Active Domain Adaptation with False Negative Prediction for Object Detection

Yuzuru Nakamura[1] Yasunori Ishii[1] Takayoshi Yamashita[2]
[1]Panasonic Holdings Corporation [2]Chubu University
{nakamura.yuzuru,ishii.yasunori}@jp.panasonic.com takayoshi@isc.chubu.ac.jp

Abstract

Domain adaptation adapts models to various scenes with different appearances. In this field, active domain adaptation is crucial in effectively sampling a limited number of data in the target domain. We propose an active domain adaptation method for object detection, focusing on quantifying the undetectability of objects. Existing methods for active sampling encounter challenges in considering undetected objects while estimating the uncertainty of model predictions. Our proposed active sampling strategy addresses this issue using an active learning approach that simultaneously accounts for uncertainty and undetectability. Our newly proposed False Negative Prediction Module evaluates the undetectability of images containing undetected objects, enabling more informed active sampling. This approach considers previously overlooked undetected objects, thereby reducing false negative errors. Moreover, using unlabeled data, our proposed method utilizes uncertainty-guided pseudo-labeling to enhance domain adaptation further. Extensive experiments demonstrate that the performance of our proposed method closely rivals that of fully supervised learning while requiring only a fraction of the labeling efforts needed for the latter.

1. Introduction

Significant differences in appearance due to factors such as lighting conditions and sensors, as observed in in-vehicle cameras, demand substantial annotation for each dataset, resulting in prolonged model deployment. Domain adaptation (DA) is an effective technique in such situations, allowing models to adapt to various scenes with different appearances. Unsupervised domain adaptation (UDA) that deals with unlabeled target domain data has been widely studied [2–4, 9, 14, 23, 35, 55, 60, 61]. UDA is an optimal approach for reducing annotation costs. However, a considerable performance gap persists between UDA and fully supervised learning, wherein all data are labeled.

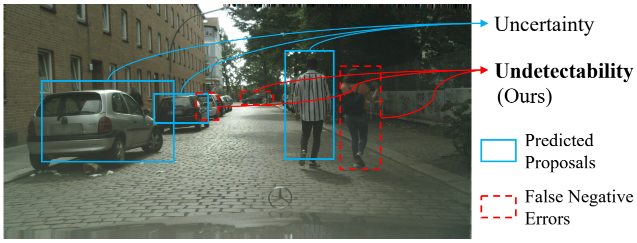

Figure 1. **Conceptual diagram of the conventional uncertainty and our proposed undetectability metrics used for the acquisition function.** Uncertainty is only estimated from predicted bounding boxes (represented by blue rectangles). In contrast, undetectability is estimated from false negative (FN) errors that the detection model cannot predict (represented by red dotted rectangles). Our proposed method performs active sampling by utilizing both uncertainty and undetectability.

Active learning (AL) [5, 10, 16, 22, 24, 26, 30, 42, 45, 46, 53, 56, 58] is a method to select a few effective samples for training, achieving high accuracy with a limited number of labeled data. However, most AL methods select samples from a single pool, assuming an independent and identically distributed (*i.i.d.*) dataset. Consequently, under domain shift, AL fails to select samples capable of enhancing performance.

Active domain adaptation (ADA) is a solution for effective sample selection under domain shift. This method has been actively studied [1, 7, 11, 18, 19, 21, 25, 29, 31, 33, 38, 40, 44, 47–51, 57, 59] and, with a limited labeling budget, achieves accuracy comparable to fully supervised learning [38]. Most ADA studies focus on image classification and semantic segmentation tasks, whereas relatively less research has been conducted on applying this method to an object detection task. Classification and segmentation tasks have a common feature: they can be regarded as classifying tasks for an image or pixel. In contrast, an object detection task predicts the location and category of an object bounding box. In ADA, classification uncertainty is taken as a criterion for sample selection. Therefore, even if ADA, designed for classification and segmentation, is applied to object detection, its contribution pertains only to classification, not the prediction or localization of the object region.

Contributions. In this paper, we propose an ADA method designed for object detection. First, we analyzed the causes of performance degradation under domain shift to identify the main challenge in DA for object detection. The analysis showed that, under domain shift, the increase in false negative (FN) errors in the target domain considerably impacts performance.

Existing AL methods for object detection estimate the uncertainty of predicted object proposals, excluding undetected objects from predictions, rendering uncertainty estimation impossible. Hence, designing an AL method that accounts for undetected objects becomes crucial for enhancing object detection performance under domain shift.

We propose an active sampling strategy integrating undetectability into the acquisition function to incorporate the aspect of undetected objects in sample selection (Figure 1). We introduce a framework and model, predicting the possibility of FN errors in images. We call this model the False Negative Prediction Module (FNPM). FNPM assesses undetectability for each image, actively selecting images that contain numerous undetected objects, thereby training the object detection model to be robust against FN errors.

Our proposed method initializes the model with UDA training. As the estimation by the acquisition function becomes inaccurate without any DA under domain shift, feature alignment is performed between source and target domains by adversarial learning [13]. Subsequently, active sampling using our acquisition function is performed, followed by training in a semi-supervised DA manner. This training includes a few labeled target domain data in addition to source and unlabeled target domain data. In addition to feature alignment, our method filters pseudo-labels by utilizing the localization uncertainty of predicted bounding boxes. A diverse range of experimental scenarios demonstrates that our proposed method achieves performance comparable to fully supervised learning, requiring only a few percent of the labeling budget. Consequently, our method, at a low labeling cost, outperforms previous UDA methods.

Our contributions can be summarized as follows:

- We proposed an ADA method designed for object detection. To the best of our knowledge, this is the first ADA method designed for object detection.
- Based on the performance analysis under domain shift, we identified the issue of undetected objects. Furthermore, we proposed the FNPM focusing on undetected objects.
- We experimentally demonstrated that even with a low labeling budget, the performance of our proposed method approaches that of fully supervised learning.

2. Related Works

Domain Adaptive Object Detection methods adapt models to various scenes with different appearances. In particular, unsupervised domain adaptation (UDA), which does not use target domain labels, has been widely studied [2–4, 14, 23, 35, 55, 61]. UDA for object detection is frequently based on adversarial learning or self-training approaches.

Adversarial learning is a method that aligns across domains in a feature space using the gradient reversal layer and domain discriminator [13]. Adaptive Teacher [23] combines domain alignment and self-training to enhance pseudo-labeling accuracy under domain shift. MGADA [60] aligns domains at various levels, such as pixel, instance, and category, through integrated multi-granularity alignment. The self-training method provides unlabeled target domain data with pseudo-labels for subsequent training. Probabilistic Teacher [3] dynamically filters out noisy pseudo-labels attributed to uncertainty. Unbiased Mean Teacher [9] uses a generative model to transform target domain images into a source-like style, thus improving the accuracy of pseudo-labeling.

