Class-aware Memory Guided Unbiased Weighting for Universal Domain Adaptive Object Detection

Qinghai Lang, Zhenwei He, Xiaowei Fu, Lei Zhang*
School of Microelectronics and Communication Engineering, Chongqing University, China
College of Computer Science and Engineering, Chongqing University of Technology, China
qhlang@cqu.edu.cn, hzw@cqut.edu.cn, xwfu@cqu.edu.cn, leizhang@cqu.edu.cn

Abstract

Cross-domain object detection aims to align the feature distributions across the source and target domains. Existing cross-domain object detectors typically rely on identical label space assumption, which, however, greatly limits their universality under class gap. This paper introduces Universal Domain Adaptive Object Detection (UDAOD) toward more practical scenarios without any prior knowledge on the category consistency. In the proposed universal setting, the category space is partially intersected (i.e., common classes) between domains. The class gap caused by source-private and target-private classes leads to serious negative transfer and degrades adaptation performance. To this end, we propose a Universal Cross-domain Faster R-CNN (UCF) with a novel unbiased weighting mechanism to effectively measure the common or private classes. Specifically, we propose a dynamic Class-aware Memory (CaM) to overcome the bias of class weights, caused by class incompleteness in a batch of UniDA. We further propose a Weight Surgery Equalization (WSE) to strengthen the polarization of the weights for common and private classes and suppress incorrect alignment. Extensive experiments under the novel UDAOD setting on multiple benchmarks including PASCAL VOC, Clipart, WaterColor, Cityscapes, and FoggyCityscapes are implemented, which shows the SOTA universality of our model.

1. Introduction

Object detection [23, 29, 30, 1] has made a great progress in the deep learning era, which relies on a lot of labeled data to extract representative features. Although excellent achievements have been made, due to the difficulty of data annotation and the diversity of domain distribution, domain adaptive object detection has attracted extensive attention.

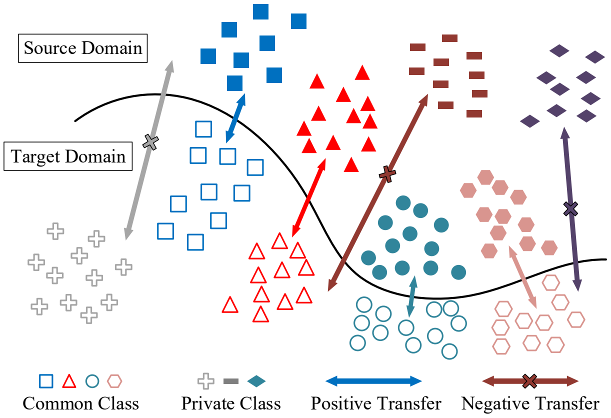


Figure 1. The Universal Domain Adaptive Object Detection (UDAOD) scenario introduced in this paper. The classes can be divided into common classes and private classes. Due to the absence in another domain, the private classes can be misaligned to another class to cause the negative transfer, which can lead to the deterioration of the detector on the target domain.

Domain adaptive object detection [4, 14, 31, 15, 42, 26] (DAOD) aims to transfer knowledge from labeled source data to unlabeled target data. However, the existing domain adaptive object detection algorithms all adopt an ideal and prior assumption that the label spaces are identical across domains (i.e. close-set problem), which greatly limits their application in the wild. For example, in real applications, it is not practical to find a source domain having the same label space as the target domain due to the diversity of detection categories. In domain adaptive image classification, recent work [2, 41] propose the partial domain adaption that requests the source label set to contain the target label set. On the contrary, open set domain adaptation [27, 32] assumes that the source label set belongs to a subset of the target domain. You *et al.* [40] proposes the universal domain adaptation (UniDA) that includes all the above scenarios. Compared to image classification task, the label spaces across domains are more likely to be unequal for detection task, because there are multiple objects in an image.

*Corresponding author (Lei Zhang)

Therefore, we propose a more generalized setting in object detection, termed **Universal Domain Adaptive Object Detection (UDAOD)**, inspired by universal domain adaptation (UniDA) that requires no prior knowledge of target label set. In other words, the label set of the target domain may differ from that of the source domain. As shown in Figure 1, the classes of two domains can be divided into two categories: (1) common classes, which are the intersection of two domain label sets and exist in both two domains; (2) private classes, including source private classes and target private classes, which are disjoint. The private classes may lead to negative transfer due to its absence in another domain, which is called **class gap** (category shift). The class gap can give rise to serious degradation of domain adaptive object detection performance. Therefore, in UDAOD, object detectors are explored to simultaneously adapt across domains and classes.

To address Universal Domain Adaptive Object Detection, we propose a **Universal Cross-domain Faster R-CNN (UCF)**, which, to the best of our knowledge, is the first work addressing the universal domain adaptation problem in object detection. To explore the domain gap, UCF aims to align the feature distribution of both source and target domains in the common label space. In detail, UCF down-weigh the importance of the private classes and up-weigh common classes adaptively. However, the previous algorithms in UniDA [2, 41, 27, 40] excessively rely on the current batch training data to evaluate the entire label space of the dataset. Due to the class incompleteness of small batch size, only a minority of classes in the current batch contribute to the weights, which are clearly class-biased and lead to incorrect alignment. Therefore, this paper proposes a novel unbiased weighting mechanism to facilitate the feature alignment of common classes and reduce the interference of private classes. Specifically, we first design a **C**lass-**a**ware **M**emory (CaM) module, which aims to get the reliable weight of each class by compensating the class incompleteness of small minibatch. Another tricky problem is, the weights between common and private classes are not so polarized, and the weights for common or private classes are unbalanced (described in Figure 4). This clearly leads to incorrect alignment between common and private classes. To this end, we propose a **W**eight **S**urgery **E**qualization (WSE) to strengthen the polarization of the weights between common and private classes from the CaM. With unbiased weights generated from both CaM and WSE, UCF effectively aligns the domains with respect to common classes, without cross-interference from the private classes. Universal detection on unlabeled target domain is then achieved. The contributions of this paper can be summarized as follows:

