Instance-Distance Active Learning for Source-Free Cross-Domain Object Detection

Kangrui Du, Yujun Qian, Juepeng Zheng✉
School of Artificial Intelligence, Sun Yat-Sen University
{dukr6, qianyj8}@mail2.sysu.edu.cn, zhengjp8@mail.sysu.edu.cn

2025 IEEE International Conference on Multimedia and Expo (ICME) | 979-8-3315-9495-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICME59968.2025.11208931

***Abstract*—*Source-Free Domain Adaptive Object Detection* (SFDA-OD) aims to apply a pre-trained detector from the source domain to unlabeled data in the target domain. Existing methods typically utilize techniques such as domain alignment and fine-tuning model parameters to achieve satisfactory performance in the target domain. However, even using SFDA approaches, the accuracy of the model on the target domain remains significantly accuracy gap compared to the fully-supervised model. In this paper, we explore a new scenario *Source-Free Active Domain Adaptive Object Detection* (SFADA-OD) to better enhance model adaptability and performance within limited annotation, and propose a novel approach named *Learning Domain Distance for Active Pick* (LDDAP). Firstly, we design graph-aware distance learning to calculate the distance between the encoded images and the predicted instances, which better evaluate the distance between domains in SFDA scenario. Secondly, we propose instance-aware active sampling to facilitate instance-level distance associated with other indexes to select the most valuable images, thereby maximizing the effectiveness of the limited labeled information. The results demonstrate a significant enhancement in performance over baseline cross-domain models and other Active Learning related methods on several well-known public datasets.In addition, our model even surpasses fully-supervised results on the dataset Pascal → Watercolor. Code is avaliable at LDDAP**

***Index Terms*—domain, distance, object detection, active learning**

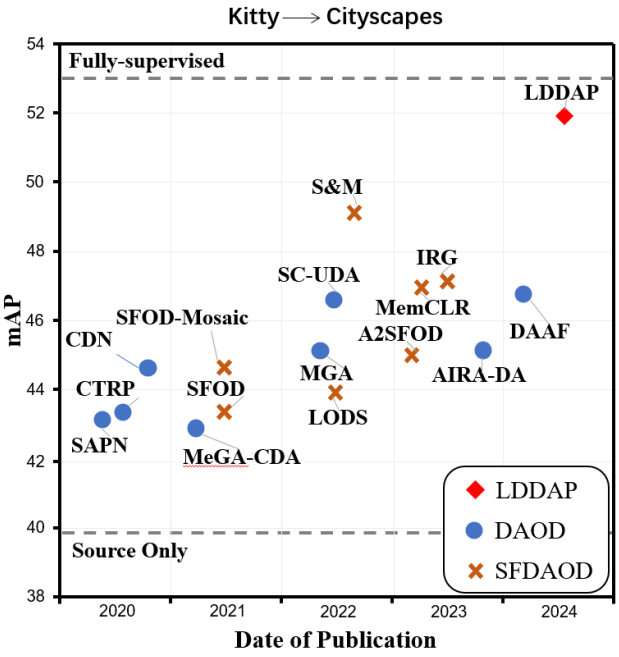

Fig. 1. The performance comparison of domain adaptation methods in object detection.

I. Introduction

Object detection, a critical task in computer vision, involves identifying and localizing objects within an image [1]–[3]. Deep learning has significantly advanced the field, enabling models to achieve remarkable performance on various object detection tasks. However, challenges arise when the training and test data distributions differ. This challenge is addressed by *Domain Adaptive Object Detection* (DA-OD) [4], which aims to mitigate performance degradation when applying a model trained on one domain (source) to another related but distinct domain (target). In many practical scenarios, accessing the source domain data during deployment can pose significant risks or breaches of privacy, especially when sensitive information is involved, *e.g.,* an autonomous driving system where the training data consists of video recordings from urban environments that include identifiable individuals and vehicles. This has led to the development of *Source-Free Domain Adaptive Object Detection* (SFDA-OD) approaches, which adapt models to new target domains without requiring access to the source domain data, thereby ensuring privacy and data security.