Despite extensive studies, UDA's performance is substantially deficient compared to that of fully supervised learning. Conversely, the performance of our proposed method approaches fully supervised learning while using only a few percent of the latter approach's labeling cost.

Active Learning (AL) for Object Detection [5, 10, 22, 24, 26, 30, 42, 45, 46, 53, 56, 58] methods aim to select effective samples for training within a limited labeling budget. Recently, a semi-supervised learning method using labeled and unlabeled data was proposed. Active Teacher [26] utilizes a teacher-student structure to generate pseudo-labels from unlabeled data and incorporates them into the student model for training. Elezi et al. [10] proposed a method that switches between self-training with pseudo-labels or inconsistency-based training based on prediction confidence determined by an acquisition function.

These AL methods sample data from a single pool, and the distribution of the sampled data is assumed to be the *i.i.d.* dataset. Therefore, the ability to sample data suitable for performance improvement is limited under domain shift.

Active Domain Adaptation (ADA) focuses on selecting effective samples for training under domain shift. This approach has been actively studied recently [1, 7, 11, 18, 19, 21, 25, 29, 31, 33, 38, 40, 44, 47–51, 57, 59]. LabOR [38] achieves accuracy comparable to fully supervised learning on a budget of a few percent of the latter in the semantic segmentation task. This approach is made possible by annotating pixels where uncertainty exists under domain shift. AADA [40] uses domain discriminator-predicted results for sampling as a diversity metric combined with uncertainty.

Previous methods mainly focus on an image classification task, and some have been applied experimentally to

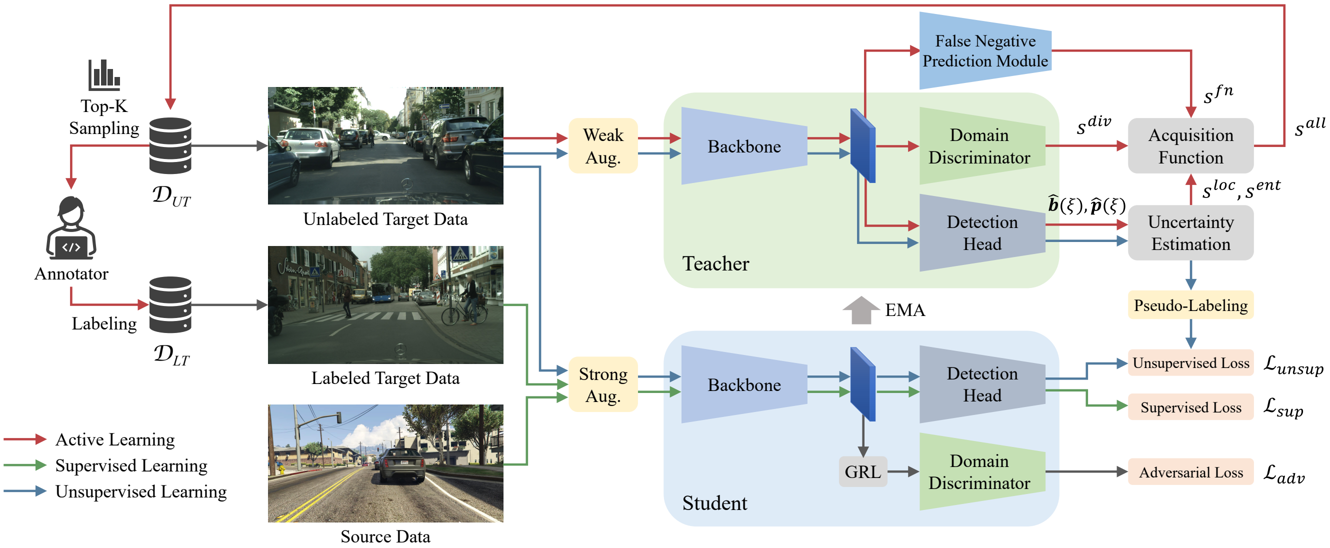

Figure 2. **Overview of our proposed method.** We propose an active domain adaptation (ADA) framework incorporating the domain adaptation (DA) and active sampling processes with the False Negative Prediction Module (FNPM). In the active sampling process, the FNPM estimates the metrics of undetectability, and our acquisition function measures the scores by considering undetectability in addition to conventional uncertainty and diversity. We then sample unlabeled target domain data based on these scores and label them. Pseudo-labels are assigned to the unlabeled target data, which are then used for training in the semi-supervised DA process.

an object detection task [40]. However, the contribution of these methods to object detection is limited because object detection involves numerous other factors besides the uncertainty of category prediction. For example, previous methods cannot estimate the uncertainty of undetected objects. Furthermore, category prediction and bounding box localization are crucial for object detection. Hence, we propose an ADA method suitable for object detection to address these issues.

3. Method

Notation. We have a set of labeled source data $\mathcal{D}_S = \{(\boldsymbol{x}_i^S, \boldsymbol{y}_i^S)\}_{i=1}^{N_S}$ and unlabeled target data $\mathcal{D}_T = \{\boldsymbol{x}_i^T\}_{i=1}^{N_T}$, where $\boldsymbol{x}_i \in \mathbb{R}^{W \times H \times 3}$ is the i-th image of width W and height H in a dataset, and $\boldsymbol{y}_i = \{\boldsymbol{b}_{i,j}, c_{i,j}\}_{j=1}^{N_{bbox}^i}$ consists of the j-th bounding box coordinates $\boldsymbol{b}_{i,j} \in \{x, y, w, h\}$ and category index $c_{i,j} \in \{1, ..., N_c\}$. $\mathcal{D}_T$ consists of a set of labeled target data $\mathcal{D}_{LT} = \{(\boldsymbol{x}_i^{LT}, \boldsymbol{y}_i^{LT})\}_{i=1}^{N_{LT}}$ and unlabeled target data $\mathcal{D}_{UT} = \{\boldsymbol{x}_i^{UT}\}_{i=1}^{N_{UT}}$. Given $\mathcal{D}_{LT} = \emptyset$ at the beginning of training, we sample images from $\mathcal{D}_{UT}$ that maximize performance through the acquisition function within a labeling budget. Then, we annotate these images and incorporate them into $\mathcal{D}_{LT}$.

