

D²ADA: Dynamic Density-aware Active Domain Adaptation for Semantic Segmentation

Tsung-Han Wu¹, Yi-Syuan Liou¹, Shao-Ji Yuan¹, Hsin-Ying Lee¹
Tung-I Chen¹, Kuan-Chih Huang¹, and Winston H. Hsu^{1,2}

¹National Taiwan University ²Mobile Drive Technology

Abstract. In the field of domain adaptation, a trade-off exists between the model performance and the number of target domain annotations. Active learning, maximizing model performance with few informative labeled data, comes in handy for such a scenario. In this work, we present **D²ADA**, a general active domain adaptation framework for semantic segmentation. To adapt the model to the target domain with minimum queried labels, we propose acquiring labels of the samples with high probability density in the target domain yet with low probability density in the source domain, complementary to the existing source domain labeled data. To further facilitate labeling efficiency, we design a dynamic scheduling policy to adjust the labeling budgets between domain exploration and model uncertainty over time. Extensive experiments show that our method outperforms existing active learning and domain adaptation baselines on two benchmarks, GTA5 \rightarrow Cityscapes and SYNTHIA \rightarrow Cityscapes. With less than 5% target domain annotations, our method reaches comparable results with that of full supervision.

Keywords: Active Learning, Domain Adaptation, Semantic Segmentation

1 Introduction

Semantic segmentation is vital for many intelligent systems, such as self-driving cars and robotics. Over the past ten years, supervised deep learning methods [5,6,7,8,22,41,45,54,56] have achieved great success assisted by rich labeled datasets. However, obtaining large-scale datasets with manual pixel-by-pixel annotation is still costly in terms of time and effort. Thus, several prior works utilized *domain adaptation* techniques to transfer the knowledge learned from the whole labeled source domain, such as simulated game environments, to real-world unlabeled target domain.

In the field of domain adaptation, a trade-off exists between model performance and the amount of target domain annotations. With sufficient target domain manual annotations, supervised learning method can reach high performance (e.g. 71.3 mIoU on Cityscapes shown in Tab. 1). However, without any target labeling, the performance of unsupervised domain adaptation (UDA) methods [4,10,12,20,23,24,26,37,42,43,53,55,59,60,61] are still far below that of

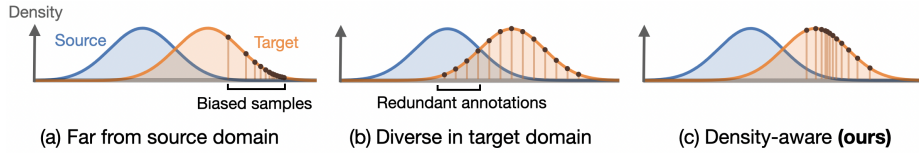


Fig. 1. Different Exploration Techniques in the ADA. (a) [40,58] proposed acquiring labels of target samples that are far from the source domain by the trained domain discriminator. However, the selected biased samples are inconsistent with real target distribution. (b) [13,31,39] proposed selecting diverse samples in the target domain with clustering techniques for label acquisition. Nonetheless, redundant annotations exist on samples that are similar to the existing labeled source domain dataset. (c) We propose a **density-aware** ADA strategy that acquires labels for samples that are representative in the target domain yet scarce the source domain, which is better than only considering either the source domain (a) or the target domain (b).

full supervision (e.g. 57.5 mIoU on GTA [32] \rightarrow Cityscapes [11] reported by [55]). Compared to the above two extreme cases, which are impractical in real-world applications, a more reasonable manner is to strike a balance between model performance and the cost of labeling efforts.

Active learning, maximizing model performance with few informative labeled data, comes in handy for such a scenario. In the past few years, several works utilized model uncertainty [14,34,36,38,44,46] or data diversity [2,21,28,35,50] as the indicator to select valuable samples for labeling. The core concept of *uncertainty-based methods* is to acquire labels for data close to the model decision boundary, and the main idea of *data diversity approaches* is to query labels for a batch of samples that are far away from each other in the feature space.

Recently, some studies leveraged uncertainty and diversity active learning methods for domain adaptation [13,31,39,40,58], named Active Domain Adaptation (ADA); however, these methods have two significant defects. First of all, existing techniques for exploring target domain distributions is not efficient. Fig. 1 (a, b) shows two typical approaches. As observed in (a), selecting samples far from source domain distributions [40,58] might lead to outliers or biased data [29]. As shown in (b), labeling diverse data in the target domain [13,31,39] might cause unnecessary annotations. In Fig. 3, we showed that the methods with 5% annotations still performed worse than ours with 2.5x fewer labels.

Furthermore, in the field of active learning, label acquisition strategies will be executed for multiple rounds [35,44]; yet, existing methods [13,31,39,40,58], combining uncertainty and diversity, did not notice the uncertainty measurement is less informative in the first few rounds in the ADA problem. As shown in Fig. 2 (a, b, c), model uncertainty might fail to detect high-confident but erroneous target domain regions (e.g. mispredict the pavement road as the sidewalk) under severe domain shift in earlier rounds. Thus, heavily relying on uncertainty cues led to poor results under low labeling budgets (see Tab. 2 and Fig. 3).

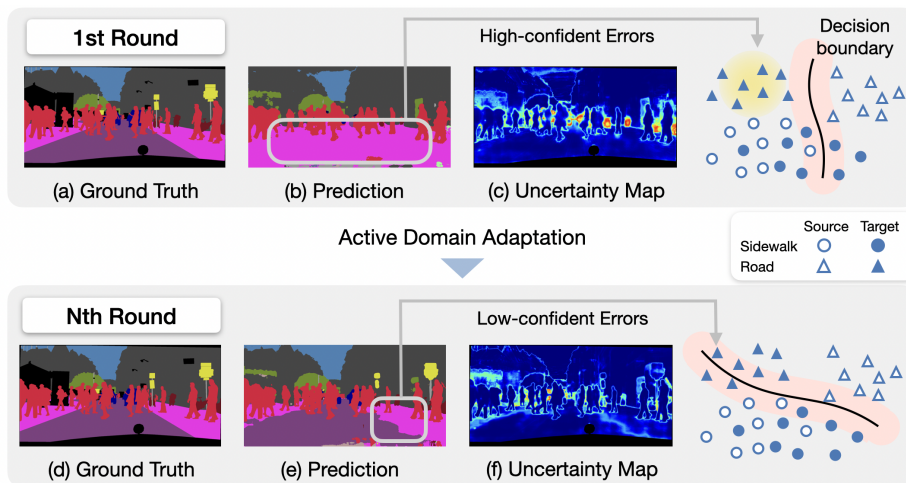


Fig. 2. Analysis of Uncertainty Criterion in the ADA. Uncertainty-based active learning methods acquire labels of data close to the decision boundary (red background). Since its inability to detect high-confident error under severe domain shift, several representative samples in the target domain far from the decision boundary will not be selected (yellow background) in earlier rounds, resulting in low label efficiency under few labeling budgets as shown in (a, b, c). However, as the two domains gradually align by fine-tuning with acquired labels, the number of high-confident but erroneous regions are rapidly reduced. Meanwhile, uncertainty measurement is able to capture the low-confident error with an accurate model in later rounds, as shown in (d, e, f). Hence, we propose a dynamic scheduling policy to pay more attention on domain exploration in earlier stages and rely more on model uncertainty later (Sec. 3.3).