In Figure 1, lots of DA-OD and SFDA-OD methods have been proposed to address domain shift issues, which have made considerable improvement compared to the Source Only. However, both DA-OD and SFDA-OD methodologies still fall short of the performance benchmark by the supervised approach. As shown in Figure 1, there still exists a large accuracy gap compared to fully-supervised method (at least 3.6% for KITTY → Cityscapes), which still faces difficulties in real-world applications. To further enhance model adaptability and performance, we integrate an active learning method to this scenario, named *Source-Free Active Domain Adaptation for Object Detection* (SFADA-OD), which helps us to be closer to the fully-supervised way within limited annotation.

However, in this newly explored scenario, two specific challenges arise that need to be addressed: (1) **Difficulty in evaluating the distance between different domains for SFDA**. Traditional domain adaptation methods exhibit significant deficiencies and challenges in handling disparate feature spaces, preserving structural information, enhancing robustness, and reducing computational complexity. These limitations hinder their ability to effectively measure distances between different domains. (2) **Inadequate consideration**

of instance-level measurement for SFADA-OD. Existing Active Learning methods deployed in ADA or SFADA focus on image-level active sampling and *do not* consider instance-level measurement metric in cross-domain object detection scenarios.

In this paper, we propose *Learning Domain Distance for Active Pick* (LDDAP), which integrates *Active Learning* (AL) with *Domain Adaptation* (DA) within a source-free object detection context. To better evaluate the distance between domains in SFDA scenario, we design graph-aware distance learning (Sec. III-A) to calculate the distance between the encoded images and the predicted instances in both models using Gromov-Wasserstein graph matching algorithm. Moreover, to tackle the problem of inadequate consideration of instance-level measurement in SFADA-OD, we propose instance-aware active sampling (Sec. III-B) to facilitate instance-level distance associated with other indexes to select the most valuable images, thereby maximizing the effectiveness of the limited labeled information.

To this end, our contributions in this paper could be summarized as follows: (1) We are the first to explore *Source-Free Active Domain Adaptation for Object Detection* (SFADA-OD) scenario. While ADA and SFADA have been explored in other tasks, our approach uniquely addresses significant accuracy gaps in current research within the field of object detection. (2) We propose a novel approach named *Learning Domain Distance for Active Pick* (LDDAP) to address two main challenges in SFADA-OD scenario, which contains graph-aware distance learning and instance-aware active sampling. (3) Our proposed LDDAP demonstrates superior performance, achieving state-of-the-art results on several well-known public datasets. This underscores the effectiveness of LDDAP in real-world scenarios.

II. Related Work

A. DA-OD & SFDA-OD

Domain Adaptive Object Detection (DA-OD) has been extensively studied in the scenario of object detection, aiming to mitigate performance degradation when models are applied to target domains different from their training (source) domains (*e.g.,* Afan [5],). Traditional DA methods leverage access to both source and target domain data to align their distributions and improve detection accuracy in the target domain (*e.g.,* GPDA [6]). However, in many practical scenarios, accessing source domain data during deployment can pose significant risks or breaches of privacy.

To address these challenges, *Source-Free Domain Adaptive Object Detection* (SFDA-OD) methods have been developed. SFDA focuses on adapting models to new target domains without requiring access to the source domain data, thereby ensuring privacy and data security [7]. Existing SFDA-OD methods often minimize the distance between domain distributions using statistical measures to find domain invariance features (*e.g.,* JS divergence [8], Wasserstein distance [9]). These methods, however, face numerous deficiencies and challenges, including difficulties in managing disparate feature spaces, maintaining structural information, enhancing robustness, and reducing computational complexity. These limitations impede their capability to effectively measure distances between different domains. To overcome this limitation, We develop a graph-aware distance learning approach to compute the distance between encoded images and predicted instances in both models, utilizing the Gromov-Wasserstein graph matching algorithm [10].

B. Active DA

To further enhance the performance and moderate the gap with fully supervised methods, incorporating *Active Learning* (AL) into DA and SFDA for object detection could be pivotal. AL aims to enhance model performance by selectively labeling the most informative data points from the target domain [11]. Existing work on combining Active Learning with Domain Adaptation (DA) or Source-Free Domain Adaptation (SFDA) has primarily focused on areas such as image classification (*e.g.,* CLUE [12]) and semantic segmentation (*e.g.,* SALAD [13]), with limited attention to *Source-Free Active Domain Adaptation for Object Detection* (SFADA-OD). Moreover, Current ADA and SFADA methods primarily focus on image-level sampling and overlook instance-level metrics in cross-domain scenarios [14]. To address this issue, we adopt instance-aware active sampling to incorporate instance-level distances along with other metrics, enabling the selection of the most valuable images and thus maximizing the effectiveness of limited labeled data [15].

