

Dual-Rate Dynamic Teacher for Source-Free Domain Adaptive Object Detection

Qi He^{1,2}, Xiao Wu^{1,2*}, Jun-Yan He³, Shuai Li⁴

¹ Southwest Jiaotong University, Chengdu, China

² Engineering Research Center of Sustainable Urban Intelligent Transportation, Chengdu, China

³ Meituan Inc., Shenzhen, China, ⁴ The Hong Kong Polytechnic University, Hong Kong, China

{qihe96, wuxiaohk, junyanhe1989}@gmail.com, novak.li@connect.polyu.hk

Abstract

Source-Free Domain Adaptive Object Detection transfers knowledge from a labeled source domain to an unlabeled target domain while preserving data privacy by restricting access to source data during adaptation. Existing approaches predominantly leverage the Mean Teacher framework for self-training in the target domain. The exponential moving average (EMA) mechanism in the Mean Teacher stabilizes the training by averaging the student weights over training steps. However, in domain adaptation, its inherent lag in responding to emerging knowledge can hinder the rapid adaptation of the student to target-domain shifts. To address this challenge, Dual-rate Dynamic Teacher (DDT) with Asynchronous EMA (AEMA) is proposed, which implements group-wise parameter updates. In contrast to traditional EMA, which simultaneously updates all parameters, AEMA dynamically decomposes teacher parameters into two functional groups based on their contributions to capture the domain shift. By applying a distinct smoothing coefficient to two groups, AEMA simultaneously enables fast adaptation and historical knowledge retention. Comprehensive experiments carried out on three widely used traffic benchmarks have demonstrated that the proposed DDT achieves superior performance, outperforming SOTA methods by a clear margin. The codes are available at <https://github.com/qih96/DDT>.

1. Introduction

Object detection is a critical component of computer vision, significantly contributing to a myriad of tasks [14, 24, 46]. Large-scale labeled datasets [20, 27] serve as the foundation for successful object detection. However, due to the domain gap between the training and testing domains, deploying the trained detector to the unseen domain often leads to performance degradation. Moreover, annotating large

*Corresponding author.

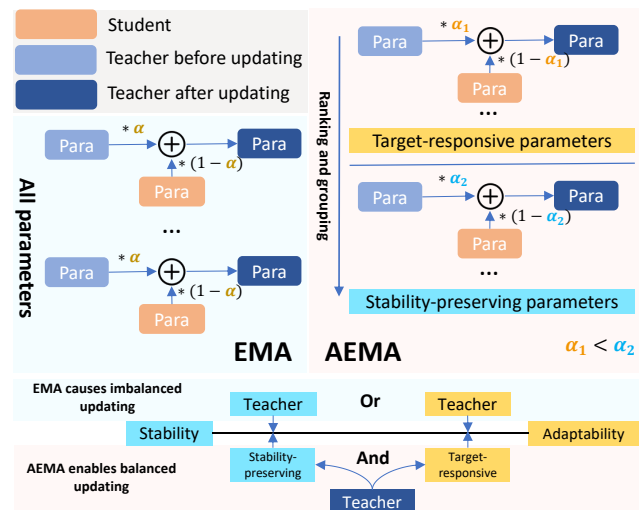


Figure 1. EMA synchronously updates parameters, while AEMA uses decoupled method to control adaptability to target domain and stability on long-term knowledge.

amounts of data for special scenes is time consuming and labor intensive. Therefore, Domain Adaptive Object Detection (DAOD) [3] has been introduced. It leverages existing labeled source domain data and unlabeled target domain data to train a well-adapted target domain detector.

In real-world applications, data security and privacy concerns often restrict the re-access of DAOD to source domain data. To address such demands, Source-Free DAOD (SF-DAOD) [19] has been proposed, which transfers knowledge of the pre-trained source detector to the unlabeled target domain, without accessing the source data during adaptation.

SF-DAOD methods [10, 17, 31] leverage the Mean Teacher [30] to generate pseudo-labels to train the unlabeled target domain. The Mean Teacher framework incorporates a teacher model that serves as a globally smoothed version of the student model. All teacher parameters are updated simultaneously based on the student parameters using an Exponential Moving Average (EMA) with a high smoothing coefficient. EMA can stabilize the teacher model

portance. The teacher is updated via AEMA to perform an asynchronous update with the student. The student is optimized with the detection loss between predictions (strong and masked images) and integrated pseudo-labels. By adjusting the smoothing coefficients based on the importance, DDT can quickly adapt to new knowledge in emerging domains without sacrificing the stability of the learned knowledge structure.

The contributions are summarized as follows:

- Dual-rate Dynamic Teacher with AEMA is proposed for SF-DAOD, which implements a dual-rate updating mechanism to balance the stability of historical knowledge and adaptability of emerging domain knowledge.
- AEMA dynamically decomposes the teacher parameters into target-responsive and stability-preserving parameters, which accelerates the adaptation of new knowledge and maintains the stability of old knowledge.
- Experimental results demonstrate that the proposed DDT achieves state-of-the-art performance in SF-DAOD and shows effectiveness in other domain adaptation tasks.

2. Related works

2.1. Source-Free Domain Adaptation

Source-Free Domain Adaptation (SFDA) aims at transferring knowledge from the source domain to the target domain without accessing the data from the source domain. All SFDA approaches utilize pseudo-labeling technology to adapt the source pre-trained model to the target domain. These methods can be grouped into two categories: non-contrastive and contrastive. Non-contrastive methods [29, 45] use pseudo-labels to align a few samples with trusted pseudo-labels with other data or leverage the target data manifold to generate high-quality pseudo-labels. Contrastive methods [32, 42] use variant-enhanced contrastive learning to improve feature representations.