To address the two problems, we present an **Dynamic Density-aware Active Domain Adaptation (D²ADA)** framework for semantic segmentation. To select the most informative target domain data for labeling, we propose a novel *density-aware selection method* to select data with the largest domain gaps. In this work, we use the term, **domain density**, to describe the prevalence of an observed sample in a specific domain. Then, we acquire the labels of regions with the largest density difference between the source and the target domain. As our intuition in Fig. 1 (c), since the model has already performed well on rich labeled source domain data, labeling few samples with high target domain density but low source domain density is sufficient for models to overcome the domain shift. Compared to prior works in Fig. 1 (a, b), the superiority of our method is demonstrated in Fig. 3. Also, it is mathematically proved to reduce the generalization bound of domain adaptation.

To maximize the advantages of domain exploration and model uncertainty, we develop a *dynamic scheduling policy* to adjust the labeling budgets between the two selection strategies over time. Since the key idea of the exploration technique is to reduce the domain gap, it greatly improves the performance in

the first few rounds when the domain gap is large but is less effective later when the two domains are almost aligned as shown in Tab. 2. In contrast, as discussed in Fig. 2, the strength of uncertainty is in later rounds (d, e, f), while its defect of ignoring high-confident but erroneous areas usually occurs in the earlier rounds (a, b, c). Based on the analysis, we designed a dynamic scheduling policy to allocate more labeling budgets for our density selection in earlier rounds and more for uncertainty selection later.

Experiments showed that our proposed **D²ADA** surpasses existing active learning and domain adaptation baselines on two widely used benchmarks, GTA5 \rightarrow Cityscapes and SYNTHIA \rightarrow Cityscapes. Moreover, with less than 5% target domain labels, our method reaches comparable results with full supervision.

To sum up, our contributions are highlighted as follows,

- We open up a new way for ADA that utilizes domain density as an active selection criterion.
- We design a dynamic scheduling policy to adjust the budgets between model uncertainty and domain exploration in the ADA problem.
- Extensive experiments demonstrate our method outperforms current state-of-the-art active learning and domain adaptation methods.

2 Related Work

We introduce domain adaptation methods for semantic segmentation with none or few target labels and briefly review the progress of deep active learning.

2.1 Unsupervised Domain Adaptation for Semantic Segmentation

Many researchers have delved into unsupervised domain adaptation (UDA) methods for semantic segmentation. One widely-used approach is to align the distribution of the source and target domain through adversarial training [4,12,24,42,43]. Another popular method is to apply self-training techniques [47,55,57,59,60,61] by regularizing model uncertainty and assigning pseudo labels to reliable target domain data. However, the performance of distribution alignment is still not satisfying [28], and self-training often suffers from label noise [55]. These defects make existing UDA methods hard to reach the performance of full supervision on labeled target domain data.

2.2 Domain Adaptation for Semantic Segmentation with Few Target Labels

To reduce the performance gap between UDA methods and supervision on the whole labeled target domain data, some recent works have explored the utilization of weak and few target domain annotations.

[30] introduced a weakly-supervised domain adaptation (WDA) framework for semantic segmentation, where image-level annotations are provided initially, and few manual or pseudo pixel-level labels are gradually added.

Some considered the semi-supervised domain adaptation (SSDA) scenario, where few pixel-level annotations are available during training. [48] attempted to align the features of the source and target domain globally and semantically. [9] reduced the domain shift by mixing domains and knowledge distillation.

Unlike the above settings, where few labels were initially given, others used active learning strategies to acquire few target annotations from humans. For example, [28] chose diverse target domain images dissimilar to multiple source domain anchors for labeling. [36] utilized the inconsistency map generated by the disagreement of two classifiers to acquire few uncertain target pixel labels.

Different from those prior works focusing on uncertain or diverse samples, our density-aware selection method acquires labels of few but vital samples that can bridge the gap between the source and target domains.

2.3 Active Learning for Domain Adaptation

Active learning can reduce the annotation effort by gradually selecting the most valuable unlabeled data to be labeled.

Conventional active learning strategies can be roughly divided into two categories: uncertainty-based methods and diversity-aware approaches. The main idea of uncertainty-based methods is to acquire labels of data close to the model decision boundary. For example, [44,46] utilized the softmax entropy, confidence, or margin of network outputs as the selection criterion. [14,15] proposed a better measurement of uncertainty with MC-dropout. On the other hand, the key concept of diversity approaches is querying annotations of a batch of samples far away from each other in the feature space. Prior works used clustering [2] or greedy selection [21,35,50] to achieve the diversity in a query batch.

In addition, several recent works have studied active domain adaptation scenarios that consider both uncertainty and diversity. [40,58] used the trained classifier and domain discriminator to select highly uncertain samples far from the source domain distribution. [13,31,39] utilized clustering to select highly uncertain samples that are diverse in target domain distribution. Very recently, [52] proposed an energy-based approach and [51] presented the concept of regional impurity to address this problem.

To the best of our knowledge, we propose the first density-aware active domain adaptation strategy to acquire labels of representative samples with high density in the target domain but low density in the source domain. Also, we developed the first dynamic scheduling policy between domain exploration and model uncertainty for active domain adaptation.

3 Method

3.1 Overview

In the Active Domain Adaptation (ADA) problem, we have a data pool of C -category semantic segmentation from two domains, comprising rich labeled data

D_S in the source domain and some data D_T in the target domain. D_T can be divided into D_T^L and D_T^U , where D_T^L contains the data in the target domain that has been labeled and D_T^U stores the remaining unlabeled data.

Given the initial D_T without any annotations, *i.e.*, $D_T^L = \emptyset$, the ADA algorithm iteratively acquires annotations of few data in D_T^U to maximize the performance improvement over the test data in the target domain. The label acquisition process will be performed several rounds and the labeling budgets in single round is a fixed number of pixels.

Our **D**ynamic **D**ensity-aware **A**ctive **D**omain **A**daptation (**D²ADA**) framework can be divided into 3 steps: (1) Train an initial model on $D_S \cup D_T$ with UDA methods as a warmup step. (2) Perform the density-aware selection to determine representative regions for each category according to the domain density and balance the budgets across classes. (Sec. 3.2). (3) Dynamically determine the labeling budgets for domain exploration and model uncertainty in each round. Then, for the top-ranked regions in D_T^U , acquire ground truth labels and move them to D_T^L . Finally, fine-tune the model on $D_S \cup D_T^L$ in a supervised manner and go back to step (2) for a new round (Sec. 3.3). The complete and detailed algorithm can be referred in the supplementary material.