III. Method

Overview: The overall framework of LDDAP proposed by us is illustrated in Figure 2. As depicted in the diagram, the essence of our approach lies in how to extract relevant metrics from the Teacher-Student structure. We start by initializing a limited set of labels. After each iteration, the trained Teacher network generates pseudo-labels for all unlabeled images, which are then used to compute metrics such as image difficulty, information, diversity and distance. Meanwhile, the Student network is employed to generate style-enhanced image encodings and instance encodings, and to calculate the distances between the image encodings and instance encodings generated by the Teacher network and their corresponding original images. Based on these metrics, our model conducts active data augmentation. In the following sections, we will provide detailed insights into the application of LDDAP in semi-supervised domain adaptation and active learning sampling.

A. Graph-aware distance learning

1) Unsupervised learning: For domain adaptation, unsupervised learning is used to achieve domain transfer through mutual learning between the student and teacher networks [17], [18]. To cost-effectively enhance network accuracy, active learning is utilized to select a small yet effective subset of labeled target domain images. A critical aspect of unsupervised learning is the estimation of the distance between the two domains using an appropriate algorithm. In this context, the

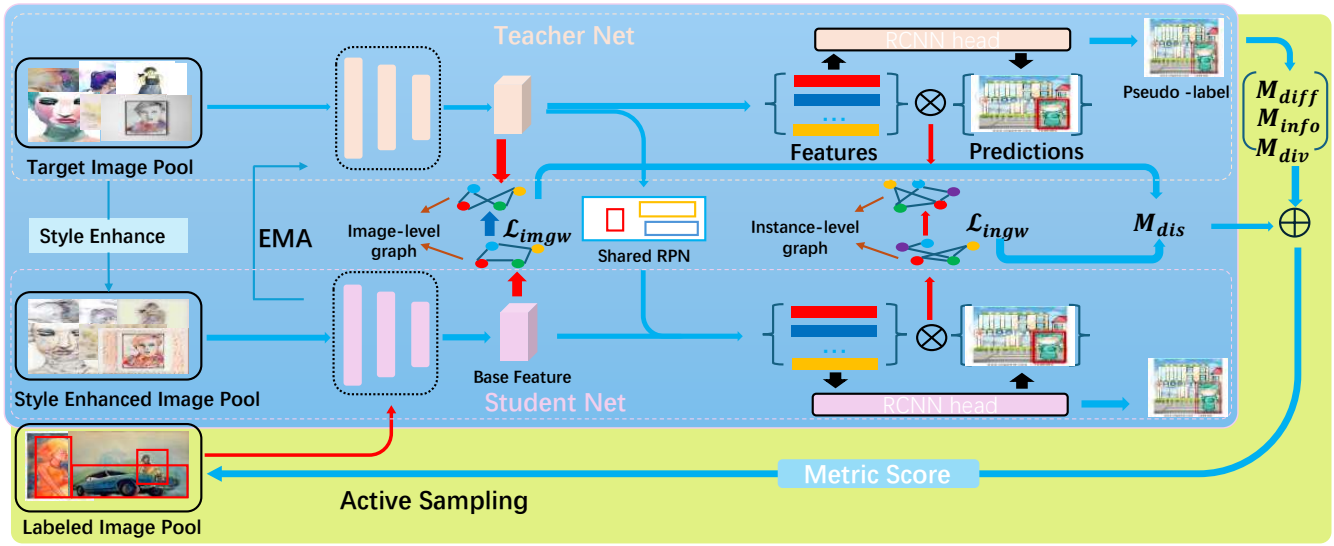

Fig. 2. We propose an overview of the *Learning Domain Distance for Active Pick* (LDDAP) method. This method employs a student-teacher architecture to train domain adaptive modules. The scores generated by the teacher network, along with the cross-domain graph distances, are used to determine the samples required for active learning. The student network undergoes supervised and unsupervised learning through gradient descent to update its parameters. These updated parameters are then transferred to the teacher network using *Exponential Moving Average* (EMA) [16] adjustments.