The base task of SFDA is classification, while Source-Free Domain Adaptive Object Detection (SF-DAOD) works for object detection. The differences between two tasks are huge, e.g., the number of objects per image. Although the training manner of SFDA and SF-DAOD is similar, SFDA methods are difficult to apply directly to SF-DAOD.

2.2. Domain Adaptive Object Detection

The goal of Domain Adaptive Object Detection (DAOD) is to transfer the knowledge of data from the labeled source domain to the unlabeled target domain in the object detection task. Mainstream DAOD methods can be mainly divided into feature alignment and self-training. Feature alignment [25, 33] uses the Gradient Reverse Layer to align the source and target domains. Self-training [39, 44] mainly leverages weak-strong augmentation and mean teacher framework for teacher-student mutual learning. CMT [1]

optimizes object-level features via contrastive learning. To mitigate the bias that the majority class brings to minority classes, CAT [16] applies an instance-level mixup, which prioritizes the model’s attention towards minority classes.

DAOD methods rely on the source data to align or calibrate the target domain. Since SF-DAOD does not use the source data during adaptation, DAOD methods (which require source data) cannot be used directly to SF-DAOD.

2.3. Source-Free Domain Adaptive Object Detection

In consideration of data security and privacy protection, Source-Free DAOD (SF-DAOD), as an extension of DAOD, aims to adapt the source domain knowledge to the target domain without accessing the source domain data during the adaptation phase. Only the source pre-trained detector and unlabeled target domain data are accessible. SF-DAOD methods [6, 17, 31] use self-training to learn the target domain data, in which weak-strong augmentation and mean-teacher framework have shown superior advantages in this field. IRG [31] uses a graph structure to learn better features of the target domain under the mean teacher. To improve the stability of the mean teacher framework, PETS [21] proposes a multi-teacher framework to integrate useful knowledge between dynamic and static teachers. LPLD [37] leverages low-confidence pseudo-labels provided by the teacher to the student to explore more knowledge about the target domain. Moreover, DRU [17] introduces a dynamic retraining-update mean teacher to overcome the noisy pseudo-labels provided by the teacher.

Existing methods overlook a critical flaw in the SF-DAOD mean teacher framework. EMA updates all parameters synchronously, creating a conflict between preserving learned knowledge and adapting to new domains. To address this, our DDT assigns different smoothing coefficients to parameters based on their roles of preserving temporal knowledge and capturing domain shift.

3. Dual-rate Dynamic Teacher

The Mean Teacher framework and the proposed method are outlined in this section.

3.1. Mean Teacher Framework

The Mean Teacher framework [30] has been widely used in semi-supervised learning and unsupervised learning to provide stable self-training for unlabeled data. It consists of a pair of identical network teacher and student. The teacher is updated from the student using EMA, which averages the student over the training steps. The teacher is able to resist disturbances during training and produce more accurate predictions than the student. Therefore, using the pseudo-labels provided by the teacher to supervise the learning of unlabeled data can cope with potential model collapse.

In SF-DAOD, the source domain data are used to train a source detector. At the beginning of adaptation, the source detector weights are utilized to initialize the teacher and the student. During the adaptation phase, pseudo-labels, which are predictions of the teacher with a high confidence (exceeding a specified threshold), are utilized to guide the student on unlabeled data from the target domain. The student is optimized with the gradient descent algorithm.

$$\theta_s \leftarrow \theta_s - \beta \cdot \frac{\partial \mathcal{L}_{det}(preds_s, PL_t)}{\partial \theta_s}, \quad (1)$$

where θ_s denotes the student parameters, β is the learning rate, $\mathcal{L}_{det}(preds_s, PL_t)$ refers to the detection loss, which contains the classification loss (\mathcal{L}_{cls}) and the regression losses (\mathcal{L}_{L1} and \mathcal{L}_{giou}), between the predictions of the student $preds_s$ and the pseudo-labels of the teacher PL_t .

The teacher updates its parameters using the EMA weights of the student.

$$\theta_t \leftarrow \alpha \cdot \theta_t + (1 - \alpha) \cdot \theta_s, \quad (2)$$

where θ_t denotes the parameters of the teacher, and α is the smoothing coefficient.

3.2. Dual-Rate Dynamic Teacher

The framework of the Dual-Rate Dynamic Teacher is illustrated in Figure 2. DDT classifies teacher parameters into two categories: target-responsive and stability-preserving parameters. It assigns a lower smoothing coefficient to target-responsive parameters to embrace rapid adaptation and gives a higher smoothing coefficient to stability-preserving parameters to maintain historical knowledge.

3.2.1. Integration of Knowledge

To enhance the acquisition of domain knowledge, the pseudo-labels generated by both teacher and student models are combined. The teacher model focuses on maintaining stable knowledge through gradual updates, whereas the student model rapidly adapts to new patterns in the target domain. This integration compensates for their differences caused by different learning strategies. By integrating the teacher’s capacity for long-term knowledge retention with the student’s aptitude for short-term pattern adaptation, this approach reduces individual model biases and strengthens the overall learning process. In particular, the teacher model’s historical consistency serves to mitigate the student model’s tendency to overfit to ephemeral patterns, while the adaptability of the student model helps alleviate the rigidity of the teacher model. The resulting integration of consistent historical understanding and flexible pattern recognition fosters more reliable and robust domain knowledge acquisition, ultimately bridging the gap between generalization and domain-specific adaptation.