3.2 Density-aware Selection

To efficiently label the most informative target domain data, unlike prior domain exploration techniques causing biased samples (Fig. 1(a)) or redundant labeling (Fig. 1(b)), our density-aware approach (Fig. 1(c)) selects data representative in the target distribution but uncommon in the source distribution. To achieve this, we introduce the concept of **domain density** to calculate the prevalence of data in a single domain and estimate the categorical domain gaps between the two domains. By selecting data with the largest domain gaps, we can adapt the model to the target domain with the least possible amount of queried annotations.

In the following section, we first introduce the definition of domain density and provide ways to estimate it for each category. Then, we elaborate on our density difference metric and its theoretical foundation. Finally, we describe the motivation and operation of our class-balanced selection.

Domain Density Estimation. Domain density is defined as a measurement of how common a region R is predicted as a certain class in a particular domain. Following prior works [3, 19, 25, 38], we divide an image into multiple regions using the widely-used SLIC [1] algorithm. To derive the domain density, we first calculate the feature z of a region R as the average feature over all the pixels within R . Similarly, a category $c \in [C] = \{1, 2, \dots, C\}$ is assigned to a region based on the predicted probability averaged across all pixels in the region. Given a domain S , a feature vector z , and a category c , we define the **source domain density**

$$d_S = p_S(z|c) \quad (1)$$

as the probability that z is classified as c in S . To estimate $p_S(z|c)$, we construct the density estimators from all the observed (z, c) pairs in the source domain dataset. The target domain follows the manner similarly.

In our implementation, a set of Gaussian Mixture Models (GMMs) are served as the density estimators. The construction of GMMs can be efficiently completed by offline and parallel execution, which takes about 0.01 seconds per region with an 8-core CPU personal computer. More discussions and details are left in the supplementary material.

Density Difference as Metric. The key idea of our density-aware method is to select unlabeled target regions that contribute most to the domain gap so that we can adapt the model from the existing labeled source domain to the target domain efficiently. In practice, we use the difference between the target and the source domain density of a single region to evaluate the domain gap.

For the i -th unlabeled target region, we first obtain its source domain density d_S^i and target domain density d_T^i by feeding (z_i, c_i) to the pre-constructed density estimators. Then, we introduce a metric to rank all regions in D_T^U :

$$\pi_i = \log\left(\frac{d_T^i}{d_S^i}\right), \quad (2)$$

where $i = 1, 2, \dots, |\#\text{regions} \in D_T^U|$. The ratio at the log scale is the log-likelihood difference of the two domain densities [16]. According to Eq. 2, regions with larger π are more prevalent in the target domain and less observable in the source domain. Therefore, we select regions with large π values to replenish the insufficient knowledge of the source domain and to maximize the exploration of the target domain. For each category c , we sorted all unlabeled regions based on their importance scores. Then, a fixed number of top-ranked regions for each category will be selected for labeling until running out of annotation budgets.

Theoretical Foundation. To provide the theoretical foundations for our proposed density difference metric, we leverage the proposition presented in [27]

$$\ell_{test} \leq \ell_{train} + \frac{M}{\sqrt{2}} \sqrt{D_{\text{KL}}(p_T(c, z) \parallel p_S(c, z))}, \quad (3)$$

where ℓ_{test} and ℓ_{train} denote the testing and training loss respectively. p_S and p_T are the joint probability distributions of the category c and the extracted feature z in the source and the target domain distribution respectively. M is a constant term and $D_{\text{KL}}(\parallel)$ is the KL divergence of the two distributions.

As shown in Eq. 3, to provide a tighter generalization bound for ℓ_{test} , we aim to minimize the KL of the two joint probability distributions as follows:

$$\begin{aligned} D_{\text{KL}}(p_T(c, z) \parallel p_S(c, z)) &= D_{\text{KL}}(p_T(c) \parallel p_S(c)) \\ &\quad + \mathbb{E}_{p_T(c)}[D_{\text{KL}}(p_T(z|c) \parallel p_S(z|c))]. \end{aligned} \quad (4)$$

The proof of Eq. 4 is provided in the supplementary material.

Since $p_T(c)$ and $p_S(c)$ are similar in most cases, we can obtain better generalization to the target domain by minimizing the expected value of the KL divergence from $p_S(z|c)$ to $p_T(z|c)$. Additionally, to deal with the label imbalance

problem and to make the model perform well for each category for semantic segmentation, we opt to reduce the categorical KL term, *i.e.*, $D_{\text{KL}}(p_T(z|c) \parallel p_S(z|c))$.

Although it is impossible to obtain the precise KL term by sampling all data points in the high dimensional continuous feature space, we could leverage the Monte Carlo simulation process to estimate the KL of the two distributions [18]. Given a sufficient number of N observed region features $\{z_i\}_{i=1}^N$ predicted to a certain category c , the KL divergence of the two conditional probability distributions can be approximated as:

$$D_{\text{KL}}(p_T(z|c) \parallel p_S(z|c)) \approx \frac{1}{N} \sum_{i=1}^N \log\left(\frac{p_T(z_i|c)}{p_S(z_i|c)}\right) = \frac{1}{N} \sum_{i=1}^N \log\left(\frac{d_T^i}{d_S^i}\right) = \frac{1}{N} \sum_{i=1}^N \pi_i, \quad (5)$$

where π is our designed metric to rank all regions in the same class (see Eq. 2).

Eq. 3, 4, 5 bridge the theory and our proposed metric. By providing annotations for data that contribute the most to the KL divergence (large π value) for each category, the domain shift between the source and target domain could be reduced, making the generalization bound of ℓ_{test} tighter.

Class-balanced Selection. With our proposed density difference metric, the labeling budget in each round can be equally divided into C categories. Nonetheless, we demonstrate that ADA can be achieved more efficiently.

We empirically observe the required budget for domain adaptation varies across classes, and thus it is inefficient to allocate the same budget to each class. For example, UDA methods can reach the performance close to fully supervised learning in some classes like “vegetarian” and “building”. However, it is difficult for them to generalize to others like “train” or “sidewalk” (see Tab. 1). Hence, we propose a class-balanced selection to fully utilize the labeling budget.