Gromov Wasserstein distance is employed for this estimation, with experimental results demonstrating its high effectiveness.

In the unsupervised segment, aligning domain styles between the source and target domains becomes crucial. Considering the image-level encodings generated after passing through the network, along with the features generated through combining these encodings with RPN, and various instance and prediction values, which named instance-level encodings, combine the distances that effectively reflect domain gaps in vector space.

Image-level Distance. For a target domain image x and its corresponding style-enhanced image$\hat{x}$, we extract basic features after the 30-th network layers and transform them into two graph vectors, where V and $\hat{V}$ represents nodes, denoting features, C and $\hat{C}$ represents edges, denoting distances between features. Here, we use cosine similarity to calculate. Based on *Gromov Wasserstein* discrepancy [9], we obtain D_{imgw} through the following equation:

$$D_{\text{imgw}} = \sum_{i,j,m,n} L(C_{i,j}, \hat{C}_{m,n}) T_{i,m} T_{j,n} \tag{1}$$

where $L(\cdot,\cdot)$ is the Kullback-Leibler divergence to measure the distance of the edges across graph, C and $\hat{C}$ here represent the similarity matrices like cosine similarity matrices between these features.. Since we input the same image, we expect to have identical features in similar regions. So we construct march matrix $T = I$, I is the identity matrix.

Instance-level Distance To enhance the discrimination ability, we utilize instance-level features as described below.

$$\tilde{f}_x = f_x \odot p, \quad \tilde{f}_{\hat{x}} = \hat{f}_x \odot \hat{p} \tag{2}$$

Here, $\tilde{f}_x \in \mathbb{R}^{R\times(C'\cdot N_c)}$ and $\tilde{f}_{\hat{x}} \in \mathbb{R}^{R\times(C'\cdot N_c)}$ are obtained through a multilinear transformation $\odot$ of the predictions and the instance-level features.

When aligning instances, we encounter the challenge that identical features may manifest in various regions, which precludes the use of the identity matrix I as a direct substitute for T. Given the inherent noise in instance features, a straightforward alignment of all features is not feasible; instead, noise mitigation is imperative. To address this, we propose the creation of a class relationship mask M designed to sift out noise and features with low confidence. For the purpose of assigning pseudo-labels to each region proposal, we introduce a confidence threshold h, governed by the following criteria:

$$l_r = \begin{cases} \arg\max_c p_{r,c} & \text{if} \quad \max_c p_{r,c} \geq h, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where l_r represents the label for the r-th region, and 0 indicates the background. h is a hyperparameter. This method filters out unreliable labels. Next, we define M as follows:

$$M_{i,m} = \begin{cases} 1 & \text{if } l_i = l_m \text{ and } l_i \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

By enforcing category consistency, M not only filters out low-confidence features but also reduces redundant alignment between features. Then, we construct a distance similarity matrix Γ using cosine similarity, measuring the similarity between different features based on the predictions of the teacher model. Finally, we obtain graph matching matrix as follows,

$$T = I + \beta M \otimes \Gamma \tag{5}$$

where $\otimes$ represents element-wise multiplication. β is a hyperparameter. The first term I is an identity matrix since there is a correspondence between the features $\tilde{f}_x$ and $\tilde{f}_{\hat{x}}$ in the same region. The second term filters out noise from Γ through M enhances the matching between features with the same category.

For L_{unsup}, we use the D computed as Eq. (1), and add a cross-entropy loss L_{ce} calculated using the pseudo-labels from the teacher network as the ground-truth for the student network, to improve the student model's discrimination ability. The calculation of L_{unsup} is as follows:

$$L_{\text{unsup}} = L_{\text{ce}} + \gamma L_{\text{imgw}} + \lambda L_{\text{ingw}} \tag{6}$$

Where $L_{\text{imgw}} = D_{\text{imgw}}, L_{\text{ingw}} = D_{\text{ingw}}$, λ and γ are hyperparameters used to balance the loss components, fixed at 0.1.