- We introduce a novel and practical Universal Domain Adaptive Object Detection (UDAOD) setting that requires no prior knowledge on the target label sets, which, to the best of our knowledge, is the first work to address the universality of DAOD in more practical scenarios under the domain and category shifts.
- We propose a Universal Cross-domain Faster RCNN (UCF) by exploring the unbiased weighting mechanism equipped with the Class-aware Memory (CaM) and Weight Surgery Equalization (WSE), which overcomes the challenges in UniDA and DAOD.
- Extensive experiments in universal settings including closed set, partial set, and open set, verify the superiority of our UCF over baselines on multiple benchmarks.

2. Related Work

2.1. Object Detection

Object detection has achieved great success in computer vision, relying on the development of convolutional neural network (CNN) [19] and a large amount of labeled training data [7, 20]. The object detection algorithms can be roughly divided into two categories: one-stage detection algorithms and two-stage detection algorithms. Due to the excellent detection speed of one-stage detection algorithms [23, 28, 34, 43, 29, 21, 28, 36], they have received extensive attention. In two-stage detection algorithms, R-CNN [10] is the first two-stage detector that extracts region proposals by classifying region of interest (ROI). Faster RCNN [30] combines Fast R-CNN [9] and Region Proposal Network (RPN) to efficiently produce object proposals. Due to its SOTA results and good scalability, most domain adaptive detection methods [12, 1, 4, 14, 3] choose it as the backbone.

2.2. Domain Adaptive Object Detection

Although the object detection algorithms perform well in a single domain, the detector performance degrades sharply when it faces the challenge of domain shift. Recent works [4, 14, 18, 38, 16, 22] mainly mitigate domain differences through adversarial learning. Chen *et al.* [4] first introduce the domain adaptive object detection (DAOD) setting and propose the DAF by learning domain-invariant features to mitigate the domain shift from the image-level alignment and instance-level alignment. After that, a large number of excellent detection algorithms emerge to overcome the problem from domain adaption. He and Zhang [14] propose a hierarchical alignment network in which multiple domain discriminators are deployed in the last three blocks of VGG-16 [35] to generalize well on the unlabeled target domain.

2.3. Universal Domain Adaption

Existing domain adaptation methods [8, 24, 39, 11, 25] in classification task generally assume that the source and

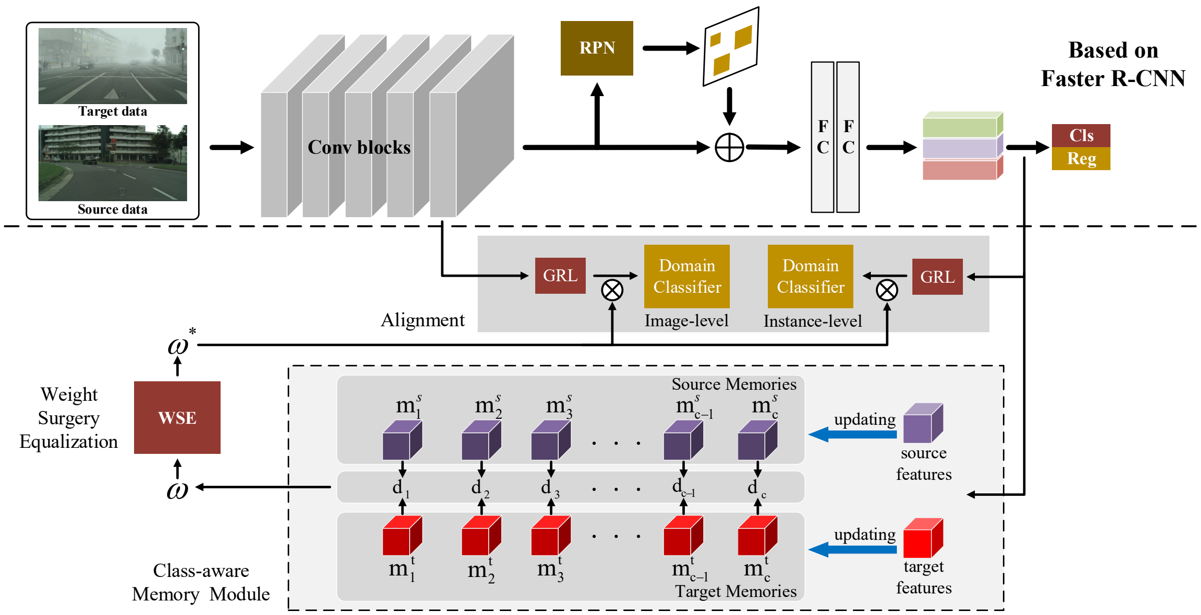


Figure 2. The architecture of our proposed UCF. The UCF is based on the Faster R-CNN with an unbiased weighting mechanism by using the Class-aware Memory module and Weight Surgery Equalization to adapt across the domains and classes.

target domains share identical label space. However, in real applications, it is usually not easy to find a source domain with identical label space as the target domain of interest. Therefore, Cao *et al.* [2] introduce the Partial Domain Adaption problem which assumes that the target label space is a subset of the source label space, and propose the weighting mechanism based on predictions in the target domain. Busto *et al.* [27] propose the Open Set Domain Adaption scene in which there is an intersection between the source and the target domain label spaces. You *et al.* [40] propose a generalized Universal Domain Adaptation (UniDA) setting that requires no prior knowledge about the label space between domains and contains all the above scenarios. However, the existing domain adaptive object detection methods [4, 14, 18, 38, 22] rely on the ideal prior knowledge about the identical label space between the source and target domains, which greatly limits their universality in the wild. Therefore, we propose a more universal setting without any prior knowledge of the target domain in DAOD.

