptation with prototype-based normalized output conditioner. *IEEE Transactions on Image Processing*, 30:9359–9371, 2021. 2

[14] Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Fsdr: Frequency space domain randomization for domain generalization. In *Proc. CVPR*, pages 6891–6902, 2021. 3

[15] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proc. ICCV*, pages 1501–1510, 2017. 4

[16] Zhengkai Jiang, Yuxi Li, Ceyuan Yang, Peng Gao, Yabiao Wang, Ying Tai, and Chengjie Wang. Prototypical contrast adaptation for domain adaptive semantic segmentation. In *Proc. ECCV*, pages 36–54, 2022. 5

[17] Guoliang Kang, Yunchao Wei, Yi Yang, Yueting Zhuang, and Alexander Hauptmann. Pixel-level cycle association: A new perspective for domain adaptive semantic segmentation. In *Proc. NeurIPS*, pages 3569–3580, 2020. 2

[18] Myeongjin Kim and Hyeran Byun. Learning texture invariant representation for domain adaptation of semantic segmentation. In *Proc. CVPR*, pages 12975–12984, 2020. 2

[19] 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 *Proc. CVPR*, pages 11593–11603, 2022. 5, 6

[20] Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan, and LINGYU DUAN. Uncertainty modeling for out-of-distribution generalization. In *Proc. ICLR*, 2021. 4

[21] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *Proc. CVPR*, pages 6936–6945, 2019. 1

[22] Qing Lian, Fengmao Lv, Lixin Duan, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *Proc. ICCV*, pages 6758–6767, 2019. 2

[23] Jian Liang, Ran He, and Tieniu Tan. A comprehensive survey on test-time adaptation under distribution shifts. *arXiv preprint arXiv:2303.15361*, 2023. 2

[24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proc. CVPR*, pages 3431–3440, 2015. 1

[25] 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. 2, 4, 5, 6, 7

[26] 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 *Proc. CVPR*, pages 2507–2516, 2019. 1, 2

[27] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *Proc. ICML*, pages 16888–16905. PMLR, 2022. 3, 4

[28] Viktor Olsson, Wilhelm Tranheden, Juliano Pinto, and Lennart Svensson. Classmix: Segmentation-based data augmentation for semi-supervised learning. In *Proc. WACV*, pages 1369–1378, 2021. 4, 7

[29] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proc. ECCV*, pages 464–479, 2018. 3

[30] Duo Peng, Yinjie Lei, Lingqiao Liu, Pingping Zhang, and Jun Liu. Global and local texture randomization for synthetic-to-real semantic segmentation. *IEEE Transactions on Image Processing*, 30:6594–6608, 2021. 3

[31] Viraj Prabhu, Shivam Khare, Deeksha Kartik, and Judy Hoffman. S4t: Source-free domain adaptation for semantic segmentation via self-supervised selective self-training. *arXiv preprint arXiv:2107.10140*, 2021. 2

[32] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *Proc. ECCV*, pages 102–118, 2016. 1

[33] 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 *Proc. CVPR*, pages 3234–3243, 2016. 1

[34] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, Ser Nam Lim, and Rama Chellappa. Learning from synthetic data: Addressing domain shift for semantic segmentation. In *Proc. CVPR*, pages 3752–3761, 2018. 2

[35] M Naseer Subhani and Mohsen Ali. Learning from scale-invariant examples for domain adaptation in semantic segmentation. In *Proc. ECCV*, pages 290–306, 2020. 2

[36] Qing Tian, Heyang Sun, Shun Peng, and Tinghuai Ma. Self-adaptive label filtering learning for unsupervised domain adaptation. *Frontiers of Computer Science*, 17(1):171308, 2023. 2

[37] Wilhelm Tranheden, Viktor Olsson, Juliano Pinto, and Lennart Svensson. Dacs: Domain adaptation via cross-domain mixed sampling. In *Proc. WACV*, pages 1379–1389, 2021. 4, 5, 6

[38] Thanh-Dat Truong, Chi Nhan Duong, Ngan Le, Son Lam Phung, Chase Rainwater, and Khoa Luu. Bimal: Bijective maximum likelihood approach to domain adaptation in semantic scene segmentation. In *Proc. ICCV*, pages 8548–8557, 2021. 1

[39] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proc. CVPR*, pages 7472–7481, 2018. 1, 2

[40] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *Proc. ICCV*, pages 1456–1465, 2019. 2

[41] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proc. CVPR*, pages 2517–2526, 2019. 1, 2

[42] Zhiqiang Wan, Lusi Li, Hepeng Li, Haibo He, and Zhen Ni. One-shot unsupervised domain adaptation for object detection. In *Proc. IJCNN*, pages 1–8. IEEE, 2020. 2

[43] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *Proc. ICLR*, 2020. 7

[44] Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Liwei Wu, Yuxi Wang, and Zhaoxiang Zhang. Pulling target to source: A new perspective on domain adaptive semantic segmentation. *arXiv preprint arXiv:2305.13752*, 2023. 2

[45] Haochen Wang, Yuchao Wang, Yujun Shen, Junsong Fan, Yuxi Wang, and Zhaoxiang Zhang. Using unreliable pseudo-labels for label-efficient semantic segmentation. *arXiv preprint arXiv:2306.02314*, 2023. 2

[46] Yuxi Wang, Jian Liang, and Zhaoxiang Zhang. Source data-free cross-domain semantic segmentation: Align, teach and propagate. *arXiv preprint arXiv:2106.11653*, 2021. 2

[47] Yuxi Wang, Junran Peng, and ZhaoXiang Zhang. Uncertainty-aware pseudo label refinery for domain adaptive semantic segmentation. In *Proc. ICCV*, pages 9092–9101, 2021. 2, 5, 6

[48] Yunyun Wang, Chao Wang, Hui Xue, and Songcan Chen. Self-corrected unsupervised domain adaptation. *Frontiers of Computer Science*, 16(5):165323, 2022. 2

[49] Xinyi Wu, Zhenyao Wu, Yuhang Lu, Lili Ju, and Song Wang. Style mixing and patchwise prototypical matching for one-shot unsupervised domain adaptive semantic segmentation. *arXiv preprint arXiv:2112.04665*, 2021. 2, 4, 5, 6, 7

[50] Zuxuan Wu, Xin Wang, Joseph E Gonzalez, Tom Goldstein, and Larry S Davis. Ace: Adapting to changing environments for semantic segmentation. In *Proc. ICCV*, pages 2121–2130, 2019. 2

[51] 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. 5

[52] Chao Yang and Ser-Nam Lim. One-shot domain adaptation for face generation. In *Proc. CVPR*, pages 5921–5930, 2020. 2

[53] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proc. CVPR*, pages 4085–4095, 2020. 1, 2, 4

[54] 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 *Proc. ICCV*, pages 2100–2110, 2019. 3

[55] 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 *Proc. CVPR*, pages 12414–12424, 2021. 2, 5, 6, 7

[56] Qiming Zhang, Jing Zhang, Wei Liu, and Dacheng Tao. Category anchor-guided unsupervised domain adaptation for semantic segmentation. In *Proc. NeurIPS*, pages 433–443, 2019. 1

[57] Bo Zhao, Jiashi Feng, Xiao Wu, and Shuicheng Yan. A survey on deep learning-based fine-grained object classification and semantic segmentation. *International Journal of Automation and Computing*, 14(2):119–135, 2017. 2

[58] 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. 2

[59] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proc. ICCV*, pages 5982–5991, 2019. 2

[60] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proc. ECCV*, pages 289–305, 2018. 2, 5, 6