Dual-Curriculum Teacher for Domain-Inconsistent Object Detection in Autonomous Driving

Longhui Yu[1]
yulonghui@stu.pku.edu.cn
Yifan Zhang[2]
yifan.zhang@u.nus.edu
Lanqing Hong[3†]
honglanqing@huawei.com
Fei Chen[3]
chen.f@huawei.com
Zhenguo Li[3]
li.zhenguo@huawei.com

[1] Peking University, Beijing, China
[2] National University of Singapore, Singapore, Singapore
[3] Huawei Noah's Ark Lab, Hong Kong, China

Abstract

Object detection for autonomous vehicles has received increasing attention in recent years, where labeled data are often expensive while unlabeled data can be collected readily, calling for research on semi-supervised learning for this area. Existing semi-supervised object detection (SSOD) methods usually assume that the labeled and unlabeled data come from the same data distribution. In autonomous driving, however, data are usually collected from different scenarios, such as different weather conditions or different times in a day. Motivated by this, we study a novel but challenging domain-inconsistent SSOD problem. It involves two kinds of distribution shifts among different domains, including (1) data distribution discrepancy, and (2) class distribution shifts, making existing SSOD methods suffer from inaccurate pseudo-labels and hurting model performance. To address this problem, we propose a novel method, namely Dual-Curriculum Teacher (DucTeacher). Specifically, DucTeacher consists of two curriculums, *i.e.,* (1) domain evolving curriculum seeks to learn from the data progressively to handle data distribution discrepancy by estimating the similarity between domains, and (2) distribution matching curriculum seeks to estimate the class distribution for each unlabeled domain to handle class distribution shifts. In this way, DucTeacher can calibrate biased pseudo-labels and handle the domain-inconsistent SSOD problem effectively. We demonstrate the advantages of DucTeacher on SODA10M, the largest publicly available semi-supervised autonomous driving dataset, and COCO, a widely used SSOD benchmark. Experiments show that DucTeacher achieves new state-of-the-art performance on SODA10M with 2.2 mAP improvement and on COCO with 0.8 mAP improvement.

† Lanqing Hong is the corresponding author.

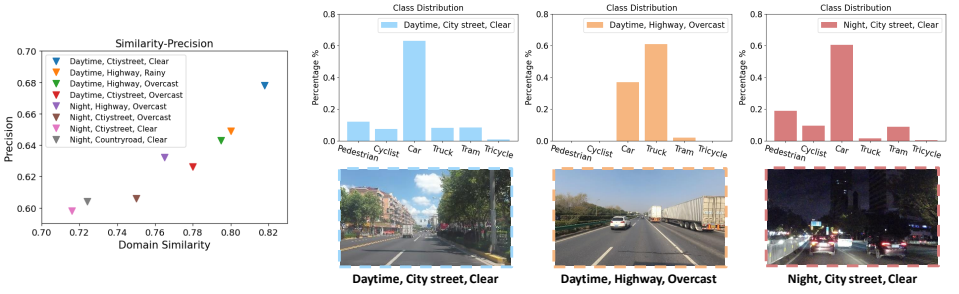
1 Introduction

Autonomous driving [13, 19, 30] has received considerable attention in recent years because of its potential to ease congestion, reduce emissions, and even save lives. However, the timeline for the real-world application of autonomous driving is still uncertain due to the unsatisfactory model performance. One main reason for the limited model performance is the limited size of labeled data, as collecting a large number of annotated data is usually expensive, especially for the tasks like object detection and segmentation. To overcome this problem, semi-supervised learning (SSL) [2, 3, 5, 16, 32, 43], a paradigm to use abundant unlabeled data to improve the model performance with limited annotated data, is a promising technology for autonomous driving.

Object detection is one of the most important tasks in autonomous driving, which provides positioning and classification for crucial targets (e.g., pedestrians) for subsequent route planning [20, 26, 33, 48]. Existing semi-supervised object detection (SSOD) mainly builds upon well-collected datasets, such as ImageNet [8] and COCO [24], assuming that labeled data and unlabeled data are independent and identically distributed (IID). In autonomous driving, however, data are usually collected from various scenarios, such as different weather conditions or different times in a day, resulting in data with different distributions. Motivated by the largest publicly available semi-supervised autonomous driving dataset, SODA10M [15], we aim for a novel setting of SSOD, where there are (1) multiple domains in the data and (2) class distribution shifts among the domains. We name it as **domain-inconsistent SSOD**. See Fig. 1(b) as illustrations.

The above domain-inconsistent setting introduces two challenges for SSOD. One is how to tackle the data distribution shift between labeled and unlabeled data. This problem has been investigated in domain adaption [11, 27, 29, 35, 45] but is usually neglected in semi-supervised learning [2, 3, 5, 16, 32, 34, 39, 43], which would result in noisy pseudo-labels and hurt the model performance. The second challenge is how to track the class distribution shifts in multiple domains. In SSL, the class distribution of the labeled data is usually adopted as a prior to calibrating the pseudo-label distribution of the unlabeled data [2, 16]. However, with class distribution shifts, existing distribution correction methods [2, 16], taking a false class distribution as a reference, would cause bias to the head classes in the labeled data and result in biased pseudo-labels. See Sec. 3 for more discussions on the challenges.