3. Universal Domain Adaptive Object Detection

In this section, we will introduce the problem setting of Universal Domain Adaptive Object Detection (UDAOD). In UDAOD, formally, the fully labeled source domain is denoted by $\mathcal{D}_s = \{(x^s_i, b^s_i, y^s_i)\}^{n_s}_i$, where x^s_i stands for the i-th image, b^s_i is the coordinate of bounding boxes, y^s_i is the category label and n_s is the number of samples from source domain. Similarly, the unlabeled target domain is denoted by $\mathcal{D}_t = \{(x^t_i)\}^{n_t}_i$. Inspired by the Universal Domain Adaption (UniDA) [40], we use $\mathcal{C}_s$ to denote the label set of the source domain and $\mathcal{C}_t$ is the label set of the target domain. The common label sets from both domains are defined as $\mathcal{C}_\wedge = \mathcal{C}_s \cap \mathcal{C}_t$. $\overline{\mathcal{C}}_s = \mathcal{C}_s \backslash \mathcal{C}_\wedge$ and $\overline{\mathcal{C}}_t = \mathcal{C}_t \backslash \mathcal{C}_\wedge$ represent the label set of the private classes from source domain and target domain, respectively. Note that the target domain is unlabeled without any prior category knowledge. Meanwhile, the coincidence rate of two label sets across domains is defined as $\xi = |\mathcal{C}_s \cap \mathcal{C}_t| / |\mathcal{C}_s \cup \mathcal{C}_t|$, which represents the degree of difference between the two domains. The task of UDAOD is to design a detector that can generalize well to unlabeled target domain no matter what the ξ is.

Besides the domain gap, the **class gap** between the source and target domains is a new challenge for UDAOD. Since the $\overline{\mathcal{C}}_s$ ($\overline{\mathcal{C}}_t$) has no intersection with another domain, the existing domain adaptive detector forcibly aligns private classes to other classes. Such a blind alignment can lead to the negative transfer and degenerate the domain-invariant feature representation, which causes the object of the target domain to be incorrectly located or classified.

4. The Proposed UCF Detector

We propose the Universal Cross-domain Faster R-CNN (UCF) to address the UDAOD problem by facilitating the alignment of the common classes while suppressing the private classes. As shown in Figure 2, the proposed UCF con-

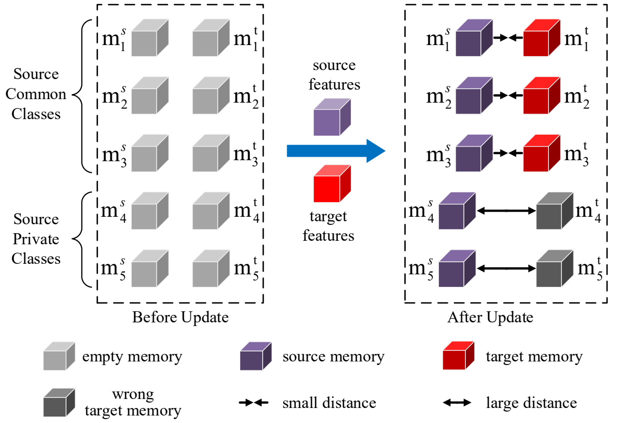


Figure 3. The update process of Class-aware Memory (CaM) module. Since both source and target domain contain the common classes, the memories are probably updated by their corresponding instance feature, such that the distances tend to be small. On the contrary, because the target instance must not belong to the source private classes, the target memory m_4^t and m_5^t store the wrong features so that they have a larger distance compared to the common classes.

tains a detection part and an adaption part.

Detection Part. The detection part adopts the powerful Faster R-CNN [30], which is a famous basic two-stage detector. The backbone of our model is set as Res101 [13].

Adaption Part. The adaption part aims to overcome the class gap in UDAOD by up-weighting the common class and down-weighting the private classes. To this end, we introduce the unbiased weighing mechanism with the proposed **C**lass-**a**ware **M**emory (CaM) and **W**eight **S**urgery **E**qualization (WSE) for unbias weighting mechanism. CaM is introduced to overcome the problem of unreliable weight caused by class incompleteness of batch samples in existing UniDA models [2, 27, 32, 40]. WSE equalizes, polarizes, and refines the class weights to avoid incorrect alignment between common and private classes. By dynamically measuring each class during the training phase, the negative transfer caused by the class gap can be alleviated. The CaM and WSE are elaborated later.

4.1. Unbiased Weighting Mechanism by CaM

Conventional UniDA methods, which are designed for the image classification task, measure the common and private classes by the entropy-based weighting based on the current batch samples. However, due to the small batch size, only a minority of classes in the current batch contribute to the weights. Therefore, the class weight is clearly biased due to the absence of some classes in the mini-batch. Particularly, in the object detection task, the batch size is much smaller (e.g., 2 input images per iteration) and obviously cannot cover the entire label space. To this end, we propose a Class-aware Memory (CaM) module in order to progressively store and the central feature (prototype) of each class, and compute the unbiased weights for UDAOD. As shown in Figure 2, we design the memories for the source and target domains, respectively, and denoted as $M^s = \{m_1^s, m_2^s, \ldots, m_c^s\}$ and $M^t = \{m_1^t, m_2^t, \ldots, m_c^t\}$. c is the number of memories, which is equal to the number of classes of the source domain.