Given that both the teacher and the student identify a significant number of identical objects, the Non-Maximum Suppression (NMS) is implemented to prevent the assignment of duplicate labels to a single object. Taking into account the pseudo-labels provided by the teacher PL_t and those of the student PL_s , the integrated pseudo-labels PL can be described as follows:

$$PL = nms(cat(PL_t, PL_s)), \quad (3)$$

where nms is NMS operation, and $cat(PL_t, PL_s)$ means concatenating detection labels from PL_t and PL_s . Pseudo-labels PL_t and PL_s are obtained by filtering predictions with thresholds of th_t and th_s , respectively.

3.2.2. Asynchronous Exponential Moving Average

AEMA achieves asynchronous updates of teacher parameters by employing parameter grouping alongside a dual rate update mechanism. It comprises two steps: gradient-aware parameter importance ranking and group-wise updating.

The parameter importance ranking requires defining a measurable proxy to quantify the significance of each parameter during domain adaptation. This is accomplished by analyzing the gradient descent algorithm. The amplitude of the gradients serves as an intrinsic indicator of the contribution of a parameter to the minimization of losses. The optimization objective minimizes the detection loss between the teacher predictions and the aggregated pseudo-labels. As a result, the importance metric can derive from backpropagation gradients. Parameters exhibiting substantial gradient magnitudes are indicative of elevated domain adaptation priority, thereby requiring prioritized updates to effectively facilitate knowledge acquisition within the target domain. In contrast, parameters with smaller gradients are identified as stability-preserving components, which play a critical role in maintaining the consistency of historical knowledge. The importance $s(\theta_t)$ can be formulated as:

$$s(\theta_t) = \left| \frac{\partial \mathcal{L}_{det}(preds_t, PL)}{\partial \theta_t} \right|, \quad (4)$$

where $|\cdot|$ denotes the absolute value. Notably, the teacher gradients are exclusively employed for parameter grouping and are not utilized within the gradient descent process.

The parameters of the teacher model can be categorized into two distinct groups based on their capacity for swift adaptation to novel domain knowledge: target-responsive parameters and stability-preserving parameters. Inspired by Network Pruning [5, 8], only a few parameters are essential for network learning. Given the importance of each parameter, the parameters of highest importance are selected as target-responsive parameters \mathcal{P}_{TR} , and others are indicated as stability-preserving parameters \mathcal{P}_{SP} .

For group-wise asynchronous updating, a reduced smoothing coefficient is allocated to target-responsive parameters to facilitate swift adaptation to distributional

	Method	Detector	person	rider	car	truck	bus	train	motor	bicycle	mAP
	Source Only	DefDETR	40.0	41.2	47.0	13.0	29.1	6.5	21.5	38.0	29.5
DAOD	SW-Faster [25]	FRCNN	32.3	42.2	47.3	23.7	41.3	27.8	28.3	35.4	34.8
	CR-DA-DET [35]	FRCNN	32.9	43.8	49.2	27.2	45.1	36.4	30.3	34.6	37.4
	TIA [43]	FRCNN	34.8	46.3	49.7	31.1	52.1	48.6	37.7	38.1	42.3
	TDD [11]	FRCNN	39.6	47.5	55.7	33.8	47.6	42.1	37.0	41.4	43.1
	CAT [16]	FRCNN	44.3	57.1	63.7	40.8	66.0	49.7	44.9	53.0	52.5
	SFA [33]	DefDETR	46.5	48.6	62.6	25.1	46.2	29.4	28.3	44.0	41.3
	MTTrans [39]	DefDETR	47.7	49.9	65.2	25.8	45.9	33.8	32.6	46.5	43.4
	DA-DETR [40]	DefDETR	49.9	50.0	63.1	24.0	45.8	37.5	31.6	46.3	43.5
MRT [44]	DefDETR	52.8	51.7	68.7	35.9	58.1	54.5	41.0	47.1	51.2	
SF-DAOD	SED (Mosaic) [19]	FRCNN	33.2	40.7	44.5	25.5	39.0	22.2	28.4	34.1	33.5
	A ² SFOD [6]	FRCNN	32.3	44.1	44.6	28.1	34.3	29.0	31.8	38.9	35.4
	LODS [18]	FRCNN	34.0	45.7	48.8	27.3	39.7	19.6	33.2	37.8	35.8
	PETS [21]	FRCNN	42.0	48.7	56.3	19.3	39.3	5.5	34.2	41.6	35.9
	IRG [31]	FRCNN	37.4	45.2	51.9	24.4	39.6	25.2	31.5	41.6	37.1
	LPLD [37]	FRCNN	39.7	49.1	56.6	29.6	46.3	26.4	36.1	43.6	40.9
	SF-UT [10]	FRCNN	40.9	48.0	58.9	29.6	51.9	50.2	36.2	44.1	45.0
	DRU [17]	DefDETR	48.3	51.5	62.5	26.2	43.2	34.1	34.2	48.6	43.6
	DDT (ours)	DefDETR	49.3	53.0	65.4	25.8	43.0	39.7	40.0	47.9	45.5

Table 1. Comparison with the state-of-the-art methods for the adaptation task of Foggy Adaptation. ‘‘Source Only’’ refers to the source-trained model. ‘‘FRCNN’’ denotes Faster R-CNN [23], and ‘‘DefDETR’’ represents Deformable DETR.

changes within the target domain, while an elevated smoothing coefficient is designated for stability-preserving parameters to maintain the integrity of historical knowledge. This dual-rate strategy reflects their distinct roles, where target-responsive parameters prioritize agility, which dynamically aligns with evolving target patterns. Because stability-preserving parameters emphasize inertia, they mitigate abrupt deviations from consolidated historical knowledge. By decoupling update rates based on parameter roles, the proposed approach balances plasticity for adaptation and stability for long-term consistency. The update of each parameter by AEMA can be described as follows.

$$\begin{cases} \theta_t^i \leftarrow \alpha_1 \cdot \theta_t^i + (1 - \alpha_1) \cdot \theta_s^i, & \text{if } \theta_t^i \in \mathcal{P}_{TR}, \\ \theta_t^i \leftarrow \alpha_2 \cdot \theta_t^i + (1 - \alpha_2) \cdot \theta_s^i, & \text{if } \theta_t^i \in \mathcal{P}_{SP}, \end{cases} \quad (5)$$

where θ_t^i and θ_s^i denote the i -th parameter of the teacher and the student, respectively. α_1 and α_2 are the smoothing coefficients for \mathcal{P}_{TR} and \mathcal{P}_{SP} , respectively.