In this work, we propose Dual-Curriculum Teacher (DucTeacher) for the challenging domain-inconsistent SSOD with data distribution shift and class distribution shift. Two curriculum strategies, i.e., distribution matching curriculum (DMC) and domain evolving curriculum (DEC), are proposed to obtain appropriate pseudo-labels for semi-supervised learning. Specifically, DEC is proposed to learn unlabeled data from different domains progressively to alleviate noisy pseudo-labels caused by the data distribution shift. It proposes a difficulty metric to measure the domain similarity and introduces unlabeled data from different domains according to the domain similarity. Moreover, DMC is proposed to adjust the class-specific and domain-adaptive thresholds for pseudo-labeling. It dynamically adjusts the thresholds for each class, to avoid introducing too many head class pseudo-labels or too few tail class pseudo-label by maintaining an unbiased pseudo-label distribution. Extensive experiments on SODA10M [15] and COCO [24] show the superiority of the proposed DucTeacher. Overall, the main contributions of this work are three-fold:



(a) Correlation of domain similarity and precision. (b) Data distribution shifts and class distribution shifts between domains.

Figure 1: (a) Correlation between domain similarity and precision of pseudo-labels. The precision of pseudo-labels produced by the model with Unbiased Teacher [26] is positively associated with the proposed domain similarity provided by DucTeacher. (b) Data distribution shifts and class distribution shifts between domains.

- We target the novel but challenging domain-inconsistent SSOD setting for practical autonomous driving, where the labeled data and unlabeled data come from different domains. Meanwhile, both the data distribution shifts and class distribution shifts happen on the data.
- We propose DucTeacher with two curriculum strategies, DEC and DMC, to provide accurate and unbiased pseudo-labels and improve the performance for semi-supervised object detection.
- In DucTeacher, we develop a novel class distribution estimation method to resist the class distribution shift on the unlabeled data, and a difficulty metric to estimate the domain similarity of unlabeled data from different domains.

2 Related Work

Curriculum Learning. Curriculum learning [1] is a learning strategy inspired by the human learning process, which aims to learn training samples from easy to hard. Prior works [6, 7, 21, 38, 44] have shown that curriculum learning can optimize the learning process and improve performance, especially when the label is noisy or the training cost is limited. However, how to design a curriculum learning strategy in Autonomous Driving for better training effective is under development

Semi-supervised Object Detection. Considering the high cost of collecting annotated data, many semi-supervised methods have been proposed for the object detection task [26, 33]. For example, STAC [33] uses a small set of labeled data to pre-train a detector and then uses the detector to generate pseudo-labels for unlabeled data. Unbiased Teacher [26] improves the pseudo-label generation by using a teacher-student mutual learning framework and alleviates the long-tailed effects of focal loss [25]. Instant-Teaching [48] uses a co-teaching [14] framework to get accurate pseudo-labels. Combating noise [37] alleviates the influences of noise pseudo-label by considering the uncertainty. MA-GCP [23] improves the consistency between the feature of the pseudo-labels and its corresponding global class feature. MUM [22] utilizes the Interpolation-regularization (IR) and proposes a more effective strong-augmentation to improve the effectiveness of the pseudo-labels. In contrast, we

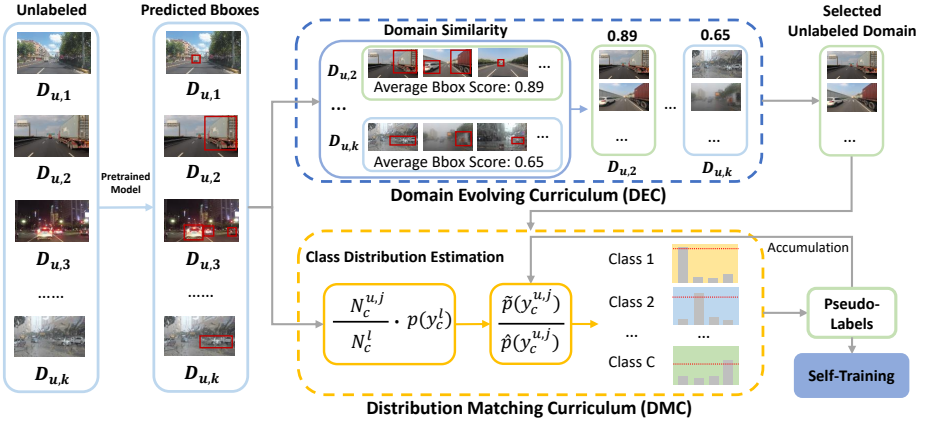


Figure 2: Illustration of the proposed DucTeacher. DucTeacher contains two curricula, DEC and DMC, to obtain accurate and unbiased pseudo-labels. While DEC controls the sampling and DMC adjusts the precision of pseudo-labels. Two curricula work with each other well to provide accurate and unbiased pseudo-labels for domain-inconsistent SSOD.

consider a more challenging setting where the labeled data and unlabeled data come from different domains, causing semi-supervised object detection more difficult.