2) Supervised learning: In this paper, we choose Faster-RCNN based on VGG16 and ResNet50 as our fully supervised detectors. We employ appropriate active sampling strategies (discussed later) to select images requiring real labels with richer information, and then utilize these real labels for training. Given an unlabeled target domain image, we first conduct a epoch of unsupervised domain adaptation to obtain corresponding metrics, selecting the most informative images to add to the labeled data pool. When the labeled data pool is not empty, the student model transitions to semi-supervised learning.

To update the parameters of the teacher model, we employ the *EMA* [16] of the student model's parameters, with the hyperparameter η set to 0.999. This approach will guide the teacher detector towards cross-domain style training.

B. Instance-aware active sampling

After a round of training in LDDAP , all images' relevant metrics are obtained, allowing us to assess which images possess higher information gain. Our evaluation metrics include Difficulty, Information, Diversity, and Cross-domain Distance.

Difficulty. Difficulty often utilized in semi-supervised learning to measure information gain, is computed through entropy calculation on predicted probability values. A higher value indicates a more challenging instance to detect. The formula is as follows:

$$M_{\text{i}}^{\text{diff}} = -\frac{1}{n_b^i} \sum_{j=1}^{n_b^i} \sum_{k=1}^{N_c} p(c_k; b_j, \theta_t) \log p(c_k; b_j, \theta_t) \tag{7}$$

where n_i is the number of predicted bounding boxes after non-maximum suppression and confidence filtering, N_c is the number of object categories, and $p(c_k; b_j, \theta_t)$ is the probability predicted by the teacher network for the k-th category on the bounding box b_j. Afterwards, we can determine whether the image is difficult for cross-domain adaptation by M^{diff} .

Information. Information is another critical metric. Unlike Difficulty, richer information in object detection implies more visual concepts present in the image, facilitating the learning of additional detection patterns. The formula is as follows.

$$M_{\text{i}}^{\text{info}} = \sum_{i=1}^{n_i^b} \text{confidence}(b_j, \theta) \tag{8}$$

where confidence(b_j, θ_t) represents the highest confidence score in the j-th bounding box predicted by the teacher network. From equation (8), we can see that the larger M^{info}, the more instances recognized by the teacher network, indicating that the image contains richer information.

Diversity. Diversity measures the number of categories within an image, evaluated by the number of categories remaining after *non-maximum suppression*(NMS) by a teacher network. It is defined as:

$$M_{\text{i}}^{\text{dive}} = \left| \{c_j\}_{j=1}^{n_b^i} \right| \tag{9}$$

where c_j represents the predicted category of the j-th bounding box, and $|\cdot|$ denotes the cardinality. More categories imply more visual concepts in this image.

Distance. Distance signifies the distance between the image in the target domain and the source domain. As our model aims to address domain shift issues, distance serves as a crucial indicator of domain alignment. The distance score M^{dis} is calculated by

$$M_{\text{i}}^{\text{dis}} = D_{\text{imgw}} + \lambda D_{\text{ingw}} \tag{10}$$

where D_{imgw} and D_{ingw} is calculated by Eq.(1). Since we believe that these two distances are equally important for domain transfer, we set λ equal to 1. A higher value indicates the image contains more abundant information across different domains, promoting alignment from the source domain towards the target domain.

Metric Fusion. With various metrics at hand, a rational combination is necessary to derive a reasonable score for image selection. Different metrics, however, may operate in different vector spaces with varying scales, potentially affecting the importance of each metric during evaluation. Thus, to mitigate the impact of different scales, we employed batch normalization with the formula

$$\hat{M}_i^m = \frac{M_i^m}{M_{\text{max}}^m} \tag{11}$$

where $m \in \{difficulty, information, diversity, distance\}$ represent the metrics,and the M_{max}^m is the maximum value of the metric.

After normalization, we have four samples in the same vector space. Then we utilize Lp regularization to combine these four samples into a single score, which is obtained by

$$M_{L_p} = \left(\sum_{i=1}^{n} w_i |\hat{M}_i|^p \right)^{1/p} \tag{12}$$

where n represent the image in unlabeled image pool, with each image corresponding to one score, and w_i are hyperparameters stand for the weights of the 4 metrics. Based on experience, we combine these three metrics using the $L1$ norm. Subsequently, we select images in descending order based on these scores. According to empirical findings, annotating 1% of the total samples per epoch has yielded outstanding results, and we finally select 10% of total images after 10 epoched.