3.1. Overview

Figure 2 shows the overall framework of our proposed method. Our proposed method involves the domain adaptation (DA) and active learning (AL) processes. The conventional uncertainty-based AL methods determine whether or not to sample predicted objects. However, these methods do not define any criteria for sampling undetected objects. Therefore, conventional AL methods cannot sufficiently improve the performance for false negative (FN) errors. In this situation, we propose an active domain adaptation (ADA) method for object detection using the False Negative Prediction Module (FNPM) to predict the number of FN objects. Our proposed acquisition function outputs a metric from the FNPM's output value, uncertainty, and diversity.

Our proposed method involves three steps: (1) Initializing the model with $(\mathcal{D}_S, \mathcal{D}_T)$ in UDA training (Section 3.2), (2) active sampling within the budget from $\mathcal{D}_{UT}$ with our acquisition function and incorporating the samples into $\mathcal{D}_{LT}$ after labeling (Section 3.3), and (3) training the model with labeled data $\mathcal{D}_S \cup \mathcal{D}_{LT}$ and unlabeled data $\mathcal{D}_{UT}$ in a semi-supervised DA manner (Section 3.4). Five rounds of steps (2) and (3) are performed. The detailed procedure is shown in Algorithm 1.

3.2. Model Initialization

We first initialize the common parameters for student and teacher models with $(\mathcal{D}_S, \mathcal{D}_T)$ in a UDA training manner. Active sampling is based on the metrics measured by the acquisition function for the target domain data. However, if the model used for the acquisition function does not have knowledge of the target domain, the acquisition function cannot correctly measure the metrics for sampling. Therefore, we first train the model adapted to the target domain using UDA. Specifically, we perform feature-level

Algorithm 1: Training procedure of proposed method

```
Input: Source data $\mathcal{D}_S$, Target data $\mathcal{D}_T = \{\mathcal{D}_{LT}, \mathcal{D}_{UT}\}$,
       where $\mathcal{D}_{LT} = \emptyset$
Output: Model parameters $\theta_t, \theta_s$ adapted on target
        domain
begin
    Pre-train the student model $\theta_s, \phi_s$ based on Eq. (1)
    $\theta_t \leftarrow \theta_s, \phi_t \leftarrow \phi_s$
    for $R$ Rounds do
        Freeze $\theta_t$ and train FNPM $\psi$ based on Eq. (4)
        for $x^{UT} \in \mathcal{D}_{UT}$ do
            Calculate the acquisition score $s^{all}$ based on
             Eq. (11)
        end for
        Select the top-K data $\mathcal{D}_{active}$ and annotate them
        $\mathcal{D}_{LT} \leftarrow \mathcal{D}_{LT} \cup \mathcal{D}_{active}$
        $\mathcal{D}_{UT} \leftarrow \mathcal{D}_{UT} \setminus \mathcal{D}_{active}$
        Unfreeze $\theta_t$
        Train the student model $\theta_s, \phi_s$ based on Eq. (12)
        Update the teacher model $\theta_t, \phi_s$ based on
         Eq. (15)
    end for
end
```

alignment across domains using adversarial learning with the gradient reversal layer (GRL) and domain discriminator [13].

Given the detection model and domain discriminator parameters θ_s and ϕ_s, respectively, of the student model, the objective loss function is defined as follows:

$$\min_{\theta_s} \max_{\phi_s} \mathcal{L}_{init} = \mathcal{L}^S_{sup} + \lambda \mathcal{L}_{adv}, \tag{1}$$

where λ is the weight of $\mathcal{L}_{adv}$. $\mathcal{L}^S_{sup}$ is the supervised loss in the source domain defined as follows:

$$\begin{aligned} \mathcal{L}_{sup} =& \mathcal{L}^{rpn}_{cls}(\boldsymbol{x}_i, \boldsymbol{c}_i) + \mathcal{L}^{rpn}_{reg}(\boldsymbol{x}_i, \boldsymbol{b}_i) \\ &+ \mathcal{L}^{roi}_{cls}(\boldsymbol{x}_i, \boldsymbol{c}_i) + \mathcal{L}^{roi}_{reg}(\boldsymbol{x}_i, \boldsymbol{b}_i), \end{aligned} \tag{2}$$

where $\mathcal{L}^{rpn}_{cls}$ and $\mathcal{L}^{rpn}_{reg}$ are the classification loss and regression loss, respectively, in the region proposal network (RPN). $\mathcal{L}^{roi}_{cls}$ and $\mathcal{L}^{roi}_{reg}$ are the classification loss and regression loss, respectively, in the region of interest (RoI) head.

The adversarial loss $\mathcal{L}_{adv}$ is defined as follows:

$$\begin{aligned} \mathcal{L}_{adv} =& -\log(1 - D(F_{enc}(\boldsymbol{x}^S_i; \theta_s); \phi_s)) \\ &- \log D(F_{enc}(\boldsymbol{x}^T_i; \theta_s); \phi_s), \end{aligned} \tag{3}$$

where F_{enc} and D are the backbone of the detection model and the domain discriminator, respectively. We train the domain discriminator to discriminate the source domain as 1 and the target domain as 0.

After training with UDA on all the data in $(\mathcal{D}_S, \mathcal{D}_T)$, the parameters of the student model are copied to the parameters (θ_t, ϕ_t) of the teacher model $(\theta_t \leftarrow \theta_s, \phi_t \leftarrow \phi_s)$, and then, we proceed to the active sampling step.

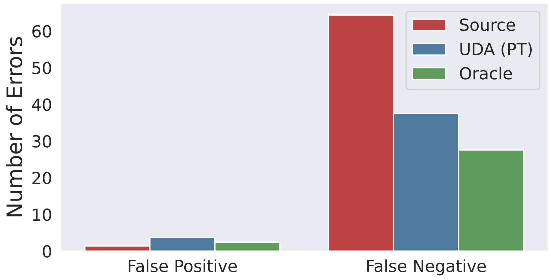

Figure 3. **Comparison of the number of false positive (FP) errors and FN errors in unsupervised domain adaptation (UDA) from KITTI to Cityscapes.** We used Probabilistic Teacher [3] as the UDA method in this analysis. Even if UDA reduced the number of FN errors, the number remained larger than that of FP errors.