Cross-domain Object Detection. Cross-domain object detection [4, 6, 10, 17, 31, 41, 49] assumes the source data and target data have different data distributions and emphasizes the high performance on target data. However, they focus on the target domain performance and ignore the performance drop on the source domain, and are limited to the single target domain. Beyond that, multi-target domain adaptation aims to address domain shifts between the source and multiple target domains. Existing studies [12, 18, 28, 42] focus on multiple domain shifts in classification tasks. However, directly applying existing multi-target domain adaptation methods to domain-inconsistent SSOD is unfavorable, since they ignore the long-tailed class imbalance within each domain [46, 47] and class distribution shifts among domains in SSOD. Moreover, to the best of our knowledge, we are the first to explore the multi domains class distribution shifts problem in semi-supervised object detection.

Discussion. Overall, the domain-inconsistent SSOD task in this paper is different from cross-domain object detection and multi-target domain adaptation. Specifically, domain-inconsistent SSOD considers multiple unlabeled domains and seeks to simultaneously handle co-variant shifts and class distribution shifts among these domains. In contrast, cross-domain object detection only considers a single target domain, while multi-target domain adaptation ignores the class distribution shifts. Hence, domain-inconsistent SSOD is more challenging.

3 Problem Definition

As discussed in Sec. 1, existing SSOD methods [20, 22, 23, 26, 33, 37, 48] are mainly limited in the IID setting, where the labeled and unlabeled data come from the same distribution. Considering the practical situation in autonomous driving, we target a novel setting named **domain-inconsistent SSOD**, as shown in Fig. 1. Specifically, there are domain shifts in the data, including (1) the data distribution shift, and (2) the class distribution shift between the

labeled and unlabeled data. Here, data distribution shift refers to the difference in pixel-level data distribution, such as daytime versus night. On the other hand, the class distribution shift refers to the difference in the class distribution in each domain. As illustrated in Fig. 1(b), the "Daytime, Highway, Overcast" domain has very few pedestrians while the "Night, City street, Clear" domain often has more pedestrians. Meanwhile, we assume that only data from a few domains are annotated, while data from most other domains are unmarked. This is common in practical autonomous driving as the labeled data are limited and cannot cover many domains [15].

Let $\mathcal{D}_L = \{(x_i^l, y_i^l)\}_{i=1}^{N_l}$ represent the labeled set including N_l labeled samples, where $y_i^l = \{c_i^l, b_i^l\}$ denotes the label that contains both the category label and the bounding box coordinate label. Let $\mathcal{D}_U = \{\mathcal{D}_{u,1}, \mathcal{D}_{u,2}, ..., \mathcal{D}_{u,k}\}$ be the unlabeled data of k domains, which may include the labeled domain. For each domain $\mathcal{D}_{u,j}$, $j = 1,2,...,k$, $\mathcal{D}_{u,j} = \{x_i^{u,j}, j\}_{i=1}^{N_{u,j}}$ represents $N_{u,j}$ unlabeled data in this domain. There is no bounding box and class annotation for $\mathcal{D}_{u,j}$, but the domain index j is assumed to be available. The domain index is easy to collect in practical applications, which represents the acquisition environment of unlabeled images (e.g., the location is *City street*, the period is *Daytime*, and the weather is *Rainy*). Under the above domain-inconsistent SSOD setting, the data distribution shift can be denoted as $p(x^l) \neq p(x^{u,j})$, that is, the distribution of the labeled data x^l is different from that of the unlabeled data $x^{u,j}$. On the other hand, the class distribution shift is denoted as $p(y^l) \neq p(y^{u,j})$, i.e., the class distributions of each domain are different.

Challenges in domain-inconsistent SSOD. The above domain-inconsistent setting introduces two challenges for SSOD. First, the data distribution shifts would let the model trained on the labeled domain predict inaccurate pseudo-labels for the unlabeled domains with large distribution gaps. As shown in Fig. 1(a), different levels of data distribution gaps result in different precision of the predicted pseudo-labels. For instance, the model trained on the labeled images of daytime tends to make more mistakes for the unlabeled images at night, compared to the unlabeled images at dusk. Second, the class distribution shifts among domains make it difficult to obtain unbiased pseudo-labels for the unlabeled domains. Specifically, the pseudo-labels would bias to the head classes in the labeled domain. A common practice to alleviate the biased pseudo-label distribution is to match the pseudo-label distribution to the class distribution of labeled data [2, 16]. However, when the unlabeled data come from multiple different domains with different class distributions, using the class distribution of the labeled data as a reference is inappropriate.