3.2.3. Training Objective

Following previous work DRU [17], the Masked Image Consistency (MIC) [12] is used to facilitate the learning of context relations of unlabeled data in the training phase. Given the strong augmented image x^S , the masked image x^M is generated by $x^M = \mathcal{M} \odot x^S$, where \mathcal{M} is a patch mask sampled from a uniform distribution, and \odot denotes element-wise multiplication.

The total training objective is formulated as follows.

$$\mathcal{L} = \mathcal{L}_{det}^S(preds_s^S, PL) + \mathcal{L}_{det}^M(preds_s^M, PL), \quad (6)$$

where $preds_s^S$ and $preds_s^M$ denote predictions of the student with x^S and x^M , respectively.

4. Experiments

4.1. Experimental Setup

Datasets. Following previous works [6, 17, 21], DDT is validated on four datasets: Cityscapes [7], FoggyCityscapes [26], BDD100K [38], and Sim10K [15]. **Cityscapes** consists of real-world urban scenes and includes eight common categories. **FoggyCityscapes** is a real-world urban scene with foggy, which is generated by applying a synthetic fog filter to the original Cityscapes images. **BDD100K** comprises 100,000 real-world images under six different conditions. Similarly to previous work [6, 17, 21], the ‘‘daytime’’ subset is extracted, with 36,728 training images and 5,258 validation images. **Sim10K** is a simulation dataset generated by the GTA-V game engine.

Adaptation Tasks. DDT is applied to three widely used adaptation tasks. 1) Normal to Foggy Adaptation: *Cityscapes* \rightarrow *FoggyCityscapes*, which means Cityscapes as the source domain and FoggyCityscapes with the highest fog density (0.02) as the target domain, 2) Cross Scene Adaptation: *Cityscapes* \rightarrow *BDD100K*, and 3) Synthetic to Real Adaptation: *Sim10K* \rightarrow *Cityscapes*.

4.2. Implement Details

The proposed DDT is built on Deformable DETR [47]. The batch size is fixed at 8 for all domain adaptation tasks. The

	Method	Detector	truck	car	rider	person	motor	bicycle	bus	mAP
	Source Only	DefDETR	18.9	58.2	28.3	42.0	15.7	18.8	21.7	29.1
DAOD	DA-Faster [3]	FRCNN	14.3	44.6	26.5	29.4	15.8	20.6	16.8	24.0
	SW-Faster [25]	FRCNN	15.2	45.7	29.5	30.2	17.1	21.2	18.4	25.3
	CR-DA-DET [35]	FRCNN	19.5	46.3	31.3	31.4	17.3	23.8	18.9	26.9
	AQT [13]	DefDETR	17.3	58.4	33.0	38.2	16.9	23.5	18.4	29.4
	O ² net [9]	DefDETR	20.4	58.6	31.2	40.4	14.9	22.7	25.0	30.5
	MTTrans [39]	DefDETR	25.1	61.5	30.1	44.1	17.7	23.0	26.9	32.6
	MRT [44]	DefDETR	24.7	63.7	30.9	48.4	20.2	22.6	25.5	33.7
SF-DAOD	SED (Mosaic) [19]	FRCNN	20.6	50.4	32.6	32.4	18.9	25.0	23.4	29.0
	PETS [21]	FRCNN	19.3	62.4	34.5	42.6	17.0	26.3	16.9	31.3
	A ² SFOD [6]	FRCNN	26.6	50.2	36.3	33.2	22.5	28.2	24.4	31.6
	LPU [4]	FRCNN	24.5	55.2	38.9	41.4	20.9	30.4	23.2	33.5
	DRU [17]	DefDETR	27.1	62.7	36.9	45.8	22.7	32.5	28.1	36.6
	DDT (ours)	DefDETR	28.7	66.1	40.3	49.4	29.3	34.1	29.1	39.6

Table 2. Results of Cross Scene Adaptation.

	Method	Detector	AP of car
	Source Only	DefDETR	48.9
DAOD	TDD [11]	FRCNN	53.4
	PT [2]	FRCNN	55.1
	SFA [33]	DefDETR	52.6
	O ² net [9]	DefDETR	54.1
	DA-DETR [40]	DefDETR	54.7
	MTTrans [39]	DefDETR	57.9
	MTM [44]	DefDETR	58.1
SF-DAOD	SED(Mosaic) [19]	FRCNN	43.1
	IRG [31]	FRCNN	43.2
	A ² SFOD [6]	FRCNN	44.0
	PETS [21]	FRCNN	57.8
	LPLD [37]	FRCNN	49.4
	SF-UT [10]	FRCNN	55.4
	DRU [17]	DefDETR	58.7
	DDT (ours)	DefDETR	60.6

Table 3. Results of Synthetic to Real Adaptation.

teacher is updated with AEMA. The smoothing coefficients are set to $\alpha_1=0.997$, $\alpha_2=0.9996$ for most tasks, except that $\alpha_1=0.9997$, $\alpha_2=0.9999$ for Cityscapes→BDD100K. The top 10% of the parameters are selected as target-responsive parameters. The thresholds of pseudo-labels are $th_s=0.5$, $th_t=0.4$. The student is optimized with the Adam optimizer with an initial learning rate of $2e^{-4}$. Following DRU [17], the MIC module is adopted, where the patch size $b=64$ and the mask ratio $r=0.5$. Mean Average Precision (mAP) with a threshold of 0.5 is used as an evaluation metric.