3.3. Active Learning Based on False Negatives

3.3.1 Analysis of Missed Detection under Domain Shift

First, we analyzed the causes of performance degradation under domain shift to clarify the primary challenge in applying UDA for object detection. Figure 3 shows a comparison between false positive (FP) errors and FN errors in UDA. When comparing Source-only (no adaptation) with Oracle, the gap in FP errors was 1 pt, whereas, for FN errors, it was more than double. This result shows that FN errors are the main cause of performance degradation under domain shift. Although the UDA method remarkably reduced FN errors compared with Source-only, the gap in FN errors remained larger than in FP errors. Therefore, FN errors persist as the main cause of performance degradation decline under domain shift, indicating the potential for further performance improvement if this issue is resolved.

The results indicate that AL should focus on sampling images containing undetected objects to effectively reduce FN errors under domain shift. However, many AL methods use predicted objects to assess uncertainty when sampling data. Therefore, uncertainty cannot be estimated for undetected objects that do not appear in the detection proposals. Hence, we propose an additional metric for the acquisition function, specifically the number of FN errors per image. This metric allows sampling to consider the uncertainty and undetectability of objects.

3.3.2 False Negative Prediction Module

The FNPM predicts the number of FN errors for an image, intuitively representing the difficulty of detecting objects. Moreover, as the FNPM outputs the count of FN errors, selected samples based on FNPM metrics contain more informative data than other samples. By integrating FNPM predictions into the acquisition function, we can quantify the undetectability of objects and actively select samples that

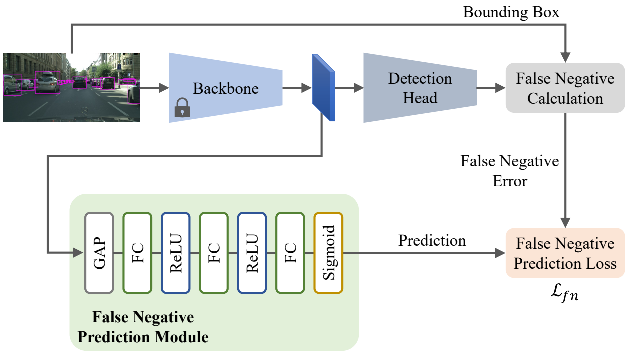

Figure 4. **Architecture of the FNPM.** The FNPM predicts the number of FN errors. We train the FNPM to predict the number of FN errors using the domain-adapted backbone.

contribute to reducing FN errors.

Since various factors can cause FN errors and deterministic approaches exhibit poor prediction accuracy, deep neural networks (DNNs) based prediction methods were proposed [32, 52]. We follow these approaches and design a branch that predicts the number of FN errors using a DNNs regression model. Figure 4 shows the architecture of the FNPM. The FNPM receives the output feature map of the backbone of the detection model and feeds this map to the global average pooling (GAP) and fully connected (FC) layers. Subsequently, the FNPM outputs predictions of the number of FN errors. As corresponding detection results are obtained from the detection model, the ground truth of the number of FN errors can be calculated. The FNPM is trained to minimize errors using the loss function, which is defined as follows:

$$\begin{aligned}\mathcal{L}_{fn} =&(G(F_{enc}(\boldsymbol{x}_i^S;\theta_t);\psi) - \mathcal{FN}(F_{head}(\boldsymbol{x}_i^S;\theta_t),\boldsymbol{y}_i^S))^2\\ &+(G(F_{enc}(\boldsymbol{x}_i^{LT};\theta_t);\psi) - \mathcal{FN}(F_{head}(\boldsymbol{x}_i^{LT};\theta_t),\boldsymbol{y}_i^{LT}))^2,\end{aligned} \tag{4}$$

where G, ψ, and F_{head} are the FNPM, its parameters, and the head of the detection model, respectively. $\mathcal{FN}(\cdot,\cdot)$ calculates the number of FN errors in the detection results by determining the number of ground truths $\boldsymbol{y}$ not assigned a detection result $F_{head}(\boldsymbol{x};\theta)$ of the same category with intersection over union (IoU) above a threshold.

During the active sampling process, although the FNPM predicts the number of FN errors for the unlabeled target domain, the challenge is that only a few labeled data in the target domain are available for training the FNPM. To leverage labeled data in the source domain, we use the adapted backbone of the detection model to extract domain-invariant features and predict these features for the target domain. Furthermore, we can reduce the number of additional parameters by using a common feature extractor.

Since the ground truths of the FN errors are calculated from the prediction results of the detection model, updating the detection model also alters the ground truths, complicating the stable training of the FNPM. Inspired by the reinforcement learning method [27, 28], we train the detection model and the FNPM alternately. As the FNPM is used only during the active sampling process, it remains unaltered during the training of the detection model. Before active sampling, we freeze the parameters of the detection model and update only the FNPM. This approach makes it possible to optimize both the detection model and FNPM simply and stably.

3.3.3 Uncertainty Estimation with MCDropout

Subsequently, we discuss the measurement of uncertainty in cases when the detection model identifies the objects in an image. Many object detection methods measure uncertainty based on class probability in a manner similar to that adopted in object recognition tasks. However, in object detection, the detection model should stabilize not only the class probabilities but also the positions of the objects. Therefore, we use variational inference with Monte Carlo Dropout (MCDropout) [12] to consider the parameters of the detection model as a probability distribution. We quantify the variance in localization predictions due to model perturbations and use it as a metric for uncertainty.

We incorporate the MCDropout layer into the detection head and reformulate the predicted bounding box coordinates $\hat{\boldsymbol{b}}_i$ and class probabilities $\hat{\boldsymbol{p}}_i$ as follows:

$$\{\hat{\boldsymbol{b}}_i(\xi),\hat{\boldsymbol{p}}_i(\xi)\} = F_{head}(\boldsymbol{x}_i;\theta_t,\xi), \text{ where } \xi \sim Ber(\eta), \tag{5}$$

where $Ber(\eta)$ is Bernoulli distribution of dropout rate η.

Variational inference allows us to obtain multiple predictions sampled under a pseudo-probability distribution. We use these means $(\hat{\boldsymbol{b}}_i^{mean},\hat{\boldsymbol{p}}_i^{mean})$ as the prediction results considering model perturbations and the variance of the predicted coordinates $\hat{\boldsymbol{b}}_i^{var}$ as the localization uncertainty. These means and variances are calculated from $\hat{\boldsymbol{b}}_{i,m} \sim \hat{\boldsymbol{b}}_i(\xi)$, $\hat{\boldsymbol{p}}_{i,m} \sim \hat{\boldsymbol{p}}_i(\xi)$ when inferred M times.

3.3.4 Acquisition Function

We propose an active sampling strategy that scores a combination of four metrics, including undetectability and localization uncertainty. In the following text, we explain each metric and the final combination of the metrics.