4 Method

In this section, we propose a dual-curriculum strategy named DucTeacher to eliminate noisy pseudo-labels in the data level and class level, as shown in Fig. 2. First, we develop Domain Evolving Curriculum (DEC) in Sec. 4.1 to introduce unlabeled data from different domains progressively according to the domain similarity, which aims to avoid noisy pseudo-labels on data from hard domains. Second, we propose Distribution Matching Curriculum (DMC) in Sec. 4.2, a curriculum learning strategy to select pseudo-labels of different classes by matching the pseudo-label distribution and ground truth class distribution.

4.1 Domain Evolving Curriculum

As shown in Fig. 1(a), learning in domains with different domain similarities would produce pseudo-labels with noise at different levels, moreover, Fig. 1(a) shows the teacher model

produces unreliable pseudo-labels y_i^u for domains with low domain similarity. Based on the observation and the idea of learning easy samples first, DEC aims to learn similar domains first and learn the dissimilar domains after the model performance has been improved, which avoids the influence of noisy pseudo-labels produced in dissimilar domains.

Domain Similarity. We define the domain similarity to measure the difficulty of each unlabeled domain. Prior work [5] shows the max prediction score can be regarded as a metric of uncertainty in image classification. In this work, we propose to use the average bboxes score $S_{u,j}$ for different unlabeled domains $D_{u,j}, j = 1,2,3,...,k$ to measure the domain similarity.

$$S_{u,j} = \frac{1}{N_{u,j}} \sum_{i=1}^{N_{u,j}} \bar{f}_\theta(y_{max}|x_i^{u,j}), \tag{1}$$

where $N_{u,j}$ represents the image number in domain $\mathcal{D}_{u,j}$, $\bar{f}_\theta(y_{max}|x_i^{u,j})$ represents the average of max class probability of the predicted bboxes for each image from the unlabeled domain $x_i^{u,j}$. A domain with a high average bboxes score represents its high self-entropy and can be regarded as an easy domain for the model.

Domain Evolving Training. After computing the similarity for each unlabeled domain. To eliminate the influence of noisy pseudo-labels in dissimilar domains, DEC selects unlabeled data from domain $\mathcal{D}_{u,j}$ with the high domain similarity $S_{u,j}$ first then the dissimilar domains.

4.2 Distribution Matching Curriculum

Current SSOD algorithms [22, 23, 26, 37, 40, 48] usually select pseudo-labels with confidence larger than a fixed pre-defined threshold. However, this strategy would cause the pseudo-labels y_i^u bias to the head classes. To reduce redundant head class pseudo-labels and encourage more neglected tail class pseudo-labels, DMC tries to match the ground truth class distribution and pseudo-label distribution by adjusting the class-specific and domain-specific thresholds at each iteration.

The intuition of DMC to dynamically adjust thresholds is that if the number of pseudo-labels produced by the teacher model is more than expected, the threshold would be raised and vice versa. The ratio of pseudo-label distribution and ground truth class distribution for each class $\widetilde{p}(y_c^{u,j})/p(y_c^{u,j})$ can be regarded as an indicator to raise or reduce the threshold for class c in the domain $\mathcal{D}_{u,j}$. Moreover, because of the class distribution shifts in unlabeled domains, the proportion of each class in different domains is also different, which is shown in Fig. 1(b) (e.g. Trams appearing on highway with a lower probability). Hence, the threshold for the same class should also be different in different domains, which drives DMC to adjust thresholds $T_c^{u,j}$ for each class at the domain level:

$$T_c^{u,j} = \tau + \mu \frac{\widetilde{p}(y_c^{u,j})}{p(y_c^{u,j})}, \tag{2}$$

where $T_c^{u,j}$ represents the threshold for class c in domain $\mathcal{D}_{u,j}$, τ is the pre-defined high threshold, and μ is the scale factor. $p(y_c^{u,j})$ is the ground truth class distribution of class c, which represents the proportion of class label c among all the ground truth labels in domain $\mathcal{D}_{u,j}$. The pseudo-label distribution $\widetilde{p}(y_c^{u,j})$ is computed by accumulating the class number of model's predictions on the unlabeled data over the training course,

$$\widetilde{p}(y_c^{u,j}) = \frac{1}{N_p^{u,j}} \sum_{i=1}^{N_p^{u,j}} \mathbb{1}(f_{\theta_t}(y_c|x_i^{u,j}) > T_c^{u,j}) \cdot \mathbb{1}(c = C), \tag{3}$$

$$C = argmax(f_{\theta_t}(y|x_i^{u,j})), \tag{4}$$

where $\widetilde{p}(y_c^{u,j})$ is the cumulative pseudo-label distribution of class c in domain $\mathcal{D}_{u,j}$ and is changed at each training iteration. $\widetilde{p}(y_c^{u,j})$ can be regarded as the proportion of pseudo-labels of class c over all the class in domain $D_{u,j}$. $f_\theta(y_c|x_i^{u,j})$ is model's prediction of the unlabeled data $x_i^{u,j}$ and $N_p^{u,j}$ is the number of pseudo-labels in domain $D_{u,j}$ over the training course.