Memory update. Specifically, we adopt the source (target) instance-level features to update the source (target) memories. As for the source domain, we update the corresponding memory according to the ground truth label of each instance. For the target domain, the prediction result of the classifier determines which memory the instance updates. The update criterion is defined as follows:

$$m_i \leftarrow (1-p) \times m_i + p \times f_{ins} \tag{1}$$

$$p = \begin{cases} score, & score < 0.5 \\ 0.5, & score \geq 0.5 \end{cases} \tag{2}$$

where f_{ins} is the source or target instance-level features, m_i is the source or target memory of the ith class, and p means the update rate. We set $p < 0.5$ to guarantee the quality of memories in a progressive manner.

Weight Computation. The weight computation process of CaM is shown in Figure 3. We adopt the $2 \times c$ memories to reliably calculate the unbiased weight. The left side of Figure 3 shows the initialized memory pairs, which come from both source and target domains. After memory update, the distances between the source and target memory pairs of common or private classes are different, as presented on the right side of Figure 3. Since both source and target domain contain the common classes, the memories are probably updated by their corresponding instance feature, such that the distances of the common class memory pairs tend to be small. On the contrary, because the target instance must not belong to the source private classes, the target memory of source private classes store the wrong features so that they have a larger distance compared to the common classes. Therefore, we can use the distance to calculate the unbiased weight, which can be written as:

$$\omega_i = e^{-\|m_i^s - m_i^t\|_2}, i \in [1, c] \tag{3}$$

where the smaller the distance between memories is, the more likely the class is to be common and should be given larger weight. In practice, it is possible that some of the weights are very small due to exponential term. Thus, we normalize the weight ω between 0 and 2 by Min-Max Normalization, i.e. $\omega \leftarrow 2(\omega - \omega_{min})/(\omega_{max} - \omega_{min})$.

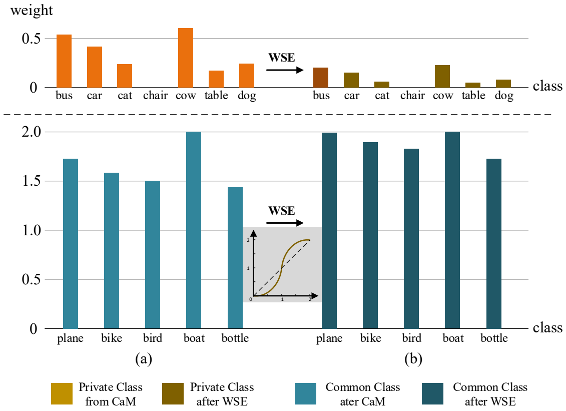


Figure 4. The motivation of our Weight Surgery Equalization (WSE). (a) shows the weighted results on the Partial Domain Adaptation (from PASCAL VOC to WaterColor) for each category by using Class-aware Memory (CaM), which is unbalanced and is unstable for universal transfer learning. (b) is the weight distribution adjusted by WSE, which is uniform and reasonable.

4.2. Weight Surgery Equalization (WSE)

By exploring the unbiased weighting mechanism via the Class-aware Memory (CaM) module, the network successfully distinguishes common classes from private classes and gives them corresponding weights. However, the weights may be unbalanced among the common classes or private classes. As shown in Figure 4 (a), 'boat' and 'bottle' are both common classes, but their weights are quite different. The unbalanced weights have two negative influences on our model. First, the small weight of common classes is approaching the large weight of private classes, i.e., polarization is not enough, such that incorrect alignment between partial common classes and partial private classes is resulted. In other words, the suppression of private classes and reinforcement of common classes are not sufficient. Second, the unbalanced weights will also result in unequalized domain alignment for common classes, while they should be aligned equally. To this end, we propose a weight surgery equalization (WSE) to strengthen the polarization between common and private classes, and simultaneously alleviate the unbalance of weights from common classes:

$$\omega^* = \frac{2}{1+e^{-\alpha(\omega-1)}} \tag{4}$$

where α represents the curvature of the function curve and the ω is from the CaM in Eq.(3). Figure 4 (b) shows the weight distribution adjusted by our proposed WSE. WSE balances the weights of common classes or private classes and achieves weight polarization between the common and private classes. Note that the background class remains unchanged and is set to 1 in our network.

4.3. Overall Objective Function

Inspired by [4], the feature alignment on both image-level and instance-level are implemented for UCF. If the weights for source and target domains are presented as ω^s and ω^t, respectively. The loss function of weighted image-level universal discriminator is written as:

$$\mathcal{L}_{img} = -\sum[\omega^s_{(u,v)} \cdot \log p^s_{(u,v)} + \omega^t_{(u,v)} \cdot \log(1 - p^t_{(u,v)})] \tag{5}$$

where (u, v) is the coordinate of feature map, p^s and p^t are the output of discriminator for source and target domains, respectively. Note that the target samples are implemented with their pseudo labels. Besides, the loss function of weighted instance-level universal discriminator can be presented as:

$$\mathcal{L}_{ins} = -\sum_m \omega^s_i \cdot \log p^s_i - \sum_n \omega^t_j \cdot \log(1 - p^t_j) \tag{6}$$

where m, n are the number of instances for the source and target domains, respectively. By combining the detection loss L_{det} and universal domain alignment loss, the total loss function of the proposed UCF detector can be written as:

$$\mathcal{L}_{UCF} = \mathcal{L}_{det} + \lambda(\mathcal{L}_{img} + \mathcal{L}_{ins}) \tag{7}$$

where λ is a trade-off parameter to balance the detection loss and adaptation loss. The adversarial learning strategy is implemented with a GRL [8], which automatically reverses the gradient during propagation.