4.3. Comparison with SOTA Methods

The proposed DDT is validated against state-of-the-art approaches in three challenging SF-DAOD adaptation tasks.

1) *Normal to Foggy Adaptation.* As weather conditions exhibit frequent fluctuations that pose substantial challenges, object detectors are required to demonstrate reliability. To address this issue, the detection model is transferred from the Cityscapes to FoggyCityscapes scenario, which is illustrated in Table 1. Compared to DRU with the same detector framework, DDT improves the performance for difficult categories (e.g., “train” and “motor”), in which more pseudo-labels are generated by the teacher with better adaptability. DDT achieves state-of-the-art performance.

2) *Cross Scene Adaptation.* The dynamic change of scene in real-world applications, especially within autonomous driving contexts, address domain shift challenges is crucial. The result of Cityscapes→BDD100K is listed in Table 2, in which the proposed DDT outperforms the previous SOTA method by 3% mAP and achieves the best performance in all categories.

3) *Synthetic to Real Adaptation.* Due to the lower cost of annotating synthetic data compared to acquiring real-world data, training object detectors on synthetic datasets, followed by the application of domain adaptation techniques to real-world scenarios, has emerged as a cost-effective approach. As demonstrated in Table 3, our DDT framework demonstrates superior performance compared to existing methodologies. DDT also outperforms the DefDETR-based DAOD methods by at least 2% mAP.

4.4. Ablation Studies

Ablation studies are conducted on the Foggy Adaptation.

Effect of individual component. The effect of the proposed modules is listed in Table 4. By applying the integration of knowledge with the Mean Teacher, performance is improved from 39.8% to 41.3%. The proposed AEMA uses inconsistent smoothing coefficients, which improves the mAP from 39.8% to 44.6%. When both integration

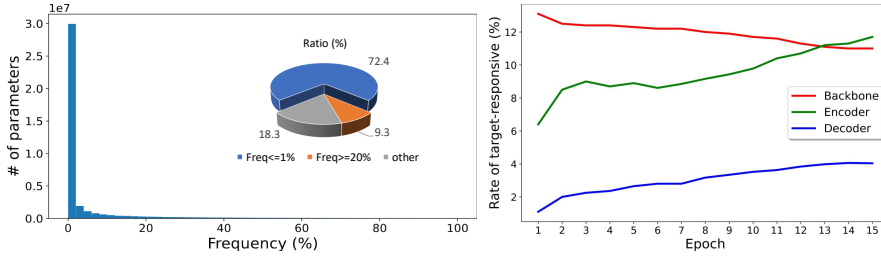


Figure 3. The statistics of target-responsive parameters.

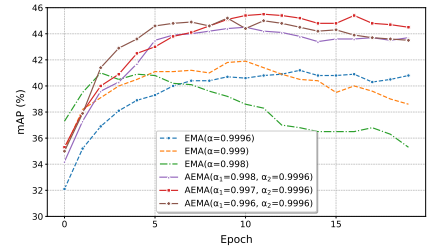


Figure 4. EMA with different coefficients.

MT	MIC	Integration	AEMA	mAP
✓	✓			39.8
✓	✓	✓		41.3
✓	✓		✓	44.6
✓	✓	✓	✓	45.5
✓				37.4
✓		✓	✓	43.2

Table 4. The effect of each component. “MT” denotes the Mean Teacher framework.

and AEMA are combined, it achieves a 5.7% improvement. In addition, DDT also improves the baseline from 37.4% to 43.2% when the MIC module is removed. This illustrates that each component significantly contributes to the enhancement of adaptive capabilities.

Student gradient guidance. As listed in Table 5, the student’s gradient remains effective in driving the AEMA update because the teacher is an ensemble of the student over training steps. However, the inherent local temporal inconsistency between the teacher and the student renders this approach less optimal compared to the direct utilization of the teacher gradient.

Analysis for target-responsive parameters. The statistics of target-responsive parameters are illustrated in Figure 3. We can see that around 70% parameters are hardly chosen (frequency $\leq 1\%$) and about 9% parameters are frequently selected (frequency $\geq 20\%$) during training. On the right, the frequency, where it is $\geq 20\%$, is provided for the backbone, encoder, and decoder of DefDETR. In the early stage of training, the target-responsive parameters are concentrated mainly on the backbone to adapt to the low-level change. They are then shifted to the encoder and decoder to enhance semantic consistency. The experiment verifies the effectiveness of AEMA. It demonstrates the capability of AEMA to facilitate hierarchical parameter updating via gradient-based decoupling, effectively addressing the adaptability-stability dilemma.

Sensitivity for target-responsive parameters. The selection ratio and the smoothing coefficient (α_1) are studied for the target-responsive parameters, which are listed in Ta-

Methods	mAP
Grouped by student gradient	45.3
Grouped by teacher gradient	45.5

Table 5. Grouping under different guidance.

	5%	10%	15%	20%
mAP	45.1	45.5	45.2	44.3

Table 6. Comparison with different selection ratios.

ble 6. Too few or too many target-responsive parameters can affect the teacher update. 10% is a suitable ratio for selecting target-responsive parameters. The results of AEMA with different α_1 are illustrated in Figure 4 (solid lines). A powerful adaptation in the early stage is achieved when $\alpha_1=0.996$. A more stable training process is obtained when $\alpha_1=0.998$, which has a relatively poor result due to the accumulation of errors. A suitable α_1 (0.997) better harmonizes adaptivity and stability, having the best performance.