Estimating Class Distribution. However, the ground truth class distribution $p(y_c^{u,j})$ in the unlabeled domain $\mathcal{D}_{u,j}$ is unavailable and because of the class distribution shifts between different domains, as shown in Fig. 1(b), the class distribution of labeled data $p(y_c^l)$ is not an accurate estimation for that of unlabeled data. To resist the class distribution shifts, we propose a simple yet effective method to estimate the accurate class distribution for different unlabeled domains. First, by evaluating the unlabeled data with the model pre-trained on a labeled domain $\mathcal{D}_l$, we can obtain the predicted bboxes on the unlabeled data from all the domains $\mathcal{D}_{u,j}, j = 1,2,...,k$. The accurate class distribution for each unlabeled domain $\mathcal{D}_{u,j}$ is computed as

$$\hat{p}(y_c^{u,j}) = p(y_c^l) \cdot \frac{\mathcal{N}_c^{u,j}}{\mathcal{N}_c^l}, \tag{5}$$

where $\hat{p}(y_c^{u,j})$ is the estimated class distribution of unlabeled domain $\mathcal{D}_{u,j}$ for class c, $p(y_c^l)$ is the ground truth class distribution of labeled data for class c, $\mathcal{N}_c^l$ and $\mathcal{N}_c^{u,j}$ are the number of predicted bboxes on the labeled domain $\mathcal{D}_l$ and the unlabeled domain $\mathcal{D}_{u,j}$ for class c. The proposed estimation method is based on this assumption, if the model is biased, the predicted results in each domain have the same bias. The proposed estimation method in Eqn. (5) can remove the bias by division and after removing the bias of model's predictions, $\mathcal{N}_c^{u,j}/\mathcal{N}_c^l$ represents the scale ratio compared to $p(y_c^l)$. Experiments in Sec. 5.3 show the proposed estimation method can get much more accurate class distribution of the unlabeled data, and the estimated class distributions can help DMC adjust the threshold more accurately.

5 Experiments

5.1 Experimental Settings

Datasets, Baselines and Evaluation Metrics. We benchmark the proposed method on the recently proposed autonomous driving dataset SODA10M [15]. There are six class labels in SODA10M (i.e., Car, Truck, Pedestrian, Tram, Cyclist, Tricycle). It has 5,000 labeled data from a single domain (i.e., Daytime, City street, Clear) and 10M unlabeled data from 48 domains [15]. We also conduct the experiments on MS-COCO [24] following [26], which can be seen in the supplementary material. In this work, we adopt STAC [33], Unbiased Teacher [26], CSD [20], and Instant-Teaching [38] as our baseline SSOD methods. Moreover, we take the state-of-the-art cross domain object detection method UMT [9] as the baseline to show the superiority of DucTeacher to tackle the multiple domain shifts. Also, we implemented the state-of-the-art multi-target domain adaptation method MT-MTDA [28] on SODA10M. We also consider the supervised baseline where the model is trained without unlabeled data. We use $AP_{50:95}$ (denoted as mAP) as the evaluation metric, which averages the ten AP values over AP_{50} to AP_{95}.

Table 1: Comparison of mAP for different semi-supervised methods on SODA10M. The value in brackets represents the mAP improvement compared to the supervised model.

Method	mAP	AP_{50}	AP_{75}	Car	Truck	Pedestrian	Cyclist	Tram	Tricycle
Supervised-only	37.9	61.6	40.4	58.3	43.2	31.0	43.2	41.3	10.5
STAC [33]	42.8 (+ 4.9)	64.8	46.0	63.4	47.5	35.7	46.4	44.4	19.6
UMT [1]	44.7 (+ 6.8)	67.5	48.2	65.1	49.9	34.6	48.1	50.2	14.3
MT-MTDA [28]	45.2 (+ 7.3)	70.4	49.4	68.6	51.8	32.4	47.5	49.4	12.5
Unbiased Teacher [26]	46.2 (+ 8.3)	70.1	50.2	67.9	53.9	33.8	50.2	55.2	16.4
MUM [22]	45.9 (+ 8.0)	71.2	49.8	66.3	53.4	35.5	48.0	48.8	23.1
DucTeacher w/o DMC	47.3 (+ 9.4)	72.1	51.7	66.9	53.6	36.6	50.3	55.8	20.0
DucTeacher w/o DEC	48.1 (+ 10.2)	73.3	52.2	67.0	53.9	37.1	50.5	55.7	21.3
DucTeacher (ours)	**48.4** (+ 10.5)	73.5	52.4	68.7	54.3	37.9	50.9	56.6	19.0

Figure 3: Visualizations of the pseudo-labels. The green and red bboxes represent objects detected or missed by models.