IV. Experiments

Datasets. In the **Pascal→Clipart** transfer scenario, **Pascal1** [19] consists of 20 categories of natural images. Similarly, **Clipart** [20] contains the same 20 categories and $1K$ clipart-style images. We use around $15K$ images from the PASCAL

VOC 2007 and 2012 training and validation sets to pre-train the source model. In the **Pascal** →**Watercolor** scenario, the **Watercolor** [20] dataset consists of $2K$ watercolor-style images and shares 6 categories with Pascal. Following prior works [21], we use the training and testing images to train and evaluate our model accordingly. For the **KITTI** → **Cityscapes** transfer scenario, **KITTI** [22] includes $7,481$ urban images distinct from Cityscapes. Captured under normal weather, **Cityscapes** [23], consisting of $2,975$ training images and 500 testing images, have a total of 8 categories. Following standard settings [24], we focus on detecting the car category and pre-train the source model using all available data.

Implementation details. To ensure fairness, we follow the experimental setup of [24], using Faster R-CNN as the base detector. First, we employ the learning rate at 0.0001 and employ the SGD optimizer to train the *Graph-aware distance learning* module. Secondary, we set ω_1 =0.1, ω_2 =0.1, ω_3=1, ω_4=1 by default in combining *Instance-aware active sampling* . During testing, we report the mean average precision (mAP) at an IoU threshold of 0.5.

A. Comparison with state-of-the-art methods

We compare our LDDAP with state-of-the-art SFOD and UDAOD methods. SFOD methods include SFOD [24], which focus on domain adaptation through data augmentation techniques. UDAOD methods are limited to CTRP [25], which uses collaborative training between region proposal localization and classification and so on.

TABLE I
DETECTION RESULTS ON **PASCAL** → **WATERCOLOR**.

Methods	Bike	Bird	Car	Cat	Dog	Person	mAP
Source Only	85.6	46.8	43.1	24.5	21.9	54.8	46.1
VDD [26]	90.0	56.6	49.2	39.5	38.8	65.3	56.6
PD [27]	95.8	54.3	48.3	42.4	35.1	65.8	56.9
UMT [28]	88.2	55.3	51.7	39.8	43.6	69.9	58.1
SFOD [24]	34.6	46.7	26.8	23.2	34.9	33.5	39.3
SOAP [29]	30.9	51.8	27.2	28.0	31.4	59.0	34.2
IRG [30]	38.5	73.4	34.4	33.2	49.2	62.3	54.1
LODS [31]	43.1	61.4	40.1	36.8	48.2	45.8	48.3
Random	68.5	57.3	56.0	51.7	52.6	75.4	60.2
LC	59.1	59.1	57.1	53.3	51.7	75.6	63.9
Entropy	87.6	71.9	65.3	49.3	55.0	78.0	67.8
MS	82.0	75.6	**72.7**	63.8	55.3	71.5	70.1
LDDAP (Ours)	**100**	**75.7**	67.5	**66.8**	**65.0**	**79.5**	**75.7**
Fully-Supervised	39.1	65.0	34.8	59.1	72.1	78.4	58.1

Pascal → **Watercolor**. In this scenario, we adapt the detector from real images to watercolor-style images. Table I shows that our method, LDDAP, achieves state-of-the-art performance with a mAP of 75.7% after adaptation, despite the large domain shift. Compared to the SFOD method, our performance improves by 21.6% at best, demonstrating the strong applicability of our method across different image styles.

KITTI → **Cityscapes**. In this scenario, we assess the adaptation performance of our method across different cameras, as shown in Table II. As shown in Table II, our method, LDDAP, achieves 52.7% in this adaptation scenario, performing comparably to many recent methods that can access source data. Compared to other SFOD methods, our method also achieves better performance.

TABLE II
DETECTION RESULTS ON **KITTI** → **CITYSCAPES**.