5. Experiments

In this section, we compare our proposed Universal Cross-domain Faster R-CNN (UCF) detector under a variety of Universal Domain adaptive Object Detection (UDAOD) settings on several datasets with different ξ, $|\mathcal{C}_s \cup \mathcal{C}_t|$, $\overline{\mathcal{C}}_s$ and $\overline{\mathcal{C}}_t$. Then, we analyze the proposed UCF with several experiments.

5.1. Experiment Setup

For fair comparison, the experiments all in this paper adopt Faster R-CNN [30] with ResNet-101 [13] backbone pre-trained on ImageNet [6]. The source domain is sufficiently annotated with bounding boxes and corresponding categories, while the target domain is completely unlabeled without any prior label knowledge. We resize the shorter side of the input image to 600 pixels. We set the hyper-parameter $\alpha = 5$ in Eq. (4) and the tradeoff parameter λ as 0.01 in Eq. (7) during the training phase. We optimize the network by using Stochastic Gradient Descent (SGD) with a momentum of 0.9 and a weight decay of 0.0005. The initial learning rate is set to 0.001 and dropped to 0.0001 after 50k iterations. Totally, 70k iterations are trained.

Table 1. Results (%) on universal adaptation from PASCAL VOC to Clipart (ξ = 0.75). In this scenario, domain private classes exist in both the source and target domain. The source private classes include: *aeroplane*, *bicycle* and *bird*. The target private classes include: *train* and *tv*. Note that the "Source Only" is the Faster R-CNN [30] without any adaption. The "DAF*" means that we just simply add the weighting method based on sample prediction to DAF [4], inspired by UniDA [2, 40]. The "UCF w/o WSE" denotes the ablation analysis without our proposed WSE.

Methods	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	bike	person	plant	sheep	sofa	mAP
Source Only [30]	31.8	41.2	31.1	34.7	5.1	33.7	23.0	20.7	8.3	43.0	52.7	49.6	40.6	17.0	13.8	29.8
DAF [4]	37.2	38.0	26.9	35.9	2.3	35.2	24.0	28.5	4.2	33.8	54.7	59.4	58.4	13.4	17.9	31.3
MAF [14]	24.2	42.9	35.1	32.3	11.0	**41.7**	22.4	**32.6**	6.7	40.0	59.1	52.7	41.0	**24.1**	17.9	32.2
HTCN [3]	25.9	**47.8**	**36.0**	32.8	11.3	39.4	51.7	18.7	**10.5**	40.9	56.3	57.9	49.4	21.3	20.4	34.7
DAF*	**38.8**	35.1	30.9	34.8	**16.8**	30.4	42.6	29.2	5.8	39.7	51.6	53.9	**54.9**	12.6	10.8	32.5
UCF w/o WSE	31.7	36.5	26.8	36.8	1.3	29.9	50.6	29.2	5.2	**42.7**	60.8	60.5	52.2	14.9	19.1	33.2
UCF	36.2	44.3	28.3	**37.1**	2.2	36.0	**61.9**	27.7	4.0	39.9	**64.7**	**64.2**	52.6	20.9	**26.9**	**36.5**

Table 2. Results (%) on universal adaptation from PASCAL VOC to Clipart (ξ = 0.50). The source private classes include: *aeroplane*, *bicycle*, *bird*, *boat* and *bottle*. The target private classes include: *plant*, *sheep*, *sofa*, *train* and *tv*.

Methods	bus	car	cat	chair	cow	table	dog	hrs	bike	prsn	mAP
Source Only	**44.3**	33.0	8.4	32.1	24.0	28.7	6.9	34.9	51.8	42.5	30.6
DAF [4]	37.5	32.8	10.2	**40.3**	27.2	31.3	4.1	41.0	55.5	52.0	33.2
MAF [14]	37.1	31.1	9.7	38.1	19.9	29.1	2.5	37.3	50.7	50.0	30.6
HTCN [3]	29.5	34.4	17.3	33.8	50.6	14.0	3.6	**46.9**	**74.7**	58.5	36.3
DAF*	32.0	31.7	**19.7**	30.5	49.5	18.5	7.5	40.0	60.1	54.0	34.4
UCF w/o WSE	32.6	35.7	3.1	28.2	49.6	32.5	**7.9**	45.5	62.5	56.1	35.4
UCF	32.4	**37.4**	4.0	33.5	**59.3**	**40.2**	4.4	39.0	58.1	**61.7**	**37.0**

Table 3. Results (%) on universal adaptation from PASCAL VOC to Clipart (ξ = 0.25). The source private classes include: *bus*, *car*, *cat*, *chair*, *cow*, *table* and *dog*. The target private classes include: *horse*, *motorbike*, *person*, *plant*, *sheep*, *sofa*, *train* and *tv*.

Methods	aero	bicycle	bird	boat	bottle	mAP
Source Only [30]	33.2	55.7	25.1	30.0	41.2	37.0
DAF [4]	31.5	42.5	25.2	**34.4**	50.8	36.9
MAF [14]	29.3	**57.0**	27.1	33.9	41.8	37.8
HTCN [3]	32.5	53.0	24.1	27.0	48.4	37.0
DAF*	33.4	46.7	25.5	33.8	46.5	37.2
UCF w/o WSE	28.2	45.4	**29.7**	32.2	54.8	38.1
UCF	**35.8**	52.9	28.6	20.8	**55.7**	**38.8**

5.2. Datasets

PASCAL VOC: PASCAL VOC [7] is a famous object detection dataset, which contains 20 categories. The image scale of the dataset is diverse. In our experiment, the training and validation splits of VOC07 and VOC12 are used as the training set, which results in about 15k images. We use the test set of VOC07 to evaluate the model.