5.2 Results

SODA10M. We benchmark the proposed DucTeacher on the SODA10M dataset [15, 36], compared with the supervised baseline and state-of-the-art SSOD methods [26, 33]. Table 1 shows the superiority of DucTeacher, which improves 10.5mAP over the supervised baseline. Compared with the state-of-the-art SSOD method, Unbiased Teacher, DucTeacher can also improve the performance by about 2.2 mAP. Table 1 also shows that compared to Unbiased Teacher, DucTeacher improves the AP performance for both the head class and the classes with poor performance. Specifically, DucTeacher outperforms Unbiased Teacher by 0.8 AP for head class "Car" and 4.1 AP, 2.6 AP for poor classes "Pedestrian" and "Tricycle", respectively. **COCO.** We also evaluate the proposed DucTeacher in classical SSOD setting on COCO [24] following Unbiased Teacher [26]. Table 2 shows that DucTeacher achieves state-of-the-art performance under different ratios of labeled data. Note that 1% means that 1% of the total image are labeled, and the others are unlabeled. Although DucTeacher is focusing on tackling the noise prediction problem caused by multiple domains and class distribution shifts among this, DucTeacher can also improve detection in COCO.

5.3 Ablation Studies

Domain Evolving Curriculum. To avoid the noisy pseudo-label produced on the unlabeled data with a drastic data distribution shift, DEC aims to learn unlabeled data from a similar domain first. As shown in Fig. 1(a), the domain similarity measured by the metric proposed in DEC is positively associated with the precision of pseudo-labels. Hence, we can use DEC to select the unlabeled data with a high domain similarity as

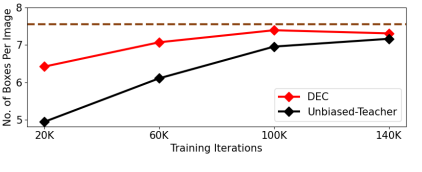


Figure 4: Number of predicted boxes per image by Unbiased Teacher and DEC.

Table 2: Comparison of mAP for different semi-supervised methods on MS-COCO under different labeled data ratios. The value in brackets represents the mAP improvement compared to the supervised model.

Method	1%	2%	5%	10%
Supervised-only	9.1	12.7	18.5	23.9
CSD [20]	10.5 (+ 1.4)	13.9 (+ 1.2)	18.6 (+ 0.1)	22.5 (- 1.4)
STAC [33]	14.0 (+ 4.9)	18.3 (+ 5.5)	24.4 (+ 5.9)	28.6 (+ 4.7)
Instant-Teaching [38]	18.0 (+ 8.9)	22.5 (+ 9.8)	26.8 (+ 8.3)	30.4 (+ 6.5)
Unbiased Teacher [26]	19.6 (+ 10.5)	23.6 (+ 10.9)	27.9 (+ 9.4)	30.9 (+ 7.0)
DucTeacher (ours)	**20.4** (+ 11.3)	**24.2** (+ 11.5)	**28.2** (+ 9.7)	**31.2** (+ 7.3)

easy unlabeled data to train first and avoid noisy pseudo-labels.

Table 1 shows that, with the help of DEC, the performance gain is 1.1 mAP compared to the Unbiased Teacher, showing the effectiveness of DEC. Moreover, Fig. 4 shows that with the help of DEC, the model would predict more instances per image, which represents DEC can make fewer False Negatives error. The False Negatives error seriously affects the self-training process since it would make the image lack annotations of some objects, which encourages the student model to predict the background class for the unmarked objects and restrains the ground truth object class. Fig. 4 also shows the number of predicted boxes per image in DEC is much more than that in Unbiased Teacher in the early training iterations. This is because the proposed DEC removes the data from dissimilar domains in the early stage, which avoids making miss detection mistakes on data from dissimilar domains and then avoids the model learning unmarked objects as a background class.

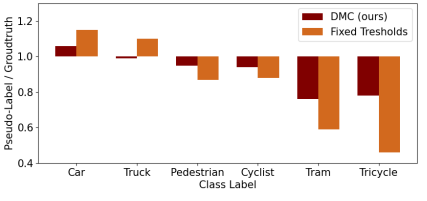


Figure 5: The ratio of pseudo-label distribution and ground truth class distribution for each class.