Undetectability. We estimate the undetectability of the detection model for images using the FNPM (Section 3.3.2). The higher the value, the more difficult the sample to predict and the more informative the sample. The undetectability metric is defined as follows:

$$s_i^{fn} = G(F_{enc}(\boldsymbol{x}_i^{UT};\theta_t);\psi). \tag{6}$$

Localization Uncertainty. We quantify the variation in the predicted coordinates of bounding boxes by variational inference (Section 3.3.3). The localization uncertainty metric is defined as follows:

$$s_i^{loc} = \frac{1}{4N^i_{bbox}} \sum_{j=1}^{N^i_{bbox}} \sum_{k\in\{x,y,w,h\}} \hat{b}^{var}_{i,j,k}. \tag{7}$$

Classification Uncertainty. We use the entropy [43] of class probabilities to estimate the uncertainty in classification. Higher entropy is assumed to be useful for training because the sample is difficult to classify. The classification uncertainty metric is defined as follows:

$$s_i^{ent} = -\frac{1}{N^i_{bbox}} \sum_{j=1}^{N^i_{bbox}} \sum_{k=1}^{N_c} \hat{p}^{mean}_{i,j,k} \log \hat{p}^{mean}_{i,j,k}. \tag{8}$$

Diversity. We use the metric based on the idea that a high density of the target domain is more critical under domain shift. Following [40], we use the domain discriminator to estimate samples that better represent the distribution of the target domain. The diversity metric is defined as follows:

$$s_i^{div} = \frac{1 - D(F_{enc}(\boldsymbol{x}_i^{UT};\theta_t);\phi_t)}{D(F_{enc}(\boldsymbol{x}_i^{UT};\theta_t);\phi_t)}. \tag{9}$$

Final Metric. These four metrics are used to calculate the final metric for each image. However, the metrics have different ranges of values, and the metrics with a large scale may become dominant. Therefore, we normalize each metric based on the following expression:

$$\hat{s}_i^l = max(0, \frac{s_i^l - (\mu(\boldsymbol{s}^l) - 3\sigma(\boldsymbol{s}^l))}{6\sigma(\boldsymbol{s}^l)}), \tag{10}$$

where $l \in \{fn, loc, ent, div\}$, and $\mu(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation, respectively. As each metric can be assumed unimodal in diverse natural images, it is scaled by 6σ based on a normal distribution.

The final metric is calculated as the product of the individual metrics as follows:

$$s_i^{all} = \hat{s}_i^{fn}\, \hat{s}_i^{loc}\, \hat{s}_i^{ent}\, \hat{s}_i^{div}. \tag{11}$$

3.4. Semi-Supervised Domain Adaptation

We use a semi-supervised learning framework to train the model because incorporating a large number of unlabeled data and a few labeled data can drastically enhance the training process and improve model accuracy. We conduct supervised learning on source and labeled target domain data and unsupervised learning on unlabeled target domain data with uncertainty-guided pseudo-labeling to perform semi-supervised learning. We use pseudo-labels with low localization uncertainty to achieve more accurate bounding box localization.

The objective loss function is defined as follows:

$$\min_{\theta_s}\max_{\phi_s} \mathcal{L}_{total} = \mathcal{L}^S_{sup} + \mathcal{L}^{LT}_{sup} + \mathcal{L}_{unsup} + \lambda\mathcal{L}_{adv}, \tag{12}$$

where $\mathcal{L}^{LT}_{sup}$ is the supervised loss in the labeled target domain, similar to Eq. (2). $\mathcal{L}_{unsup}$ is the unsupervised loss in the unlabeled target domain defined as follows:

$$\begin{aligned}\mathcal{L}_{unsup} =& \frac{1}{N^i_{bbox}} \sum_{j=1}^{N^i_{bbox}} \mathbb{I}_{bbox}(\hat{\boldsymbol{b}}^{var}_{i,j}) \\ & [\mathcal{L}^{rpn}_{cls}(\boldsymbol{x}_i'^{UT}, c^{PL}_{i,j}) + \mathcal{L}^{roi}_{cls}(\boldsymbol{x}_i'^{UT}, c^{PL}_{i,j})],\end{aligned} \tag{13}$$

where $c^{PL}_{i,j}$ is the pseudo-label and $\mathbb{I}_{bbox}(\cdot)$ is the indicator function defined as follows:

$$\mathbb{I}_{bbox}(\hat{\boldsymbol{b}}^{var}_{i,j}) = \begin{cases} 1, & \text{if } \frac{1}{4} \sum_{k\in\{x,y,w,h\}} \hat{b}^{var}_{i,j,k} \leq \gamma \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

where γ is the threshold for using pseudo-labels with variance less than a specific value.

Finally, after the student model is updated once by Eq. (12), the teacher model is updated by using the exponential moving average (EMA) [41]:

$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s, \quad \phi_t \leftarrow \alpha\phi_t + (1-\alpha)\phi_t, \tag{15}$$

where α is the update ratio.

4. Experiments

4.1. Experimental Settings

4.1.1 Datasets

We evaluated our proposed method in four domain adaptation scenarios using five datasets. The adaptation scenarios from domain X to Y are represented as X → Y.

Cityscapes (C) [6] is a clear-weather scene dataset captured using in-vehicle cameras in urban areas. It contains 2,975 training and 500 validation images, with bounding boxes derived from instance segmentation masks.

Foggy Cityscapes (F) [36] is a pseudo-dense foggy dataset generated from Cityscapes, maintaining the same number of images. There are three levels of fog density (0.005, 0.01, and 0.02). We used a split of 0.02 fog density level.

BDD100k (B) [54] is a large-scale dataset of in-vehicle camera images and contains 100k images. We used a daytime subset of this dataset, 36,728 training images and 5,258 validation images.

SIM10k (S) [20] is a dataset of 10,000 images synthesized by the game engine. We used all 10,000 images in training, including 58,071 bounding boxes.

KITTI (K) [15] is a dataset of in-vehicle cameras captured in a different scene from Cityscapes. We used 7,481 images for training.

4.1.2 Implementation Details

Network Architecture. We used Faster-RCNN [34] as the detection model, utilizing a BatchNorm-free VGG16 [39] backbone pre-trained on ImageNet [8] following [3]. The architecture of the domain discriminator followed that used in [17].

Pre-processing. The image size was resized to 600 on the shorter side while maintaining the aspect ratio constant. The transformation of strong and weak augmentation proceeded as follows [3].