Comparison with EMA. AEMA is compared with EMA with different smoothing coefficients, which is shown in Figure 4 (dashed line). The teacher model with a lower smoothing coefficient (e.g., $\alpha=0.998$) initially achieves good performance by rapidly incorporating new knowledge from target-domain pseudo-labels. However, in later training stages, this aggressive adaptation amplifies noise from error-prone pseudo-labels, destabilizing historical knowledge and causing performance degradation. In contrast, a higher smoothing coefficient (e.g., $\alpha=0.9996$) stabilizes the teacher model but sacrifices adaptability. The slow teacher updates prevent the student from learning domain-specific features, as the pseudo-labels fail to reflect emerging knowledge. This phenomenon highlights the need for a balanced approach to harmonize stability and adaptability. Using group-wise updates to meet demand, AEMA achieves obvious advantages compared to traditional EMA in SF-DAOD.

Sensitivity for thresholds of pseudo-labels. As shown in Table 7, varying the thresholds for the student and teacher models impacts performance. The combination of a threshold of 0.5 for the student and 0.4 for the teacher achieves the highest mAP. The student model lacks long-term knowledge, which requires a higher threshold to effectively filter out noisy predictions compared to the teacher model.

Effect of the number of groups. From Table 8, we can see that two groups achieve the best performance. To main-

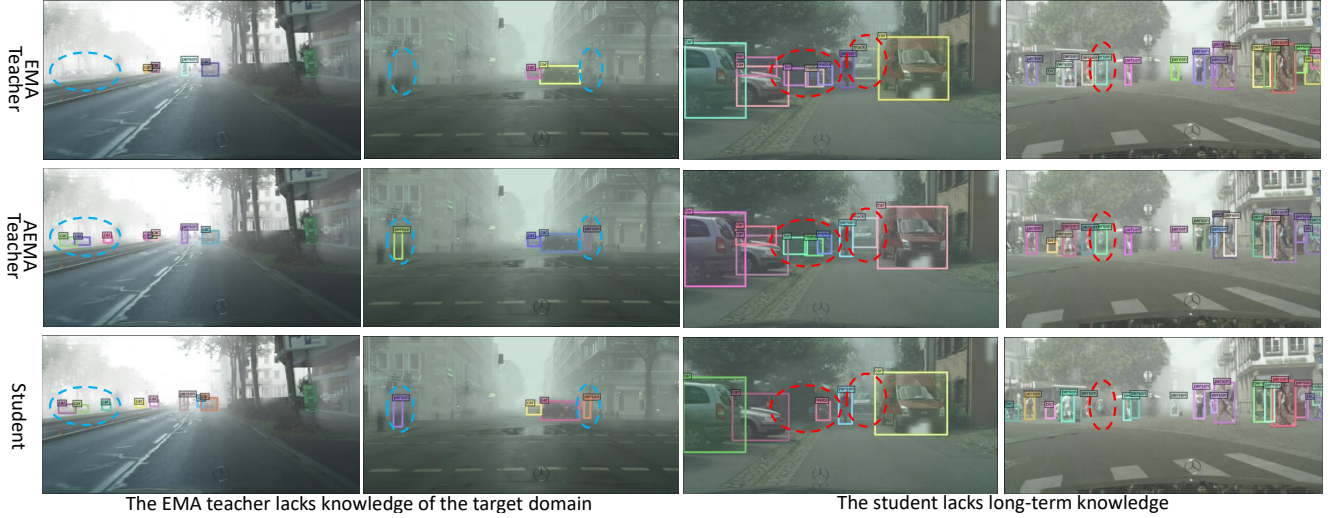


Figure 5. The cyan oval dashed boxes indicate objects missed by the EMA teacher due to insufficient target-domain knowledge, while the red oval dashed boxes show failures of the student model caused by inadequate historical knowledge preservation.

Thresholds	mAP
$th_s = 0.4, th_t = 0.4$	44.4
$th_s = 0.5, th_t = 0.5$	42.9
$th_s = 0.4, th_t = 0.5$	44.6
$th_s = 0.5, th_t = 0.4$	45.5

Table 7. Different thresholds of pseudo-labels.

# of groups	mAP
1	41.3
2	45.5
3	44.7
4	44.0

Table 8. Effect of the number of groups.

tain the robustness of the hyperparameter, linear interpolation between α_1 and α_2 is used to obtain group-specific α . For example, the parameter grouping follows [10%, 10%, 80%] for 3 groups, and the corresponding α are $[\alpha_1, (\alpha_1 + \alpha_2)/2, \alpha_2]$.

Generalization ability of AEMA. To study the generalization of AEMA, it is applied to SFDA (the same adaptation mode, different base tasks) and DAOD (the same base task, different adaptation modes). 1) *Application of AEMA to SFDA.* The proposed AEMA is integrated into the Mean Teacher-based TransDA [36] for SFDA. As presented in Table 9, TransDA+AEMA improves performance. 2) *Application of AEMA to DAOD.* As listed in Table 10, integrated MRT+AEMA increases performance. It indicates that AEMA is robust to the improvement of Mean Teacher in cross-domain learning.

Qualitative visualization. The detection results are visualized to demonstrate the quality of the proposed DDT. The EMA teacher ($\alpha=0.9996$) is updated during DDT training. As shown in Figure 5, the EMA teacher fails to detect certain objects (cyan oval dashed boxes) due to insufficient emerging knowledge detected by the AEMA teacher and the student, while the student misses others (red oval

Method	Avg
TransDA [36]	83.0
TransDA + AEMA	84.2

Table 9. Average accuracy on VisDA [22] for SFDA.