Distribution Matching Curriculum. DMC is proposed to introduce pseudo-labels with the cut of dynamical thresholds for each class and obtain unbiased pseudo-labels. As shown in Fig. 5, using the fixed thresholds for each class to select pseudo-labels would produce redundant head class pseudo-labels and restrain the produce of tail class pseudo-labels, which causes a high ratio of pseudo-label distribution to ground-truth class distribution for head class Car and low ratios for tail classes Tram and Tricycle. Compared with the Fixed thresholds, the ratio of pseudo-label distribution produced by DMC to ground-truth class distribution is much closer to 1, which represents DMC can inhibit the production of over-confident pseudo-labels of head class, and encourage much more neglected pseudo-labels of tail class. Table 1 also shows with the effect of DMC, DucTeacher can achieve 1.9 AP improvement compared to Unbiased Teacher, moreover, improve 4.9 AP for tail class (Tricycle). Fig. 3 shows DucTeacher can detect poor class objects neglected by Unbiased Teacher and this kind of pre-

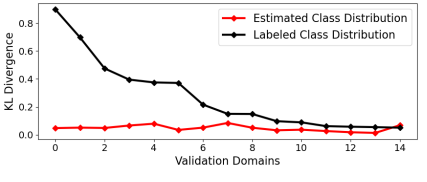


Figure 6: KL divergence between the ground truth class distribution and estimated class distribution (red line), class distribution of labeled data (black line).

dicted bboxes would further promote the learning course for poor classes in a self-training framework, which improves the recognition ability for poor classes.

Estimating Class Distribution. To rectify the pseudo-label distribution, we need to accurately estimate the class distributions of multiple unlabeled domains. To verify the effectiveness of the proposed class distribution estimation method, we estimate the class distribution of the validation set and observe the difference between estimated class distributions and ground truth class distributions. Fig. 6 shows the KL divergence between estimated class distributions and ground truth class distributions is lower than 0.2 for all the unlabeled domains. Moreover, Fig. 6 shows when existing class distribution shifts, class distribution of labeled data would have a high KL divergence with that of unlabeled domains (black line), which means regarding the class distributions of labeled data and unlabeled data as the same is inaccurate.

6 Conclusions

In this paper, we introduce a practical semi-supervised object detection setting for autonomous driving, named as Domain-Inconsistent Semi-Supervised Object Detection, where the labeled data and unlabeled data come from multiple different domains. Moreover, we point out there are two serious problems, input data distribution shifts and class distribution shifts. Furthermore, we propose DucTeacher with two curricula, DEC and DMC, to obtain accurate and unbiased pseudo-labels. Experiments show our DucTeacher achieves satisfactory performance on both the domain-inconsistent SSOD dataset SODA10M and the classical SSOD dataset COCO. Beyond DucTeacher, how to design a more effective data curriculum strategy, such as without using domain labels, is interesting. Second, extending DucTeacher to the 3D autonomous driving scene is also significant. Third, due to the large training cost of training an autonomous driving model, combining the advantages of semi-supervised object detection and continual object detection to improve both the precision and training efficiency is considerable.

Acknowledgement

We gratefully acknowledge the support of MindSpore, CANN (Compute Architecture for Neural Networks) and Ascend AI Processor used for this research.

References

[1] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[2] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019.

[3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249*, 2019.

[4] Qi Cai, Yingwei Pan, Chong-Wah Ngo, Xinmei Tian, Lingyu Duan, and Ting Yao. Exploring object relation in mean teacher for cross-domain detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11457–11466, 2019.

[5] Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. Curriculum labeling: Self-paced pseudo-labeling for semi-supervised learning. *arXiv preprint arXiv:2001.06001*, 8, 2020.

[6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3339–3348, 2018.

[7] Jaehoon Choi, Minki Jeong, Taekyung Kim, and Changick Kim. Pseudo-labeling curriculum for unsupervised domain adaptation. *arXiv preprint arXiv:1908.00262*, 2019.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee, 2009.

[9] Jinhong Deng, Wen Li, Yuhua Chen, and Lixin Duan. Unbiased mean teacher for cross-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4091–4101, 2021.

[10] Di Feng, Christian Haase-Schütz, Lars Rosenbaum, Heinz Hertlein, Claudius Glaeser, Fabian Timm, Werner Wiesbeck, and Klaus Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2020.

[11] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

[12] Behnam Gholami, Pritish Sahu, Ognjen Rudovic, Konstantinos Bousmalis, and Vladimir Pavlovic. Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Transactions on Image Processing*, 29:3993–4002, 2020.

[13] Sorin Grigorescu, Bogdan Trasnea, Tiberiu Cocias, and Gigel Macesanu. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37(3): 362–386, 2020.

[14] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *arXiv preprint arXiv:1804.06872*, 2018.

[15] Jianhua Han, Xiwen Liang, Hang Xu, Kai Chen, Lanqing Hong, Chaoqiang Ye, Wei Zhang, Zhenguo Li, Xiaodan Liang, and Chunjing Xu. Soda10m: Towards large-scale object detection benchmark for autonomous driving. *arXiv preprint arXiv:2106.11118*, 2021.

[16] Ruifei He, Jihan Yang, and Xiaojuan Qi. Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6930–6940, 2021.

[17] Han-Kai Hsu, Chun-Han Yao, Yi-Hsuan Tsai, Wei-Chih Hung, Hung-Yu Tseng, Maneesh Singh, and Ming-Hsuan Yang. Progressive domain adaptation for object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 749–757, 2020.