Clipart and WaterColor: The Clipart and Watercolor [17] are constructed by the Amazon Mechanical Turk, which is introduced for the DAOD. Similar to the Pascal VOC, the Clipart contains 1000 images and 20 categories. WaterColor has 2000 images of 6 categories. Half of the datasets are introduced for training while the remaining is used for the test.

Cityscapes and FoggyCityscapes: The Cityscapes [5] dataset is collected from street scenes of 27 cities in normal weather, which contains 2,975 images for training and 500 images for validation. The Foggy Cityscapes [33] is derived from the Cityscapes [5] by manually adding fog to simulate the foggy weather, which shares the same annotations as the Cityscapes. Notably, the target labels are unavailable in training process.

5.3. Experimental Results

In this section, we set up five universal scenarios with different ξ on these datasets. We compare the proposed UCF with DAF [4] (baseline model), MAF [14], HTCN [3] and DAF*. MAF [14] proposes a multi-layer alignment structure and HTCN [3] designs multiple masks to improve transferability, which all perform well on closed sets and lead to later research. So it shall be valuable to study the performance of these methods in the practical UDAOD setting. DAF* means that we just simply add the weighting method based on sample prediction to DAF [4], inspired by UniDA [2, 32, 40]. Note that these models are compared under the same experimental setting, and all adopt ResNet-101 [13] as the backbone.

Table 4. Results (%) on universal adaptation from PASCAL VOC to WaterColor in partial scenario ($\mathcal{C}_s \supset \mathcal{C}_t$).

Methods	bicycle	bird	car	cat	dog	person	mAP
Source Only [30]	82.4	51.7	48.4	39.9	30.7	59.2	52.0
DAF [4]	73.4	51.9	43.1	35.6	28.8	63.1	49.3
MAF [14]	70.4	50.3	44.3	36.7	30.6	62.9	49.2
HTCN [3]	74.1	49.8	**51.9**	35.3	**35.3**	**66.0**	52.1
DAF*	81.2	50.4	45.1	35.6	33.9	63.9	51.7
UCF w/o WSE	84.4	50.8	49.2	33.5	31.4	65.1	52.4
UCF	**84.8**	**52.1**	49.8	**40.6**	33.8	63.2	**54.1**

Table 5. Results (%) on universal adaptation from WaterColor to PASCAL VOC in open set scenario ($\mathcal{C}_s \subset \mathcal{C}_t$).

Methods	bicycle	bird	car	cat	dog	person	mAP
Source Only [30]	29.8	50.2	47.1	**62.2**	51.5	57.8	49.8
DAF [4]	29.5	53.8	50.6	58.1	48.1	56.5	49.4
MAF [14]	28.5	50.0	46.8	59.4	50.2	58.6	48.9
HTCN [3]	26.4	43.0	46.5	50.8	44.0	53.9	44.1
DAF*	**36.2**	**54.7**	51.2	59.0	51.9	59.8	52.1
UCF w/o WSE	30.4	53.9	**53.8**	60.9	**55.7**	59.8	52.4
UCF	34.8	52.0	**53.8**	61.9	54.2	**60.5**	**52.9**

5.3.1 Transfer from PASCAL VOC to Clipart

Scenario Setting. In practical applications, the label space of the source domain is often unequal to that of the target domain. Both of dataset PASCAL VOC [7] and dataset Clipart [17] have 20 categories. Therefore, we use the PASCAL VOC as the source domain and the Clipart as the target domain, and we select some classes as the common classes or private classes. Specifically, we design three experiments for this scenario (from PASCAL VOC to Clipart) with different ξ: (1) ξ = 0.75, in which common classes $\mathcal{C}_\wedge$ are 15 classes, source private classes $\overline{\mathcal{C}}_s$ are 3 classes and target private classes $\overline{\mathcal{C}}_t$ are 2 classes; (2) ξ = 0.50, in which $\mathcal{C}_\wedge$ = 10 and both of $\overline{\mathcal{C}}_s$ and $\overline{\mathcal{C}}_t$ are 5; (3) ξ = 0.25, in which $\mathcal{C}_\wedge$ = 5, $\overline{\mathcal{C}}_s$ = 12 and $\overline{\mathcal{C}}_t$ = 13.

Results. The experimental results are shown in Table 1, 2 and 3 respectively. The experiments show that whatever this ξ is, our UCF can achieve state-of-the-art results among all compared methods. The proposed UCF clearly outperforms the baseline model [4] by +6.7%, +3.8%, and +1.9% with different ξ. Note that our UCF also can surpass the MAF [14] and HTCN [3], even if they have a multi-layer alignment structure and additional adaptation modules. And both source and target domains have their own private classes in these scenarios, which lead to more serious negative transfer. However, our model still performs well by using the proposed unbiased weighting mechanism.

5.3.2 Transfer from PASCAL VOC to WaterColor

Scenario Setting. We conduct the partial domain adaptive object detection scenario, in which the target label set is completely a subset of the source label set ($\mathcal{C}_s \supset \mathcal{C}_t$). WaterColor [17] dataset contains 6 categories in common with PASCAL VOC [7]. Therefore, we adopt the PASCAL VOC as the source domain and the WaterColor as the target domain in the partial domain adaptation. The 6 categories in WaterColor are the common classes $\mathcal{C}_\wedge$, and the remaining categories in the source domain are the source private classes $\overline{\mathcal{C}}_s$. In this scenario, there is no target private class $\overline{\mathcal{C}}_t$. Here the coincidence factor ξ = 0.30.