Method	mAP
MRT [44]	51.2
MRT + AEMA	51.6

Table 10. Normal to foggy adaptation in DAOD manner.

dashed boxes) requiring historical consistency detected by the EMA and AEMA teachers. Our AEMA resolves this duality by balancing emerging knowledge learning and historical knowledge preservation. Therefore, the AEMA teacher is capable of detecting objects that the EMA teacher and the student do not detect, exhibiting enhanced robustness in domain adaptation tasks.

5. Conclusion

To address the limitations of Exponential Moving Average (EMA) caused by domain adaptation, we propose a novel Dual-rate Dynamic Teacher (DDT) framework, equipped with an Asynchronous EMA (AEMA) updating strategy. The AEMA strategy enables asynchronous updates of teacher parameters to capture the domain change, achieving a balance between rapid adaptation to the target domain and retention of historical knowledge. In addition, the integration of knowledge between the teacher and the student captures a comprehensive domain knowledge to improve the student’s optimization and teacher updating. Experimental results in three adaptation tasks validated the effectiveness of DDT, and extensive ablation studies provided deeper insight into its mechanisms, underscoring its potential for robust domain adaptation.

6. Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62372387), and Special Research Funding under Yibin Municipal-University Dual Agreement (No. YBSCXY2024010012, YBSCXY2024010006).

References

- [1] Shengcao Cao, Dhiraj Joshi, Liang-Yan Gui, and Yu-Xiong Wang. Contrastive mean teacher for domain adaptive object detectors. In *CVPR*, pages 23839–23848, 2023. 3
- [2] Meilin Chen, Weijie Chen, Shicai Yang, Jie Song, Xinchao Wang, Lei Zhang, Yunfeng Yan, Donglian Qi, Yueting Zhuang, Di Xie, et al. Learning domain adaptive object detection with probabilistic teacher. In *ICML*, 2022. 6
- [3] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *CVPR*, pages 3339–3348, 2018. 1, 6
- [4] Zhihong Chen, Zilei Wang, and Yixin Zhang. Exploiting low-confidence pseudo-labels for source-free object detection. In *ACM MM*, pages 5370–5379, 2023. 2, 6
- [5] Hongrong Cheng, Miao Zhang, and Javen Qinfeng Shi. A survey on deep neural network pruning: Taxonomy, comparison, analysis, and recommendations. *IEEE TPAMI*, 2024. 4
- [6] Qiaosong Chu, Shuyan Li, Guangyi Chen, Kai Li, and Xiu Li. Adversarial alignment for source free object detection. In *AAAI*, pages 452–460, 2023. 3, 5, 6
- [7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, pages 3213–3223, 2016. 5
- [8] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *ICLR*, 2019. 2, 4
- [9] Kaixiong Gong, Shuang Li, Shugang Li, Rui Zhang, Chi Harold Liu, and Qiang Chen. Improving transferability for domain adaptive detection transformers. In *ACM MM*, pages 1543–1551, 2022. 6
- [10] Yan Hao, Florent Forest, and Olga Fink. Simplifying source-free domain adaptation for object detection: Effective self-training strategies and performance insights. In *ECCV*, pages 196–213, 2024. 1, 5, 6
- [11] Mengzhe He, Yali Wang, Jiayi Wu, Yiru Wang, Hanqing Li, Bo Li, Weihao Gan, Wei Wu, and Yu Qiao. Cross domain object detection by target-perceived dual branch distillation. In *CVPR*, pages 9570–9580, 2022. 5, 6
- [12] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *CVPR*, pages 11721–11732, 2023. 5
- [13] Wei-Jie Huang, Yu-Lin Lu, Shih-Yao Lin, Yusheng Xie, and Yen-Yu Lin. Aqt: Adversarial query transformers for domain adaptive object detection. In *IJCAI*, pages 972–979, 2022. 6
- [14] Jinbae Im, JeongYeon Nam, Nokyoung Park, Hyungmin Lee, and Seunghyun Park. Egtr: Extracting graph from transformer for scene graph generation. In *CVPR*, pages 24229–24238, 2024. 1
- [15] Matthew Johnson-Roberson, Charles Barto, Rounak Mehta, Sharath Nittur Sridhar, Karl Rosaen, and Ram Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? *arXiv preprint arXiv:1610.01983*, 2016. 5
- [16] Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, and Robby T Tan. Cat: Exploiting inter-class dynamics for domain adaptive object detection. In *CVPR*, pages 16541–16550, 2024. 3, 5
- [17] Trinh Le Ba Khanh, Huy-Hung Nguyen, Long Hoang Pham, Duong Nguyen-Ngoc Tran, and Jae Wook Jeon. Dynamic retraining-updating mean teacher for source-free object detection. In *ECCV*, pages 328–344, 2024. 1, 2, 3, 5, 6
- [18] Shuaifeng Li, Mao Ye, Xiatian Zhu, Lihua Zhou, and Lin Xiong. Source-free object detection by learning to overlook domain style. In *CVPR*, pages 8014–8023, 2022. 5
- [19] 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 *AAAI*, pages 8474–8481, 2021. 1, 5, 6
- [20] 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 *ECCV*, pages 740–755, 2014. 1
- [21] Qipeng Liu, LuoJun Lin, Zhifeng Shen, and Zhifeng Yang. Periodically exchange teacher-student for source-free object detection. In *ICCV*, pages 6414–6424, 2023. 2, 3, 5, 6
- [22] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 8
- [23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE TPAMI*, 39(6):1137–1149, 2016. 5
- [24] Noam Rotstein, David Bensaid, Shaked Brody, Roy Ganz, and Ron Kimmel. Fusecap: Leveraging large language models for enriched fused image captions. In *WACV*, pages 5689–5700, 2024. 1
- [25] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *CVPR*, pages 6956–6965, 2019. 3, 5, 6
- [26] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *IJCV*, 126:973–992, 2018. 5
- [27] Shuai Shao, Zeming Li, Tianyuan Zhang, Chao Peng, Gang Yu, Xiangyu Zhang, Jing Li, and Jian Sun. Objects365: A large-scale, high-quality dataset for object detection. In *ICCV*, pages 8430–8439, 2019. 1
- [28] Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, and Muhammad Zeshan Afzal. Sparse semi-detr: sparse learnable queries for semi-supervised object detection. In *CVPR*, pages 5840–5850, 2024. 2