[18] Takashi Isobe, Xu Jia, Shuaijun Chen, Jianzhong He, Yongjie Shi, Jianzhuang Liu, Huchuan Lu, and Shengjin Wang. Multi-target domain adaptation with collaborative consistency learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8187–8196, 2021.

[19] Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger, et al. Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends® in Computer Graphics and Vision*, 12(1–3):1–308, 2020.

[20] Jisoo Jeong, Seungeui Lee, Jeesoo Kim, and Nojun Kwak. Consistency-based semi-supervised learning for object detection. *Advances in neural information processing systems*, 32:10759–10768, 2019.

[21] Lu Jiang, Deyu Meng, Qian Zhao, Shiguang Shan, and Alexander G Hauptmann. Self-paced curriculum learning. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[22] JongMok Kim, Jooyoung Jang, Seunghyeon Seo, Jisoo Jeong, Jongkeun Na, and Nojun Kwak. Mum: Mix image tiles and unmix feature tiles for semi-supervised object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14512–14521, 2022.

[23] Aoxue Li, Peng Yuan, and Zhenguo Li. Semi-supervised object detection via multi-instance alignment with global class prototypes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9809–9818, 2022.

[24] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

[25] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988, 2017.

[26] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Peter Vajda. Unbiased teacher for semi-supervised object detection. *arXiv preprint arXiv:2102.09480*, 2021.

[27] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

[28] Le Thanh Nguyen-Meidine, Atif Belal, Madhu Kiran, Jose Dolz, Louis-Antoine Blais-Morin, and Eric Granger. Unsupervised multi-target domain adaptation through knowledge distillation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1339–1347, 2021.

[29] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE transactions on neural networks*, 22(2):199–210, 2010.

[30] Scott Drew Pendleton, Hans Andersen, Xinxin Du, Xiaotong Shen, Malika Meghjani, You Hong Eng, Daniela Rus, and Marcelo H Ang. Perception, planning, control, and coordination for autonomous vehicles. *Machines*, 5(1):6, 2017.

[31] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6956–6965, 2019.

[32] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020.

[33] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A simple semi-supervised learning framework for object detection. *arXiv preprint arXiv:2005.04757*, 2020.

[34] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780*, 2017.

[35] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7167–7176, 2017.

[36] Liyuan Wang, Xingxing Zhang, Kuo Yang, Longhui Yu, Chongxuan Li, Lanqing Hong, Shifeng Zhang, Zhenguo Li, Yi Zhong, and Jun Zhu. Memory replay with data compression for continual learning. *arXiv preprint arXiv:2202.06592*, 2022.

[37] Zhenyu Wang, Ya-Li Li, Ye Guo, and Shengjin Wang. Combating noise: Semi-supervised learning by region uncertainty quantification. *Advances in Neural Information Processing Systems*, 34:9534–9545, 2021.

[38] Xiaoxia Wu, Ethan Dyer, and Behnam Neyshabur. When do curricula work? *arXiv preprint arXiv:2012.03107*, 2020.

[39] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V Le. Unsupervised data augmentation for consistency training. *arXiv preprint arXiv:1904.12848*, 2019.

[40] Qize Yang, Xihan Wei, Biao Wang, Xian-Sheng Hua, and Lei Zhang. Interactive self-training with mean teachers for semi-supervised object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5941–5950, 2021.

[41] Fuxun Yu, Di Wang, Yinpeng Chen, Nikolaos Karianakis, Tong Shen, Pei Yu, Dimitrios Lymberopoulos, Sidi Lu, Weisong Shi, and Xiang Chen. Unsupervised domain adaptation for object detection via cross-domain semi-supervised learning. *arXiv preprint arXiv:1911.07158*, 2019.

[42] Huanhuan Yu, Menglei Hu, and Songcan Chen. Multi-target unsupervised domain adaptation without exactly shared categories. *arXiv preprint arXiv:1809.00852*, 2018.

[43] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *arXiv preprint arXiv:2110.08263*, 2021.

[44] Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2020–2030, 2017.

[45] Yifan Zhang, Ying Wei, Qingyao Wu, Peilin Zhao, Shuaicheng Niu, Junzhou Huang, and Mingkui Tan. Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Transactions on Image Processing*, 29:7834–7844, 2020.

[46] Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. Deep long-tailed learning: A survey. *arXiv preprint arXiv:2110.04596*, 2021.

[47] Yifan Zhang, Bryan Hooi, Lanqing Hong, and Jiashi Feng. Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition. In *Advances in Neural Information Processing Systems*, 2022.

[48] Qiang Zhou, Chaohui Yu, Zhibin Wang, Qi Qian, and Hao Li. Instant-teaching: An end-to-end semi-supervised object detection framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4081–4090, 2021.

[49] Xinge Zhu, Jiangmiao Pang, Ceyuan Yang, Jianping Shi, and Dahua Lin. Adapting object detectors via selective cross-domain alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 687–696, 2019.