Optimization. For the detection models, we used SGD with a momentum of 0.9, a weight decay of 10^{-4}, an initial learning rate of 0.02 with a warm-up, training for 40k iterations, and reducing the learning rate by ten at 30k and 35k iterations. For the FNPM, we used SGD with an initial learning rate of 10^{-4}, training for 2k iterations with a cosine annealing scheduler before active sampling.

Training Configuration. The batch size was set to four for each domain data. Five rounds of active sampling were performed at 5k, 10k, 15k, 20k, and 25k iterations. The variance threshold γ was set to 0.1, dropout rate η to 0.1, number of variational inferences M to 10, EMA rate α to 0.9996, and weight of adversarial loss λ to 0.01.

Evaluation Metrics. We evaluated our method with the average precision (AP) at the intersection over union (IoU) threshold of 0.5. In the **C** → **F** and **C** → **B** scenarios, we report the mean AP (mAP) over all classes.

4.2. Comparison with Domain Adaptation Methods

Table 1 shows the quantitative results of comparing our proposed method with the state-of-the-art unsupervised domain adaptation (UDA) and active domain adaptation (ADA) methods. Row "Source-only" and "Oracle" display the results of supervised learning using the source domain dataset and fully labeled target domain dataset, respectively. Among previous ADA methods, we reproduced AADA as it focuses on object detection. Our proposed method outperformed state-of-the-art UDA methods in most cases while requiring only 1% of the labeling budget. Particularly, in the **K** → **C** scenario, the performance improved by more than 10 pt with only a 1% increase in the labeling budget. Furthermore, with 5% of the labeling budget, our method achieved nearly the same performance as Oracle, displaying an accuracy difference ratio of 98%.

For a fair comparison of our method with the UDA methods, we evaluated the model under 0% of the labeling budget setting by solely using pseudo-labels with low localization uncertainty for training without active sampling. The evaluation results show that our method is competitive with the state-of-the-art methods in the UDA setting. Therefore, our proposed method can also provide a strong UDA baseline. Our proposed method achieved the highest performance among all the methods in the **C** → **B** scenario.

Table 1. **Comparison with the state-of-the-art UDA and ADA methods.** Even with only 1% of the labeling budget (Ours (1%)), the performance of our proposed method exceeded that of almost all conventional methods in terms of accuracy. Furthermore, our proposed method achieved the performance nearly equivalent to that of Oracle while using 5% of the labeling budget (Ours (5%)). AADA† represents an evaluation through re-implementation.

Methods		**C** → **F**	**C** → **B**	**S** → **C**	**K** → **C**
Source-only		15.5	26.8	40.3	31.3
UDA	DA-Faster [4]	27.6	—	39.0	38.5
	SWDA [35]	34.3	—	47.7	—
	PT [3]	42.7	34.9	55.1	60.2
	AT [23]	50.9	—	—	—
	MGADA [60]	44.3	—	49.8	45.2
	CMT [2]	50.3	—	—	64.3
	CSDA [14]	45.0	—	56.9	48.6
	NSA-UDA [61]	**52.7**	35.5	56.3	55.6
	Ours (0%)	52.0	36.3	55.0	53.9
ADA	AADA† [40] (1%)	37.2	33.1	57.9	47.7
	AADA† (5%)	38.5	41.0	58.4	53.7
	Ours (1%)	52.2	45.1	61.8	64.2
	Ours (5%)	51.8	**50.9**	**65.6**	**67.9**
Oracle		46.3	52.3	68.9	68.9

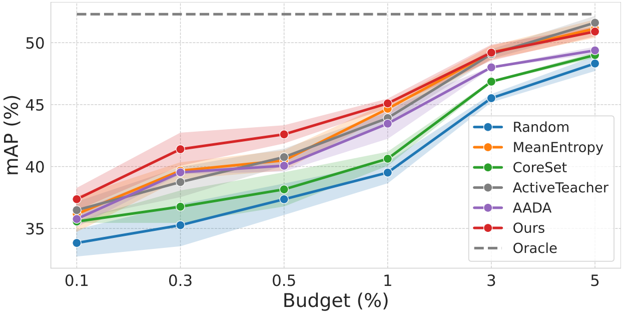

Figure 5. **Comparison of active sampling strategies in the adaptation from Cityscapes to BDD100k.** The horizontal axis represents the labeling budget, and the vertical axis represents mAP. Our proposed strategy outperformed the conventional strategies across almost all budgets, achieving a performance close to that of Oracle with 5% of the labeling budget.

4.3. Comparison with Active Sampling Strategies

We evaluated our proposed method with different active sampling strategies and budgets. Figure 5 shows the performance of the active sampling strategy on budgets in the **C** → **B** scenario, comparing our method with random sampling, MeanEntropy [43], CoreSet [37], ActiveTeacher [26], and AADA [40] as the baseline for ADA.

Our proposed strategy consistently outperformed previous strategies across almost all budgets. In particular, there was a substantial performance gain against AADA when the budget was less than or equal to 1%, witnessing 2.5 pt improvements at 0.5% of the labeling budget. The reason for

Table 2. **Comparison of the performances of our proposed metric under varying active sampling strategies.** The most accurate number is represented in bold. In MeanEntropy [43] and AADA [40], we underline the results with higher accuracy for the with and without s^{fn}.

Strategies	w/ s^{fn}	Budget (%)					
		0.1	0.3	0.5	1	3	5
Random		33.8	35.3	37.4	39.5	45.5	48.3
MeanEntropy		36.2	37.1	40.5	44.7	**49.2**	**51.1**
	✓	**38.4**	39.3	41.7	43.9	48.4	51.0
AADA		35.8	39.5	40.1	43.5	48.0	49.4
	✓	37.4	38.3	40.7	44.0	48.4	50.4
Ours	✓	37.4	**41.4**	**42.6**	**45.1**	**49.2**	50.9

these improvements is that FN errors increased when using the previous strategies with small budgets, as the model was not sufficiently adapted to the target domain. Therefore, the samples that reduce the FN errors are not selected using the previous strategies. In contrast, our undetectability metric effectively selects the samples that reduce FN errors. Thus, our proposed active sampling strategy is particularly effective under low-budget conditions.

We also evaluated the effectiveness of our undetectability metric s^{fn}. Table 2 shows the results of a performance comparison when s^{fn} was incorporated into previous strategies. In MeanEntropy, our proposed metric improved accuracy when the budget was low. This result indicates that a high budget is needed to sample targets with low-class probabilities, which leads to the occurrence of FN errors. In AADA, our proposed metric was generally effective. This result indicates that the diversity metric used in AADA is less effective in suppressing FN errors.