Methods	AP on car	Methods	AP on car
Source Only	39.2	AIRA-DA [32]	45.2
CTRP [25]	43.6	SC-UDA [33]	46.4
CDN [34]	44.9	DAAF [35]	46.7
SFOD [24]	43.6	S&M [36]	49.7
IRG [30]	46.9	RPL [37]	47.8
Random	47.7	Entropy	48.1
LC	39.1	MC	48.0
LDDAP (Ours)	**49.8**	Fully-Supervised	53.3

Pascal → **Clipart**. In this scenario, we transfer the object detector from real images to clipart-style images, addressing a significant domain shift. Our proposed method, LDDAP, achieves state-of-the-art performance with a mAP of 63.6%, an improvement of 18.4% over the SFOD method (from 45.2% to 63.6%). Compared to the state-of-the-art methods that can access target data, our approach boosts the mAP by 19.5% (from 44.1% to 63.6%). Due to space constraints, the table and our qualitative comparison will be presented in the supplementary materials.

B. Ablation Study

We conducted an ablation study in the **Pascal**→**Clipart** transfer scenario. We compared the effects of different metrics in our supplementary materials. From the table, it can be seen that the four metrics, namely *difficulty, information, diversity, and distance*, are all beneficial to the domain transfer results. Moreover, linearly combining them can achieve a better result, which strongly demonstrates their effectiveness and helps the object detector perform better across domains.

V. CONCLUSION

We propose a new active learning strategy for cross-domain style transfer based on a teacher-student model, called *Learning Domain Distance for Active Pick* (LDDAP). This strategy selects the most valuable labeled images to enable the target detector to adapt more quickly to new scenes. Based on this strategy, we use an efficient method to calculate the inter-domain distance and study sample selection from the perspectives of *difficulty, information, diversity, and distance*. Experimental results demonstrate the superior performance of LDDAP, indicating that it effectively improves the performance of the baseline network while reducing labeling costs, even surpassing fully-supervised methods, confirming the effectiveness of our approach. Furthermore, quantitative and qualitative analyses provide useful insights for data annotation in practical applications.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (42401415) and Jiangsu Innovation Capacity Building Program (Project BM2022028).

REFERENCES

[1] Juepeng Zheng, Haohuan Fu, Weijia Li, et al., "Growing status observation for oil palm trees using unmanned aerial vehicle (uav) images," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 95–121, 2021.

[2] Juepeng Zheng, Shuai Yuan, Wenzhao Wu, Weijia Li, et al., "Surveying coconut trees using high-resolution satellite imagery in remote atolls of the pacific ocean," *Remote Sens. of Environ.*, vol. 287, pp. 113485, 2023.

[3] Juepeng Zheng, Shuai Yuan, Weijia Li, et al., "A review of individual tree crown detection and delineation from optical remote sensing images: Current progress and future," *IEEE Geosci. Remote Sens. Mag.*, 2024.

[4] Juepeng Zheng, Haohuan Fu, et al., "Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 154–177, 2020.

[5] Hongsong Wang, Shengcai Liao, and Ling Shao, "Afan: Augmented feature alignment network for cross-domain object detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 4046–4056, 2021.

[6] Jing Sun, Zhihui Wang, Wei Wang, Haojie Li, and Fuming Sun, "Domain adaptation with geometrical preservation and distribution alignment," *Neurocomputing*, vol. 454, pp. 152–167, 2021.

[7] Jogendra Nath Kundu, Naveen Venkat, et al., "Universal source-free domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4544–4553.

[8] Mehran Khodabandeh, Arash Vahdat, Mani Ranjbar, and William G Macready, "A robust learning approach to domain adaptive object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 480–490.

[9] Gabriel Peyré, Marco Cuturi, and Justin M. Solomon, "Gromov-wasserstein averaging of kernel and distance matrices," in *International Conference on Machine Learning*, 2016.

[10] Hongteng Xu, Dixin Luo, Hongyuan Zha, et al., "Gromov-wasserstein learning for graph matching and node embedding," in *International conference on machine learning*. PMLR, 2019, pp. 6932–6941.

[11] Richard M Felder and Rebecca Brent, "Active learning: An introduction," *ASQ higher education brief*, vol. 2, no. 4, pp. 1–5, 2009.

[12] Viraj Prabhu, Arjun Chandrasekaran, Kate Saenko, and Judy Hoffman, "Active domain adaptation via clustering uncertainty-weighted embeddings," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8505–8514.

[13] Divya Kothandaraman, Sumit Shekhar, Abhilasha Sancheti, Manoj Ghuhan, Tripti Shukla, and Dinesh Manocha, "Salad: Source-free active label-agnostic domain adaptation for classification, segmentation and detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 382–391.