Intuitively, we would like to spend more budgets on hard categories, while fewer selection budgets for already well-aligned ones. According to Eq. 5, the average π scores of a particular category can be regarded as an approximation of categorical KL divergence from the source domain to the target domain. Therefore, we use this indicator to balance the selection budget across classes. Suppose the total labeling budget of our density selection in each round is B^d pixels, for a certain category c , the allocated budget $B^{d,c}$ is as follows:

$$B^{d,c} = \frac{w_c}{\sum_{i \in [C]} w_i} \cdot B^d, \quad w_c = \sigma(D_{\text{KL}}(p_T(z|c) \parallel p_S(z|c))). \quad (6)$$

In the Eq. 6, σ is a normalization function to make the weighting term falls within a reasonable range. The equation ensures that more labels are assigned to classes with larger domain shift, while less are distributed to well-aligned ones.

3.3 Dynamic Scheduling Policy

As described in Sec. 1, prior ADA practices acquired labels considering both uncertainty and domain exploration equally over time. However, these practices,

not aware of the defects of both methods under different stages, caused poor labeling efficiency, especially under low labeling budgets in earlier rounds.

As shown in the comparison of Fig. 2 (a, b, c) and (d, e, f), uncertainty indicator is less effective under severe domain shift but provides informative cues when more target domain labeled data are acquired. Conversely, since the design concept of domain exploration is to conquer the domain shift, it is effective in the earlier rounds in ADA but the growth rapidly slows down in the later rounds as shown in Tab. 2 (a, c). As a result, we develop a dynamic scheduling policy to benefit from both selection strategies and facilitate labeling efficiency.

Let B be the constant labeling budgets of each active selection round. In the n -th round, B can be divided to B_n^d for the density-aware approach and B_n^u for an uncertainty-based method. Our budget allocation arrangements for the two approaches are as follows:

$$\begin{aligned}\lambda &= \alpha \cdot 2^{-\beta(n-1)}, \\ B_n^d &= \lambda B, \quad B_n^u = (1 - \lambda)B,\end{aligned}\tag{7}$$

where α is the balance coefficient between density-aware and the uncertainty-based method, and β is the decay rate of the labeling budget of the density-aware approach. The reason for using half decay is to reflect the observation of the rapid domain shift reduction, which is discussed in Sec. 4.4. The above formula ensures that the density-aware method is heavily relied on when the domain gaps are large in the first few rounds, and the uncertainty-based method is mainly used when the gaps are diminished in the later rounds. The effectiveness of the designed policy is shown in Tab. 2. In our implementation, we use the conventional softmax entropy [44] as the uncertainty measurement.

After deciding the labeling budgets B_n^d and B_n^u for density-aware and uncertainty methods with our dynamic scheduling policy in each round, we select regions with the two selection strategies. Then, we acquire ground truth labels for these regions and move them from the unlabeled target set D_T^U to the labeled target set D_T^L . Finally, we fine-tune the model with $D_S \cup D_T^L$ under full supervision and then execute the next active selection round.

4 Experiments

4.1 Experimental Settings

We elaborate the datasets, networks and active learning protocol in our experiments. The implementation details are reported in the supplementary material.

Datasets. We evaluated various active learning and domain adaptation methods on two widely-used domain adaptive semantic segmentation benchmarks: GTA5 \rightarrow Cityscapes and SYNTHIA \rightarrow Cityscapes. Cityscapes [11] is a real-world self-driving dataset containing 2975 labeled training images, 500 labeled validation images and 1225 test images. GTA5 [32] is a synthetic dataset consisting of 24966 annotated images, which shares 19 semantic categories with Cityscapes. SYNTHIA dataset [33] has 9400 synthetic annotated images, which share 16

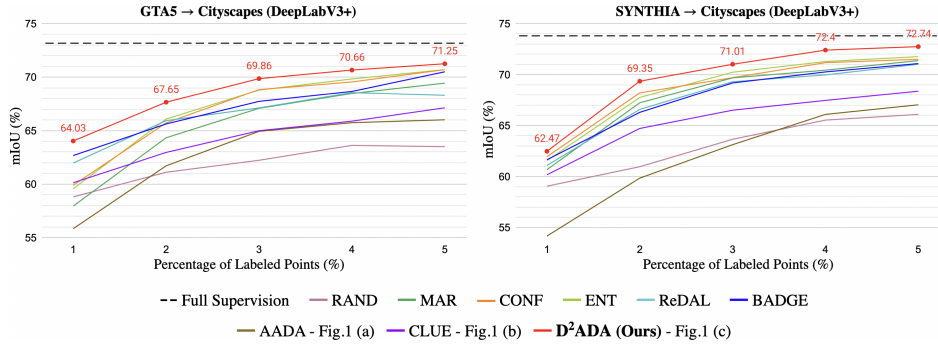


Fig. 3. Comparison with different active learning baselines. On two widely-used benchmarks, our **D²ADA** outperforms 8 existing active learning strategies, including uncertainty-based methods (MAR, CONF, ENT), hybrid approaches (ReDAL, BADGE), and ADA practices. The result suggested that compared to prior domain exploration methods shown in Fig. 1 (a, b), our density-aware method (c) achieves significant improvement. More explanations and observations are reported in Sec. 4.2.

semantic categories with Cityscapes. For a fair comparison, we evaluated all methods on the Cityscapes validation split.

Models. We used DeepLabV3+ [8], a semantic segmentation model, to evaluate all active learning strategies. To fairly compare our method with existing domain adaptation approaches, we report the results on both DeepLabV2 [6] and DeepLabV3+ [8]. The two models are based on the ResNet-101 [17] backbone.

Active Learning Protocol. For all experiments, we first adopted [42] to train the network with adversarial training as a warmup step. Then we performed N -round active selection comprising of the following steps: (1) Select a small number of regions with B pixels from the unlabeled Cityscapes training split D_T^U based on different active selection strategies. (2) Acquire the labels of these selected data and add them to D_T^L . (3) Fine-tune the model with $D_S \cup D_T^L$ in a supervised manner. In our setting, we choose $N = 5$, $B = 1\%$ of the total number of pixels in Cityscapes training split for both tasks.

4.2 Comparison with Active Learning Baselines

We compared **D²ADA** with 8 other active learning strategies, including random region selection (RAND), uncertainty-based methods (CONF [44], MAR [44], and ENT [44]), hybrid methods (BADGE [2] and ReDAL [50]), and existing Active Domain Adaptation baselines (AADA [40] and CLUE [31]). The implementation details are described in the supplementary material.

The experimental results are shown in Fig. 3. In each subplot, the x-axis indicates the percentage of total target domain annotated points and the y-axis is the corresponding mIoU score achieved by the network.

Table 1. Comparison with different domain adaptation approaches on (a) GTA5 \rightarrow Cityscapes and (b) SYNTHIA \rightarrow Cityscapes. Our proposed **D²ADA** outperforms any existing methods on the overall mIoU and most per-class IoU with only 5% target annotations with different network backbones. mIoU* in (b) denotes the averaged scores across 13 categories used in [42]. To fairly compare all methods under the same number of annotations, we draw a chart in Fig. 4.