4.4. Ablation Studies

4.4.1 Contribution of Components

We investigated the contribution of components within our framework. Our framework comprises components such as adversarial learning (AD), pseudo-labeling (PL), mean teacher (MT), uncertainty-aware PL (UP), and strong augmentation (SA). We compared the performances delivered by the model trained with a 0.5% budget in the **C** → **F** and **K** → **C** scenarios (Table 3).

SA was the most beneficial component for enhancing performance in both scenarios within our framework. AD and PL followed in terms of impactful contributions in the **C** → **F** scenario, highlighting the efficacy of feature-level alignment across domains. MT and UP did not show any significant differences. This result indicates that Cityscapes and Foggy Cityscapes share the same information on object regions; hence, a simple pseudo-label with a confidence threshold is sufficient for prediction.

Conversely, PL, MT, and UP, in that order, contributed to the **K** → **C** scenario, indicating that the prediction accuracy of pseudo-label is crucial. Implementation of UP in our framework led to an improvement of the performance by 3 pt owing to the effectiveness of uncertainty in improving the prediction accuracy of the pseudo-label. In contrast, the contribution of AD was low. Pseudo-labels are more effective than feature-level alignment in this scenario owing to the large instance-level gaps.

Table 3. Comparison of the performances of our proposed method under varying usage conditions of the components

AD	PL	MT	UP	SA	**C** → **F**	**K** → **C**
	✓	✓	✓	✓	38.2	63.2
✓				✓	40.1	53.9
✓	✓		✓	✓	51.1	54.2
✓	✓	✓		✓	50.5	60.9
✓	✓	✓	✓		30.2	49.9
✓	✓	✓	✓	✓	51.6	63.1

Table 4. Comparison of the performances of our proposed method under varying changing each indicator of the acquisition function

s^{fn}	s^{loc}	s^{ent}	s^{div}	**S** → **C**	**K** → **C**
	✓	✓	✓	57.4	61.8
✓		✓	✓	59.2	62.4
✓	✓		✓	60.6	61.2
✓	✓	✓		59.5	62.8
✓	✓	✓	✓	58.4	63.1

4.4.2 Effectiveness of False Negative Prediction

We investigated the contribution of each metric in the active sampling strategy (Table 4). For this experiment, we used a model trained with a 0.5% labeling budget in the **S** → **C** and **K** → **C** scenarios. In the **S** → **C** scenario, s^{fn} provides the largest contribution, as evidenced by the remarkably lower performance when s^{fn} is omitted, and the notable performance gap observed between scenarios with and without s^{fn}. In the **K** → **C** scenario, we observed a result similar to that in the **S** → **C** scenario. These results show that this pattern is effective in various scenarios.

5. Conclusion

We proposed an active domain adaptation method designed for object detection. Through a performance analysis under domain shift, we identified the challenge of undetected objects and proposed the False Negative Prediction Module, focusing on addressing this issue. Experimental verification in in-vehicle camera scenarios, characterized by significant domain gaps, demonstrated the effectiveness of our method. Thus, we believe that our method performs well across various datasets.

References

[1] Sharat Agarwal, Saket Anand, and Chetan Arora. Reducing annotation effort by identifying and labeling contextually diverse classes for semantic segmentation under domain shift. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5904–5913, 2023. 1, 2

[2] Shengcao Cao, Dhiraj Joshi, Liang-Yan Gui, and Yu-Xiong Wang. Contrastive mean teacher for domain adaptive object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23839–23848, 2023. 1, 2, 7

[3] Meilin Chen, Weijie Chen, Shicai Yang, Jie Song, Xinchao Wang, Lei Zhang, Yunfeng Yan, Donglian Qi, Yueting Zhuang, Di Xie, and Shiliang Pu. Learning domain adaptive object detection with probabilistic teacher. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 3040–3055, 2022. 2, 4, 7

[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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3339–3348, 2018. 1, 2, 7

[5] Jiwoong Choi, Ismail Elezi, Hyuk-Jae Lee, Clement Farabet, and Jose M. Alvarez. Active learning for deep object detection via probabilistic modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10264–10273, 2021. 1, 2

[6] 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 (CVPR)*, pages 3213–3223, 2016. 6

[7] Antoine de mathelin, François Deheeger, Mathilde Mougeot, and Nicolas Vayatis. Discrepancy-based active learning for domain adaptation. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022. 1, 2

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009. 7

[9] Jinhong Deng, Wen Li, Yuhua Chen, and Lixin Duan. Unbiased mean teacher for cross-domain object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4091–4101, 2021. 1, 2

[10] Ismail Elezi, Zhiding Yu, Anima Anandkumar, Laura Leal-Taixé, and Jose M. Alvarez. Not all labels are equal: Rationalizing the labeling costs for training object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14492–14501, 2022. 1, 2

[11] Bo Fu, Zhangjie Cao, Jianmin Wang, and Mingsheng Long. Transferable query selection for active domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7272–7281, 2021. 1, 2

[12] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pages 1050–1059, 2016. 5

[13] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 1180–1189, 2015. 2, 4

[14] Changlong Gao, Chengxu Liu, Yujie Dun, and Xueming Qian. Csda: Learning category-scale joint feature for domain adaptive object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11421–11430, 2023. 1, 2, 7

[15] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361, 2012. 6

[16] Denis Gudovskiy, Alec Hodgkinson, Takuya Yamaguchi, and Sotaro Tsukizawa. Deep active learning for biased datasets via fisher kernel self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9041–9049, 2020. 1

[17] Han-Kai Hsu, Chun-Han Yao, Yi-Hsuan Tsai, Wei-Chih Hung, Hung-Yu Tseng, Maneesh Singh, and Ming-Hsuan Yang. Progressive domain adaptation for object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 749–757, 2020. 7

[18] Duojun Huang, Jichang Li, Weikai Chen, Junshi Huang, Zhenhua Chai, and Guanbin Li. Divide and adapt: Active domain adaptation via customized learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7651–7660, 2023. 1, 2

[19] Sehyun Hwang, Sohyun Lee, Sungyeon Kim, Jungseul Ok, and Suha Kwak. Combating label distribution shift for active domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 549–566, 2022. 1, 2

[20] 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? In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 746–753, 2017. 6

[21] 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 (WACV)*, pages 382–391, 2023. 1, 2

[22] Suraj Kothawade, Saikat Ghosh, Sumit Shekhar, Yu Xiang, and Rishabh Iyer. Talisman: Targeted active learning for object detection with rare classes and slices using submodular

mutual information. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 1–16, 2022. 1, 2