[14] Munan Ning, Donghuan Lu, Yujia Xie, Dongdong Chen, Dong Wei, Yefeng Zheng, Yonghong Tian, Shuicheng Yan, and Li Yuan, "Madav2: Advanced multi-anchor based active domain adaptation segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[15] Zhongzheng Ren, Zhiding Yu, Xiaodong Yang, Ming-Yu Liu, Yong Jae Lee, Alexander G Schwing, and Jan Kautz, "Instance-aware, context-focused, and memory-efficient weakly supervised object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10598–10607.

[16] Antti Tarvainen and Harri Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," 2018.

[17] Qingmei Li, Yibin Wen, et al., "Hyunida: Breaking label set constraints for universal domain adaptation in cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 2024.

[18] Wentang Chen, Yibin Wen, et al., "Ban: A universal paradigm for cross-scene classification under noisy annotations from rgb and hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, 2025.

[19] Mark Everingham and John Winn, "The pascal visual object classes challenge 2012 (voc2012) development kit," *Pattern Anal. Stat. Model. Comput. Learn., Tech. Rep*, vol. 2007, no. 1-45, pp. 5, 2012.

[20] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa, "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5001–5009.

[21] 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*, 2019, pp. 6956–6965.

[22] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[23] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[24] Xianfeng Li, Weijie Chen, Di Xie, Shicai Yang, Peng Yuan, Shiliang Pu, and Yueting Zhuang, "A free lunch for unsupervised domain adaptive object detection without source data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 8474–8481.

[25] Ganlong Zhao, Guanbin Li, Ruijia Xu, and Liang Lin, "Collaborative training between region proposal localization and classification for domain adaptive object detection," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 2020, pp. 86–102.

[26] Aming Wu, Rui Liu, Yahong Han, Linchao Zhu, and Yi Yang, "Vector-decomposed disentanglement for domain-invariant object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9342–9351.

[27] Aming Wu, Yahong Han, Linchao Zhu, and Yi Yang, "Instance-invariant domain adaptive object detection via progressive disentanglement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4178–4193, 2021.

[28] 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*, 2021, pp. 4091–4101.

[29] Lin Xiong, Mao Ye, Dan Zhang, Yan Gan, Xue Li, and Yingying Zhu, "Source data-free domain adaptation of object detector through domain-specific perturbation," *International Journal of Intelligent Systems*, vol. 36, no. 8, pp. 3746–3766, 2021.

[30] Vibashan VS, Poojan Oza, and Vishal M Patel, "Instance relation graph guided source-free domain adaptive object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3520–3530.

[31] Shuaifeng Li, Mao Ye, Xiatian Zhu, Lihua Zhou, and Lin Xiong, "Source-free object detection by learning to overlook domain style," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8014–8023.

[32] Kunyang Sun, Wei Lin, Haoqin Shi, Yu Liu, Zhengming Zhang, Yongming Huang, and Horst Bischof, "Aira-da: Adversarial image reconstruction alignments for unsupervised domain adaptive object detection," *IEEE Robotics and Automation Letters*, 2023.

[33] Fuxun Yu, Di Wang, Yinpeng Chen, Nikolaos Karianakis, Tong Shen, Pei Yu, Dimitrios Lymberopoulos, Sidi Lu, Weisong Shi, and Xiang Chen, "Sc-uda: Style and content gaps aware unsupervised domain adaptation for object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 382–391.

[34] Peng Su, Kun Wang, Xingyu Zeng, Shixiang Tang, Dapeng Chen, Di Qiu, and Xiaogang Wang, "Adapting object detectors with conditional domain normalization," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. Springer, 2020, pp. 403–419.

[35] Xiaoyong Yu et al., "Domain adaptation of anchor-free object detection for urban traffic," *Neurocomputing*, vol. 582, pp. 127477, 2024.

[36] Peng Yuan, Weijie Chen, Shicai Yang, Yunyi Xuan, Di Xie, Yueting Zhuang, and Shiliang Pu, "Simulation-and-mining: Towards accurate source-free unsupervised domain adaptive object detection," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3843–3847.

[37] Siqi Zhang, Lu Zhang, and Zhiyong Liu, "Refined pseudo labeling for source-free domain adaptive object detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.