(a) GTA5 \rightarrow Cityscapes																					
Method		Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
UDA	AdaptSeg [42]	86.5	36.0	79.9	23.4	23.3	35.2	14.8	14.8	83.4	33.3	75.6	58.5	27.6	73.7	32.5	35.4	3.9	30.1	28.1	42.8
	ADVENT [43]	89.9	36.5	81.2	29.2	25.2	28.5	32.3	22.4	83.9	34.0	77.1	57.4	27.9	83.7	29.4	39.1	1.5	28.4	23.3	48.4
	SIMDA [49]	90.6	44.7	84.8	34.3	28.7	31.6	35.0	37.6	84.7	43.3	85.3	57.0	31.5	83.8	42.6	48.5	1.9	30.4	39.0	49.2
	CBST [60]	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
	CRST [61]	91.0	55.4	80.0	33.7	21.4	37.3	32.9	24.5	85.0	34.1	80.8	57.7	24.6	84.1	27.8	30.1	26.9	26.0	42.3	47.1
	LTIR [20]	92.9	55.0	85.3	34.2	31.1	34.9	40.7	34.0	85.2	40.1	87.1	61.0	31.1	82.5	32.3	42.9	0.3	36.4	46.1	50.2
	FDA [53]	92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	53.1	16.9	27.7	46.4	50.5
	TPLD [37]	94.2	60.5	82.8	36.6	16.6	39.3	29.0	25.5	85.6	44.9	84.4	60.6	27.4	84.1	37.0	47.0	31.2	36.1	50.3	51.2
	IAST [26]	93.8	57.8	85.1	39.5	26.7	26.2	43.1	34.7	84.9	32.9	88.0	62.6	29.0	87.3	39.2	49.6	23.2	34.7	39.6	51.5
	ProDA [55]	87.8	56.0	79.7	46.3	44.8	45.6	53.5	53.5	88.6	45.2	83.1	70.7	39.2	88.8	45.5	59.4	1.0	48.9	56.4	57.5
	CAMDA [59]	91.3	46.0	84.5	34.4	29.7	32.6	35.8	36.4	84.5	43.2	82.0	60.0	32.2	83.2	35.0	46.7	0.0	33.7	42.2	49.2
	DPL-Dual [10]	92.8	54.4	86.2	41.6	32.7	36.4	49.0	34.0	85.8	41.3	86.0	63.2	34.2	87.2	39.3	44.5	18.7	42.6	43.1	53.3
BAPA-Net [23]	94.4	61.0	88.0	26.8	39.9	38.3	46.1	55.3	87.8	46.1	89.4	68.8	40.0	90.2	60.4	59.0	0.0	45.1	54.2	57.4	
WDA	WDA [30] (Point)	94.0	62.7	86.3	36.5	32.8	38.4	44.9	51.0	86.1	43.4	87.7	66.4	36.5	87.9	44.1	58.8	23.2	35.6	55.9	56.1
SSDA	ASS [48] (+50 city)	94.3	63.0	84.5	26.8	28.0	38.4	35.5	48.7	87.1	39.2	88.8	62.2	16.3	87.6	23.2	39.2	7.2	24.4	58.1	50.4
ADA	LabOR [36] (V2, 2.2%)	96.6	77.0	89.6	47.8	50.7	48.0	56.6	63.5	89.5	57.8	91.6	72.0	47.3	91.7	62.1	61.9	48.9	47.9	65.3	66.6
	MADA [28] (V3+, 5%)	95.1	69.8	88.5	43.3	48.7	45.7	53.3	59.2	89.1	46.7	91.5	73.9	50.1	91.2	60.6	56.9	48.4	51.6	68.7	64.9
	RIPU [51] (V3+, 5%)	97.0	77.3	90.4	54.6	53.2	47.7	55.9	64.1	90.2	59.2	93.2	75.0	54.8	92.7	73.0	79.7	68.9	55.5	70.3	71.2
	D²ADA (V2, 5%)	96.3	73.6	89.3	50.0	52.3	48.0	56.9	64.7	89.3	53.9	92.3	73.9	52.9	91.8	69.7	78.9	62.7	57.7	71.1	69.7
	D²ADA (V3+, 5%)	97.0	77.8	90.0	46.0	55.0	52.7	58.7	65.8	90.4	58.9	92.1	75.7	54.4	92.3	69.0	78.0	68.5	59.1	72.3	71.3
Target Only	DeepLabV2 [6]	97.4	79.5	90.3	51.1	52.4	49.0	57.5	68.0	90.5	58.1	93.1	75.1	53.9	92.7	72.0	80.2	65.0	58.1	71.1	71.3
	DeepLabV3+ [8]	97.6	81.3	91.1	49.8	57.6	53.8	59.6	69.1	91.2	60.5	94.4	76.7	55.6	93.3	75.8	79.9	72.9	57.7	72.2	73.2
(b) SYNTHIA \rightarrow Cityscapes																					
Method		Road	SW	Build	Wall*	Fence*	Pole*	TL	TS	Veg.	Sky	PR	Rider	Car	Bus	Motor	Bike	mIoU	mIoU*		
UDA	AdaptSeg [42]	79.2	37.2	78.8	-	-	-	9.9	10.5	78.2	80.5	53.5	19.6	67.0	29.5	21.6	31.3	-	45.9		
	ADVENT [43]	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	84.1	57.9	23.8	73.3	36.4	14.2	33.0	41.2	48.0		
	SIMDA [49]	83.0	44.0	80.3	-	-	-	17.1	15.8	80.5	81.8	59.9	33.1	70.2	37.3	28.5	45.8	-	52.1		
	CBST [60]	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		
	CRST [61]	67.7	32.2	73.9	10.7	1.6	37.4	22.2	31.2	80.8	80.5	60.8	29.1	82.8	25.0	19.4	45.3	43.8	50.1		
	LTIR [20]	92.6	53.2	79.2	-	-	-	1.6	7.5	78.6	84.4	52.6	20.0	82.1	34.8	14.6	39.4	-	49.3		
	FDA [53]	79.3	35.0	73.2	-	-	-	19.9	24.0	61.7	82.6	61.4	31.1	83.9	40.8	38.4	51.1	-	52.5		
	TPLD [37]	80.9	44.3	82.2	19.9	0.3	40.6	20.5	30.1	77.2	80.9	60.6	25.5	84.8	41.1	24.7	43.7	47.3	53.5		
	IAST [26]	81.9	41.5	83.3	17.7	4.6	32.3	30.9	28.8	83.4	85.0	65.5	30.8	86.5	38.2	33.1	52.7	49.8	57.0		
	ProDA [55]	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		
	CAMDA [59]	82.5	42.2	81.3	-	-	-	18.3	15.9	80.6	83.5	61.4	33.2	72.9	39.3	26.6	43.9	-	52.4		
	DPL-Dual [10]	87.5	45.7	82.8	13.3	0.6	33.2	22.0	20.1	83.1	86.0	56.6	21.9	83.1	40.3	29.8	45.7	47.0	54.2		
BAPA-Net [23]	91.7	53.8	83.9	22.4	0.8	34.9	30.5	42.8	86.6	88.2	66.0	34.1	86.6	51.3	29.4	50.5	53.3	61.2			
WDA	WDA [30] (Point)	94.9	63.2	85.0	27.3	24.2	34.9	37.3	50.8	84.4	88.2	60.6	36.3	86.4	43.2	36.5	61.3	57.2	63.7		
SSDA	ASS [48] (+50 city)	94.1	63.9	87.6	-	-	-	18.1	37.1	87.5	89.7	64.6	37.0	87.4	38.6	23.2	59.6	-	60.7		
ADA	MADA [28] (V3+, 5%)	96.5	74.6	88.8	45.9	43.8	46.7	52.4	60.5	89.7	92.2	74.1	51.2	90.9	60.3	52.4	69.4	68.1	73.3		
	RIPU [51] (V3+, 5%)	97.0	78.9	89.9	47.2	50.7	48.5	55.2	63.9	91.1	93.0	74.4	54.1	92.9	79.9	55.3	71.0	71.4	76.7		
	D²ADA (V2, 5%)	96.4	76.3	89.1	42.5	47.7	48.0	55.6	66.5	89.5	91.7	75.1	55.2	91.4	77.0	58.0	71.8	70.6	76.3		
	D²ADA (V3+, 5%)	96.7	76.8	90.3	48.7	51.1	54.2	58.3	68.0	90.4	93.4	77.4	56.4	92.5	77.5	58.9	73.3	72.7	77.7		
Target Only	DeepLabV2 [6]	97.4	79.5	90.3	51.1	52.4	49.0	57.5	68.0	90.5	93.1	75.1	53.9	92.7	80.2	58.1	71.1	72.5	77.5		
	DeepLabV3+ [8]	97.6	81.3	91.1	49.8	57.6	53.8	59.6	69.1	91.2	94.4	76.7	55.6	93.3	79.9	57.7	72.2	73.8	78.4		