Results. Table 4 shows the results on the partial domain adaptive object detection scenario from PASCAL VOC [7] to WaterColor [17]. Specifically, our proposed UCF achieves a remarkable increase of +4.8% over the DAF [4]. Note that the DAF [4] achieves lower accuracy than the Source Only [30] without any adaptation (from 52.0% to 49.3%), which is mainly caused by the negative transfer from private classes. However, our UCF can mitigate the negative transfer and outperform the Source Only [30] (+2.1% mAP) due to the unbiased weighting mechanism, including CaM and WSE.

5.3.3 Transfer from WaterColor to PASCAL VOC

Scenario Setting. In this scenario, we conduct the experiment of open set domain adaptive object detection, inspired by [32], which requires that the source label set is completely a subset of the target label set ($\mathcal{C}_s \subset \mathcal{C}_t$). Therefore, the WaterColor [17] is used as the labeled source domain and the dataset PASCAL VOC [7] is used as the fully unlabeled target domain. We adopt the 6 categories in WaterColor [17] dataset as the common classes $\mathcal{C}_\wedge$, and the remaining 14 categories in PASCAL VOC [7] are the target private classes $\overline{\mathcal{C}}_t$. Here ξ = 0.30.

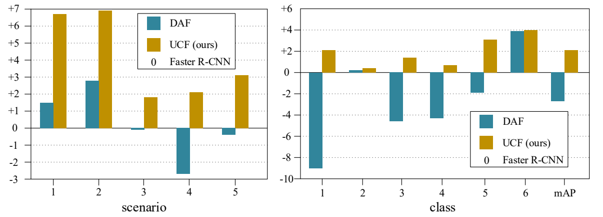
Results. As shown in Table 5, compared with the baseline model, the mAP of our UCF is improved by 3.5% (from 49.4% to 52.9%). Note that the MAF [14] and HTCN [3] perform worse than other methods, including Faster R-CNN [30]. The two methods adopt the multi-layer alignment structure. The negative transfer of private classes leads to the incorrect alignment in the shallow layer, which is bound to affect the learning for deep-layer features. As can be seen, the propoesd UCF can still achieve SOTA results without the WSE component.

5.3.4 The Closed Set Adaptation

Further, as shown in Table 6, we conduct the experiment on the closed set ($\mathcal{C}_s = \mathcal{C}_t$) from Cityscapes [5] to FoggyCityscapes [33] by comparing the two methods [4, 14]. Experimental result shows that our UCF outperforms the two methods, which indicates that our unbiased weighting

Table 6. Results (%) on adaptation from Cityscapes to FoggyCityscapes in closed set scenario ($\mathcal{C}_s = \mathcal{C}_t$). Note that UCF+ adopts a multi-layer structure for fair comparison with MAF [14].

Methods	prson	rider	car	trunk	bus	train	mcyc	bicy	mAP
Source Only [30]	17.8	23.6	27.1	11.9	23.8	9.1	14.4	22.8	18.8
DAF [4]	25.0	31.0	**40.5**	**22.1**	35.3	**20.2**	**20.0**	**27.1**	27.6
UCF	**29.2**	**38.4**	39.6	17.5	**36.2**	19.2	16.6	26.9	**28.0**
MAF [14]	**28.2**	39.5	43.9	**23.8**	**39.9**	33.3	**29.2**	33.9	34.0
UCF+	**28.2**	**42.0**	**45.1**	19.2	39.5	**34.6**	28.9	**36.1**	**34.2**



(a) Negative Transfer in Each Scenario

(b) Negative Transfer in Each Class

Figure 5. The negative transfer influence. (a) is the negative transfer including PASCAL-Clipart (ξ = 0.75, 0.50 and 0.25), PASCAL-WaterColor and WaterColor-PASCAL (they correspond to scenarios 1 to 5 successively). (b) is the negative transfer in each class on the scenario from PASCAL VOC to WaterColor.

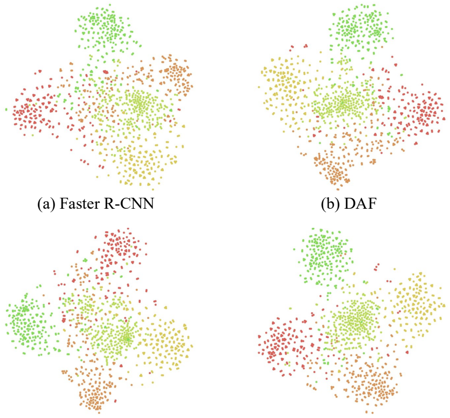
mechanism does not affect the performance of the network in the close-set scenario.

5.4. Further Empirical Analysis

Ablation Study. We conduct the ablation study by removing the Weight Surgery Equalization (WSE) in all scenarios. We can see that the proposed WSE is designed reasonably and when it is removed, the performance drops accordingly. And we compare our UCF with the DAF* that just simply adds the weighting method based on sample prediction into DAF [4], inspired by UniDA [2, 41, 40]. All the experimental results show that our method significantly exceeds DAF*. The reason is that the unbiased weighting mechanism we proposed is more reliable due to the central feature of each category, not the current batch samples that only belong to a subset of the entire label space.

Analysis for Negative Transfer. As shown in Figure 5, (a) is the performance change of DAF [4] and our UCF comparing to Faster R-CNN [30] (i.e., the Abscissa axis) with respect to each scenario. The adaptive module from the DAF [4] does not work in some scenarios, and even degrade performance. And (b) is the comparison in each class on the partial scenario, in which DAF has a significant negative transfer in most classes. However, the UCF can overcome the negative transfer in any scenario or class and improve the performance by our unbiased weighting mechanism.

Results of Visualization. We plot the t-SNE embeddings of the feature representations of the target domain in Figure 6. Figure 6 shows the results of Faster R-CNN [30], DAF [4], DAF*, and our UCF, where different color stand for different categories. We observe that features learned by Faster R-CNN, DAF, and DAF* are not clustered as clearly as our UCF, which may be caused by the negative transfer. Benefited by the unbiased weighing mechanism, including CaM and WSE, our UCF can suppress the negative transfer and provide a discriminative feature distribution.