- [29] Song Tang, An Chang, Fabian Zhang, Xiatian Zhu, Mao Ye, and Changshui Zhang. Source-free domain adaptation via target prediction distribution searching. *IJCV*, 132(3):654–672, 2024. [3](#)
- [30] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NeurIPS*, 2017. [1](#), [3](#)
- [31] Vibashan VS, Poojan Oza, and Vishal M Patel. Instance relation graph guided source-free domain adaptive object detection. In *CVPR*, pages 3520–3530, 2023. [1](#), [3](#), [5](#), [6](#)
- [32] Jing Wang, Wonho Bae, Jiahong Chen, Kuangen Zhang, Leonid Sigal, and Clarence W de Silva. What has been overlooked in contrastive source-free domain adaptation: Leveraging source-informed latent augmentation within neighborhood context. In *ICLR*, 2025. [3](#)
- [33] Wen Wang, Yang Cao, Jing Zhang, Fengxiang He, Zheng-Jun Zha, Yonggang Wen, and Dacheng Tao. Exploring sequence feature alignment for domain adaptive detection transformers. In *ACM MM*, pages 1730–1738, 2021. [3](#), [5](#), [6](#)
- [34] Xinjiang Wang, Xingyi Yang, Shilong Zhang, Yijiang Li, Litong Feng, Shijie Fang, Chengqi Lyu, Kai Chen, and Wayne Zhang. Consistent-teacher: Towards reducing inconsistent pseudo-targets in semi-supervised object detection. In *CVPR*, pages 3240–3249, 2023. [2](#)
- [35] Chang-Dong Xu, Xing-Ran Zhao, Xin Jin, and Xiu-Shen Wei. Exploring categorical regularization for domain adaptive object detection. In *CVPR*, pages 11724–11733, 2020. [5](#), [6](#)
- [36] Guanglei Yang, Hao Tang, Zhun Zhong, Mingli Ding, Ling Shao, Nicu Sebe, and Elisa Ricci. Transformer-based source-free domain adaptation. *arXiv preprint arXiv:2105.14138*, 2021. [8](#)
- [37] Ilhoon Yoon, Hyeongjun Kwon, Jin Kim, Junyoung Park, Hyunsung Jang, and Kwanghoon Sohn. Enhancing source-free domain adaptive object detection with low-confidence pseudo label distillation. In *ECCV*, pages 337–353, 2024. [3](#), [5](#), [6](#)
- [38] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *CVPR*, pages 2636–2645, 2020. [5](#)
- [39] Jinze Yu, Jiaming Liu, Xiaobao Wei, Haoyi Zhou, Yohei Nakata, Denis Gudovskiy, Tomoyuki Okuno, Jianxin Li, Kurt Keutzer, and Shanghang Zhang. Mtrains: Cross-domain object detection with mean teacher transformer. In *ECCV*, pages 629–645, 2022. [3](#), [5](#), [6](#)
- [40] Jingyi Zhang, Jiaying Huang, Zhipeng Luo, Gongjie Zhang, Xiaoqin Zhang, and Shijian Lu. Da-detr: Domain adaptive detection transformer with information fusion. In *CVPR*, pages 23787–23798, 2023. [5](#), [6](#)
- [41] Jiacheng Zhang, Xiangru Lin, Wei Zhang, Kuo Wang, Xiao Tan, Junyu Han, Errui Ding, Jingdong Wang, and Guanbin Li. Semi-detr: Semi-supervised object detection with detection transformers. In *CVPR*, pages 23809–23818, 2023. [2](#)
- [42] Ziyi Zhang, Weikai Chen, Hui Cheng, Zhen Li, Siyuan Li, Liang Lin, and Guanbin Li. Divide and contrast: Source-free domain adaptation via adaptive contrastive learning. In *NeurIPS*, pages 5137–5149, 2022. [3](#)
- [43] Liang Zhao and Limin Wang. Task-specific inconsistency alignment for domain adaptive object detection. In *CVPR*, pages 14217–14226, 2022. [5](#)
- [44] Zijing Zhao, Sitong Wei, Qingchao Chen, Dehui Li, Yifan Yang, Yuxin Peng, and Yang Liu. Masked retraining teacher-student framework for domain adaptive object detection. In *ICCV*, pages 19039–19049, 2023. [3](#), [5](#), [6](#), [8](#)
- [45] Lihua Zhou, Nianxin Li, Mao Ye, Xiatian Zhu, and Song Tang. Source-free domain adaptation with class prototype discovery. *PR*, 145:109974, 2024. [3](#)
- [46] Fangrui Zhu, Yiming Xie, Weidi Xie, and Huaizu Jiang. Diagnosing human-object interaction detectors. *IJCV*, pages 1–18, 2025. [1](#)
- [47] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. In *ICLR*, 2021. [5](#)