The improvement of our proposed **D²ADA** is significant compared to prior active learning strategies. Under 5% labels, **D²ADA** surpassed the second-

placed approach by 0.55 mIoU in $\text{GTA} \rightarrow \text{Cityscapes}$, whereas the gap between the second and fourth place was only 0.2. Similarly, in $\text{SYNTHIA} \rightarrow \text{Cityscapes}$, **D²ADA** outperformed second place by 0.98 mIoU, while the difference between the second and fourth place was 0.39. Moreover, our method consistently prevailed over other methods by more than 0.5 mIoU under all cases. The results verify our method makes huge progress in active domain adaptation problems.

On the $\text{GTA5} \rightarrow \text{Cityscapes}$, our proposed **D²ADA** surpasses uncertainty-based active learning methods (ENT, CONF, and MAR) for over 3% mIoU at the first round. This suggests that density as a selection metric clearly benefits ADA over uncertainty against domain shift in the first few rounds.

Compared with existing ADA methods (AADA, CLUE) or hybrid active learning approaches (ReDAL, BADGE), our method also achieves better performance under any tasks and with any amount of budget. On the two tasks, AADA even performs worse than random selection in the early stage. We conjecture that the selected biased samples might be inconsistent with the real target distribution as shown in Fig.1 (a). As for CLUE, the methods with 5% annotations still performed worse than our method with merely 2.5x fewer labels (2% annotations) on the two tasks. We infer the main reason might be that they only considered choosing diverse target domain data but did not avoid the selection that were likely to appear in the source distribution as observed in Fig.1 (b).

4.3 Comparison with Domain Adaptation Methods

We compared our **D²ADA** with various domain adaptation methods, including unsupervised domain adaptations (UDAs) [10,20,23,26,37,42,43,49,53,55,59,60] weakly-supervised domain adaptation (WDA) [30], semi-supervised domain adaptation (SSDA) [9,48] and active domain adaptations (ADA) [28,36,51,52]¹.

Tab. 1 (a, b) shows the result of $\text{GTA5} \rightarrow \text{Cityscapes}$ and $\text{SYNTHIA} \rightarrow \text{Cityscapes}$ respectively. Compared with the current state-of-the-art UDA method [55], our method achieves an advantage of more than 10% mIoU with only 5% annotation effort. Furthermore, the performance of our method can achieve over 97% of the result of full supervision on target labeled data on both DeepLabV2 and DeepLabV3+ models. On the two tasks, compared with other label-efficient approaches (WDA, SSDA, ADA), our method not only performs better on the overall mIoU score but also gains improvement on small objects, such as “traffic sign (TS)” or “bicycle”, and difficult categories, like “fence” and “pole”.

Since the labeling budgets used in prior label-efficient methods differ, Fig. 4 makes a fair comparison of these methods. The x-axis indicates the percentage of total target domain annotated points and the y-axis is the corresponding mIoU score achieved by the network. Obviously, our approach achieves higher performance over existing methods with the same number of labeled pixels.

Furthermore, according to statistics, our selected regions contain only 3.1 different categories on average, with 50% regions less than two categories. Thus,

¹ We report the official results of the UDA, WDA, SSDA and ADA methods and train the DeepLabV2 and DeepLabV3+ ourselves.

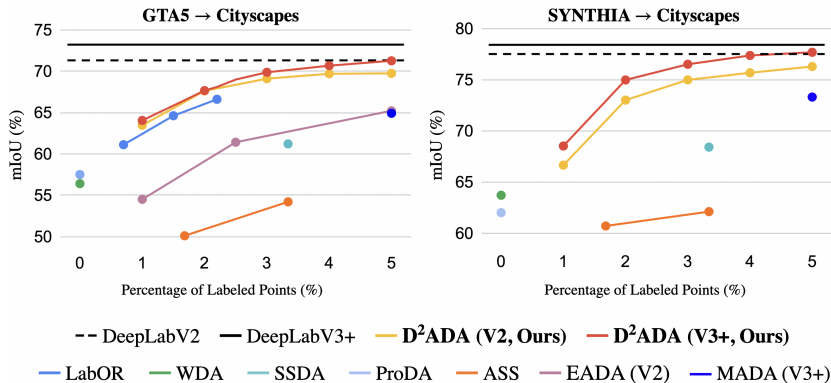


Fig. 4. Comparison with various label-efficient domain adaptation methods. We compare our method with the state-of-the-art UDA method, ProDA [55] as well as various label-efficient approaches, including WDA [30], SSDA [9], ASS [48], MADA [28], LabOR [36] and EADA [52]. On two tasks and networks, our proposed **D²ADA** achieves the best result under the same number of labels.