(a) Faster R-CNN

(b) DAF

(c) DAF*

(d) UCF

Figure 6. The t-SNE plot of Faster R-CNN, DAF, DAF*, and our UCF for the PASCAL VOC to Clipart (ξ = 0.25) task on the target domain. Different colors stand for different categories.

6. Conclusion

In this paper, we introduce a novel Universal Domain Adaptive Object Detection (UDAOD) setting, which requires no prior knowledge on the label set of target domains. In order to address the UDAOD problem, we propose the Universal Cross-domain Faster R-CNN (UCF), which, to the best of our knowledge, is the first work deploying the universal domain adaptive problem in object detection. UCF aims to reduce the domain gap by aligning the feature in common label space. In order to overcome the class incompleteness of conventional UniDA, we introduce the Class-aware Memory (CaM) module with dynamically updated central features. Besides, Weight Surgery Equalization (WSE) is proposed to balance the weights provided by CaM. A thorough evaluation shows that existing methods requiring prior knowledge on the target label set cannot generalize well in the general UDAOD setting while the proposed UCF works stably and achieves SOTA results.

Acknowledgments

This work was partially supported by National Key R&D Program of China (2021YFB3100800), National Natural Science Fund of China (62271090), Chongqing Natural Science Fund (cstc2021jcyj-jqX0023). This work is also supported by Huawei computational power of Chongqing Artificial Intelligence Innovation Center.

References

[1] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *CVPR*, pages 6154–6162, 2018.

[2] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *ECCV*, pages 135–150, 2018.

[3] Chaoqi Chen, Zebiao Zheng, Xinghao Ding, Yue Huang, and Qi Dou. Harmonizing transferability and discriminability for adapting object detectors. In *CVPR*, pages 8869–8878, 2020.

[4] 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.

[5] 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.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. Ieee, 2009.

[7] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 88(2):303–338, 2010.

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

[9] Ross Girshick. Fast r-cnn. In *ICCV*, December 2015.

[10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, pages 580–587, 2014.

[11] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, pages 2961–2969, 2017.

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

[14] Zhenwei He and Lei Zhang. Multi-adversarial faster-rcnn for unrestricted object detection. In *ICCV*, pages 6668–6677, 2019.

[15] Cheng-Chun Hsu, Yi-Hsuan Tsai, Yen-Yu Lin, and Ming-Hsuan Yang. Every pixel matters: Center-aware feature alignment for domain adaptive object detector. In *ECCV*, pages 733–748. Springer, 2020.

[16] 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 *WACV*, pages 749–757, 2020.

[17] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *CVPR*, pages 5001–5009, 2018.

[18] Taekyung Kim, Minki Jeong, Seunghyeon Kim, Seokeon Choi, and Changick Kim. Diversify and match: A domain adaptive representation learning paradigm for object detection. In *CVPR*, pages 12456–12465, 2019.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[20] TY Lin, M Maire, S Belongie, J Hays, P Perona, D Ramanan, and CL Zitnick. Microsoft coco: Common objects in context. *ECCV*, pages 740–755, 2014.

[21] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017.

[22] Feng Liu, Xiaosong Zhang, Fang Wan, Xiangyang Ji, and Qixiang Ye. Domain contrast for domain adaptive object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[23] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *ECCV*, pages 21–37. Springer, 2016.

[24] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *ICML*, pages 97–105. PMLR, 2015.

[25] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. *arXiv preprint arXiv:1602.04433*, 2016.

[26] Dang-Khoa Nguyen, Wei-Lun Tseng, and Hong-Han Shuai. Domain-adaptive object detection via uncertainty-aware distribution alignment. In *ACM MM*, pages 2499–2507, 2020.

[27] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. In *ICCV*, pages 754–763, 2017.

[28] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, pages 779–788, 2016.

[29] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[30] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.

[31] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In *CVPR*, June 2019.

[32] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *ECCV*, pages 153–168, 2018.

[33] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018.

[34] Zhiqiang Shen, Zhuang Liu, Jianguo Li, Yu-Gang Jiang, Yurong Chen, and Xiangyang Xue. Dsod: Learning deeply supervised object detectors from scratch. In *ICCV*, pages 1919–1927, 2017.

[35] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[36] Keyang Wang and Lei Zhang. Single-shot two-pronged detector with rectified iou loss. In *ACM MM*, pages 1311–1319, 2020.

[37] Yu Wang, Rui Zhang, Shuo Zhang, Miao Li, YangYang Xia, XiShan Zhang, and ShaoLi Liu. Domain-specific suppression for adaptive object detection. In *CVPR*, pages 9603–9612, 2021.

[38] Minghao Xu, Hang Wang, Bingbing Ni, Qi Tian, and Wenjun Zhang. Cross-domain detection via graph-induced prototype alignment. In *CVPR*, pages 12355–12364, 2020.

[39] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *CVPR*, pages 2272–2281, 2017.

[40] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *CVPR*, pages 2720–2729, 2019.

[41] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona. Importance weighted adversarial nets for partial domain adaptation. In *CVPR*, pages 8156–8164, 2018.

[42] Yangtao Zheng, Di Huang, Songtao Liu, and Yunhong Wang. Cross-domain object detection through coarse-to-fine feature adaptation. In *CVPR*, pages 13766–13775, 2020.

[43] Chenchen Zhu, Yihui He, and Marios Savvides. Feature selective anchor-free module for single-shot object detection. In *CVPR*, pages 840–849, 2019.