[23] Yu-Jhe Li, Xiaoliang Dai, Chih-Yao Ma, Yen-Cheng Liu, Kan Chen, Bichen Wu, Zijian He, Kris Kitani, and Peter Vajda. Cross-domain adaptive teacher for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7581–7590, 2022. 1, 2, 7

[24] Mengyao Lyu, Jundong Zhou, Hui Chen, Yijie Huang, Dongdong Yu, Yaqian Li, Yandong Guo, Yuchen Guo, Liuyu Xiang, and Guiguang Ding. Box-level active detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23766–23775, 2023. 1, 2

[25] Xinhong Ma, Junyu Gao, and Changsheng Xu. Active universal domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8968–8977, 2021. 1, 2

[26] Peng Mi, Jianghang Lin, Yiyi Zhou, Yunhang Shen, Gen Luo, Xiaoshuai Sun, Liujuan Cao, Rongrong Fu, Qiang Xu, and Rongrong Ji. Active teacher for semi-supervised object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14482–14491, 2022. 1, 2, 7

[27] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. In *Proceedings of the Advances in Neural Information Processing Systems Workshops (NeurIPSW)*, 2013. 5

[28] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529–533, 2015. 5

[29] Munan Ning, Donghuan Lu, Dong Wei, Cheng Bian, Chenglang Yuan, Shuang Yu, Kai Ma, and Yefeng Zheng. Multi-anchor active domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9112–9122, 2021. 1, 2

[30] Younghyun Park, Wonjeong Choi, Soyeong Kim, Dong-Jun Han, and Jaekyun Moon. Active learning for object detection with evidential deep learning and hierarchical uncertainty aggregation. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2023. 1, 2

[31] 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 (ICCV)*, pages 8505–8514, 2021. 1, 2

[32] Quazi Marufur Rahman, Niko Sünderhauf, and Feras Dayoub. Did you miss the sign? a false negative alarm system for traffic sign detectors. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3748–3753, 2019. 5

[33] Harsh Rangwani, Arihant Jain, Sumukh K Aithal, and R. Venkatesh Babu. S3vaada: Submodular subset selection for virtual adversarial active domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7516–7525, 2021. 1, 2

[34] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2015. 7

[35] 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 (CVPR)*, pages 6956–6965, 2019. 1, 2, 7

[36] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision (IJCV)*, 126:973–992, 2018. 6

[37] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018. 7

[38] Inkyu Shin, Dong-Jin Kim, Jae Won Cho, Sanghyun Woo, Kwanyong Park, and In So Kweon. Labor: Labeling only if required for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8588–8598, 2021. 1, 2

[39] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015. 7

[40] Jong-Chyi Su, Yi-Hsuan Tsai, Kihyuk Sohn, Buyu Liu, Subhransu Maji, and Manmohan Chandraker. Active adversarial domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 739–748, 2020. 1, 2, 3, 6, 7, 8

[41] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2017. 6

[42] Huy V. Vo, Oriane Siméoni, Spyros Gidaris, Andrei Bursuc, Patrick Pérez, and Jean Ponce. Active learning strategies for weakly-supervised object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 211–230, 2022. 1, 2

[43] Dan Wang and Yi Shang. A new active labeling method for deep learning. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pages 112–119, 2014. 6, 7, 8

[44] Fan Wang, Zhongyi Han, Zhiyan Zhang, Rundong He, and Yilong Yin. Mhpl: Minimum happy points learning for active source free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20008–20018, 2023. 1, 2

[45] Yuting Wang, Velibor Ilic, Jiatong Li, Branislav Kisačanin, and Vladimir Pavlovic. Alwod: Active learning for weakly-supervised object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6459–6469, 2023. 1, 2

[46] Jiaxi Wu, Jiaxin Chen, and Di Huang. Entropy-based active learning for object detection with progressive diversity constraint. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9397–9406, 2022. 1, 2

[47] Tsung-Han Wu, Yi-Syuan Liou, Shao-Ji Yuan, Hsin-Ying Lee, Tung-I Chen, Kuan-Chih Huang, and Winston H. Hsu. D2ada: Dynamic density-aware active domain adaptation for semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 449–467, 2022. 1, 2

[48] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, and Xinjing Cheng. Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8068–8078, 2022.

[49] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8708–8716, 2022.

[50] Ming Xie, Yuxi Li, Yabiao Wang, Zekun Luo, Zhenye Gan, Zhongyi Sun, Mingmin Chi, Chengjie Wang, and Pei Wang. Learning distinctive margin toward active domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7993–8002, 2022.

[51] Mixue Xie, Shuang Li, Rui Zhang, and Chi Harold Liu. Dirichlet-based uncertainty calibration for active domain adaptation. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2023. 1, 2

[52] Qinghua Yang, Hui Chen, Zhe Chen, and Junzhe Su. Introspective false negative prediction for black-box object detectors in autonomous driving. *Sensors*, 21(8):2819, 2021. 5

[53] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 93–102, 2019. 1, 2

[54] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2636–2645, 2020. 6

[55] Jinze Yu, Jiaming Liu, Xiaobao Wei, Haoyi Zhou, Yohei Nakata, Denis Gudovskiy, Tomoyuki Okuno, Jianxin Li, Kurt Keutzer, and Shanghang Zhang. Mttrans: Cross-domain object detection with mean teacher transformer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 629–645, 2022. 1, 2

[56] Weiping Yu, Sijie Zhu, Taojiannan Yang, and Chen Chen. Consistency-based active learning for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3951–3960, 2022. 1, 2

[57] Jiakang Yuan, Bo Zhang, Xiangchao Yan, Tao Chen, Botian Shi, Yikang Li, and Yu Qiao. Bi3d: Bi-domain active learning for cross-domain 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15599–15608, 2023. 1, 2

[58] Tianning Yuan, Fang Wan, Mengying Fu, Jianzhuang Liu, Songcen Xu, Xiangyang Ji, and Qixiang Ye. Multiple instance active learning for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5330–5339, 2021. 1, 2

[59] Hao Zhang and Ruimao Zhang. Active domain adaptation with multi-level contrastive units for semantic segmentation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 1640–1657, 2022. 1, 2

[60] Wenzhang Zhou, Dawei Du, Libo Zhang, Tiejian Luo, and Yanjun Wu. Multi-granularity alignment domain adaptation for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9581–9590, 2022. 1, 2, 7

[61] Wenzhang Zhou, Heng Fan, Tiejian Luo, and Libo Zhang. Unsupervised domain adaptive detection with network stability analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6986–6995, 2023. 1, 2, 7