Table 2. Ablation studies. The three columns on the right show the model performance when 1%, 3%, and 5% target annotations are acquired by different active selection strategies. The results validate the effectiveness of our density-aware method, class-balanced selection, and dynamic scheduling policy as the discussion in Sec. 4.4. It also demonstrates uncertainty perform worse with low labeling budgets (1%, 3%).

	Components				mIoU wrt. labels		
	uncertainty	density aware	class balance	dynamic scheduling	1%	3%	5%
(a)	✓				59.97	68.79	70.70
(b)		✓			62.75	68.81	69.82
(c)		✓	✓		64.03	69.38	70.69
(d)	✓	✓	✓		63.30	69.66	71.15
(e)	✓	✓	✓	✓	64.03	69.86	71.25

compared with prior works selecting scattered pixels [36] or small regions [51], our method, selecting superpixel-level regions, are more friendly for annotators to label (by drawing few polygons) and would require much less labeling effort.

4.4 Ablation Studies

We also conduct experiments to validate the effectiveness of all the proposed components with DeepLabV3+ backbone on **GTA5 → Cityscapes**. Extensive experiments about hyper-parameter settings and other in-depth analyses are left in the supplementary material.

Tab. 2 demonstrates the effectiveness of our density-aware selection, class-balanced selection, and dynamic scheduling policy. First, from the comparison of

(a, b), with low budgets (1%, 3%), our density-aware selection solves the defect of uncertainty metrics, echoing our motivations stated in Fig. 2.

Second, as shown in Fig. 5, due to the class-balanced selection, our method queries more labels on minor classes but fewer labels on well-learned classes. Tab. 2 (b, c) also shows our class-balanced selection boosts the performance by over 0.5 mIoU in all cases. These results suggest that spending more labeling budgets on hard categories but few budgets on well-aligned ones is effective.

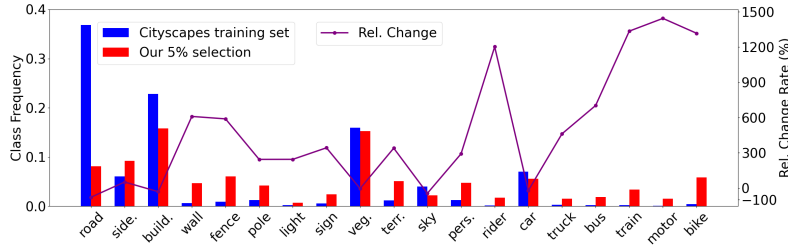


Fig. 5. The label distribution of our selected regions. Following [51], we draw histograms to compare the label distribution of the original target domain data (blue) and our 5% selected regions (red). The purple line chart shows the relative changes from the target dataset to our selection. Evidently, our class-balanced selection acquires more labels on minor classes, like 12x more labels on “train”, “motor”, and “bike”.

Third, as shown in Tab. 2 (a, e) and (c, e), given any budget, our dynamic scheduling policy surpasses the density or uncertainty methods alone. Also, the comparison of (d) and (e) validates our policy outperforms the “equally distributed” budget allocation under all situations. The results indicate that our dynamic scheduling strategy can take advantage of the two methods, further improving the labeling efficiency.

5 Conclusion

We propose **D²ADA**, a novel active domain adaptation method for semantic segmentation. By labeling samples with high probability density in the target domain yet with low probability density in the source domain, models can conquer the domain shift with limited annotation effort. Furthermore, we design the first dynamic scheduling policy to balance labeling budgets between domain exploration and model uncertainty for the ADA problem.

Acknowledgement

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant MOST 110-2634-F-002-051, Mobile Drive Technology Co., Ltd (MobileDrive), and Industrial Technology Research Institute (ITRI). We are grateful to the National Center for High-performance Computing.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slc superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* **34**(11), 2274–2282 (2012) [6](#)
2. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse, uncertain gradient lower bounds. In: *ICLR* (2020) [2](#), [5](#), [10](#)
3. Casanova, A., Pinheiro, P.O., Rostamzadeh, N., Pal, C.J.: Reinforced active learning for image segmentation. In: *International Conference on Learning Representations* (2020), <https://openreview.net/forum?id=SkGC6TNFvr> [6](#)
4. Chang, W.L., Wang, H.P., Peng, W.H., Chiu, W.C.: All about structure: Adapting structural information across domains for boosting semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1900–1909 (2019) [1](#), [4](#)
5. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: 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) [1](#)
6. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K.P., Yuille, A.L.: Semantic image segmentation with deep convolutional nets and fully connected crfs. *CoRR abs/1412.7062* (2015) [1](#), [10](#), [11](#)
7. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017) [1](#)
8. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*. pp. 801–818 (2018) [1](#), [10](#), [11](#)
9. Chen, S., Jia, X., He, J., Shi, Y., Liu, J.: Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 11018–11027 (2021) [5](#), [12](#), [13](#)
10. Cheng, Y., Wei, F., Bao, J., Chen, D., Wen, F., Zhang, W.: Dual path learning for domain adaptation of semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 9082–9091 (2021) [1](#), [11](#), [12](#)
11. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3213–3223 (2016) [2](#), [9](#)
12. Du, L., Tan, J., Yang, H., Feng, J., Xue, X., Zheng, Q., Ye, X., Zhang, X.: Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 982–991 (2019) [1](#), [4](#)
13. Fu, B., Cao, Z., Wang, J., Long, M.: Transferable query selection for active domain adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7272–7281 (2021) [2](#), [5](#)
14. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: *international conference on machine learning*. pp. 1050–1059 (2016) [2](#), [5](#)

15. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: International Conference on Machine Learning. pp. 1183–1192 (2017) [5](#)
16. Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: The elements of statistical learning: data mining, inference, and prediction, vol. 2. Springer (2009) [7](#)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) [10](#)
18. Hershey, J.R., Olsen, P.A.: Approximating the kullback leibler divergence between gaussian mixture models. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. vol. 4, pp. IV–317. IEEE (2007) [8](#)
19. Kasarla, T., Nagendar, G., Hegde, G.M., Balasubramanian, V., Jawahar, C.: Region-based active learning for efficient labeling in semantic segmentation. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1109–1117. IEEE (2019) [6](#)
20. Kim, M., Byun, H.: Learning texture invariant representation for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12975–12984 (2020) [1](#), [11](#), [12](#)
21. Kirsch, A., van Amersfoort, J., Gal, Y.: Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. In: Advances in Neural Information Processing Systems. pp. 7026–7037 (2019) [2](#), [5](#)
22. Lin, G., Milan, A., Shen, C., Reid, I.: Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1925–1934 (2017) [1](#)
23. Liu, Y., Deng, J., Gao, X., Li, W., Duan, L.: Bapa-net: Boundary adaptation and prototype alignment for cross-domain semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8801–8811 (2021) [1](#), [11](#), [12](#)
24. Luo, Y., Zheng, L., Guan, T., Yu, J., Yang, Y.: Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2507–2516 (2019) [1](#), [4](#)
25. Mackowiak, R., Lenz, P., Ghorri, O., Diego, F., Lange, O., Rother, C.: CEREALS - cost-effective region-based active learning for semantic segmentation. In: British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018. p. 121. BMVA Press (2018), <http://bmvc2018.org/contents/papers/0437.pdf> [6](#)
26. Mei, K., Zhu, C., Zou, J., Zhang, S.: Instance adaptive self-training for unsupervised domain adaptation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVI 16. pp. 415–430. Springer (2020) [1](#), [11](#), [12](#)
27. Nguyen, A.T., Tran, T., Gal, Y., Torr, P., Baydin, A.G.: KL guided domain adaptation. In: International Conference on Learning Representations (2022), <https://openreview.net/forum?id=0JzqULIVVd> [7](#)
28. Ning, M., Lu, D., Wei, D., Bian, C., Yuan, C., Yu, S., Ma, K., Zheng, Y.: Multi-anchor active domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9112–9122 (2021) [2](#), [4](#), [5](#), [11](#), [12](#), [13](#)
29. Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., Snoek, J.: Can you trust your model’s uncertainty? evaluating

- predictive uncertainty under dataset shift. *Advances in neural information processing systems* **32** (2019) 2
30. Paul, S., Tsai, Y.H., Schuster, S., Roy-Chowdhury, A.K., Chandraker, M.: Domain adaptive semantic segmentation using weak labels. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX* 16. pp. 571–587. Springer (2020) 4, 11, 12, 13
 31. Prabhu, V., Chandrasekaran, A., Saenko, K., Hoffman, J.: Active domain adaptation via clustering uncertainty-weighted embeddings. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 8505–8514 (2021) 2, 5, 10
 32. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: *European conference on computer vision*. pp. 102–118. Springer (2016) 2, 9
 33. Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M.: The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3234–3243 (2016) 9
 34. Roth, D., Small, K.: Margin-based active learning for structured output spaces. In: *European Conference on Machine Learning*. pp. 413–424. Springer (2006) 2
 35. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: *International Conference on Learning Representations* (2018), <https://openreview.net/forum?id=H1aIuk-RW> 2, 5
 36. Shin, I., Kim, D.J., Cho, J.W., Woo, S., Park, K., Kweon, I.S.: Labor: Labeling only if required for domain adaptive semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 8588–8598 (2021) 2, 5, 11, 12, 13
 37. Shin, I., Woo, S., Pan, F., Kweon, I.S.: Two-phase pseudo label densification for self-training based domain adaptation. In: *European conference on computer vision*. pp. 532–548. Springer (2020) 1, 11, 12
 38. Siddiqui, Y., Valentin, J., Nießner, M.: Viewal: Active learning with viewpoint entropy for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9433–9443 (2020) 2, 6
 39. Singh, A., Doraiswamy, N., Takamuku, S., Bhalerao, M., Dutta, T., Biswas, S., Chepuri, A., Vengatesan, B., Natori, N.: Improving semi-supervised domain adaptation using effective target selection and semantics. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2709–2718 (2021) 2, 5
 40. Su, J.C., Tsai, Y.H., Sohn, K., Liu, B., Maji, S., Chandraker, M.: Active adversarial domain adaptation. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 739–748 (2020) 2, 5, 10
 41. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 5693–5703 (2019) 1
 42. Tsai, Y.H., Hung, W.C., Schuster, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7472–7481 (2018) 1, 4, 10, 11, 12
 43. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2517–2526 (2019) 1, 4, 11, 12

44. Wang, D., Shang, Y.: A new active labeling method for deep learning. In: 2014 International Joint Conference on Neural Networks (IJCNN). pp. 112–119 (2014). <https://doi.org/10.1109/IJCNN.2014.6889457> 2, 5, 9, 10
45. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al.: Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* (2020) 1
46. Wang, K., Zhang, D., Li, Y., Zhang, R., Lin, L.: Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology* **27**(12), 2591–2600 (2016) 2, 5
47. Wang, Y., Peng, J., Zhang, Z.: Uncertainty-aware pseudo label refinery for domain adaptive semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 9092–9101 (2021) 4
48. Wang, Z., Wei, Y., Feris, R., Xiong, J., Hwu, W.M., Huang, T.S., Shi, H.: Alleviating semantic-level shift: A semi-supervised domain adaptation method for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 936–937 (2020) 5, 11, 12, 13
49. Wang, Z., Yu, M., Wei, Y., Feris, R., Xiong, J., Hwu, W.m., Huang, T.S., Shi, H.: Differential treatment for stuff and things: A simple unsupervised domain adaptation method for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12635–12644 (2020) 11, 12
50. Wu, T.H., Liu, Y.C., Huang, Y.K., Lee, H.Y., Su, H.T., Huang, P.C., Hsu, W.H.: Redal: Region-based and diversity-aware active learning for point cloud semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 15510–15519 (2021) 2, 5, 10
51. Xie, B., Yuan, L., Li, S., Liu, C.H., Cheng, X.: Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 8068–8078 (2022) 5, 11, 12, 13, 14
52. Xie, B., Yuan, L., Li, S., Liu, C.H., Cheng, X., Wang, G.: Active learning for domain adaptation: An energy-based approach. In: *Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)* (2022) 5, 12, 13
53. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4085–4095 (2020) 1, 11, 12
54. Yuan, Y., Chen, X., Chen, X., Wang, J.: Segmentation transformer: Object-contextual representations for semantic segmentation. In: *European Conference on Computer Vision (ECCV)*. vol. 1 (2021) 1
55. Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: 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*. pp. 12414–12424 (2021) 1, 2, 4, 11, 12, 13
56. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2881–2890 (2017) 1
57. Zheng, Z., Yang, Y.: Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision* **129**(4), 1106–1120 (2021) 4

- 58. Zhou, F., Shui, C., Yang, S., Huang, B., Wang, B., Chaib-draa, B.: Discriminative active learning for domain adaptation. *Knowledge-Based Systems* **222**, 106986 (2021) [2](#), [5](#)
- 59. Zhou, Q., Feng, Z., Gu, Q., Pang, J., Cheng, G., Lu, X., Shi, J., Ma, L.: Context-aware mixup for domain adaptive semantic segmentation. *arXiv preprint arXiv:2108.03557* (2021) [1](#), [4](#), [11](#), [12](#)
- 60. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: *Proceedings of the European conference on computer vision (ECCV)*. pp. 289–305 (2018) [1](#), [4](#), [11](#), [12](#)
- 61. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 5982–5991 (2019) [1](#), [4](#), [11](#)