# Bridge Past and Future: Overcoming Information Asymmetry in Incremental Object Detection

Qijie Mo[1,3], Yipeng Gao[1,3], Shenghao Fu[1,3], Junkai Yan[1,3], Ancong Wu[1,3,⋆], and Wei-Shi Zheng[1,2,3,⋆]

[1] School of Computer Science and Engineering, Sun Yat-sen University, China
[2] Peng Cheng Laboratory, Shenzhen, China
[3] Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
{moqj3,gaoyp23,fushh7,yanjk3}@mail2.sysu.edu.cn, wuanc@mail.sysu.edu.cn, wszheng@ieee.org

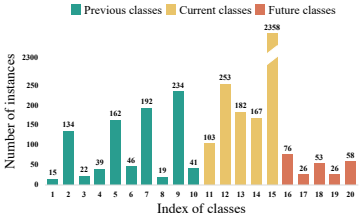
**Abstract.** In incremental object detection, knowledge distillation has been proven to be an effective way to alleviate catastrophic forgetting. However, previous works focused on preserving the knowledge of old models, ignoring that images could simultaneously contain categories from past, present, and future stages. The co-occurrence of objects makes the optimization objectives inconsistent across different stages since the definition for foreground objects differs across various stages, which limits the model's performance greatly. To overcome this problem, we propose a method called "Bridge Past and Future" (BPF), which aligns models across stages, ensuring consistent optimization directions. In addition, we propose a novel Distillation with Future (DwF) loss, fully leveraging the background probability to mitigate the forgetting of old classes while ensuring a high level of adaptability in learning new classes. Extensive experiments are conducted on both Pascal VOC and MS COCO benchmarks. Without memory, BPF outperforms current state-of-the-art methods under various settings. The code is available at https://github.com/iSEE-Laboratory/BPF.

**Keywords:** Object Detection · Incremental Learning · Knowledge Distillation

## 1 Introduction

Object detection [3, 14, 36, 53, 61, 68] is a fundamental computer vision task and has significantly given rise to many sub-directions [2, 15, 16, 19, 27, 44, 47, 63]. General detection models typically assume full access to all interest classes during training and require exhaustive labeled data from the start. However, in dynamic real-world applications, where new object categories may appear continually, object detection models are expected to adapt to these ongoing changing classes, especially when there is limited storage to save all the data from the beginning

⋆ Corresponding Author




(a) Distribution of instances in the second stage of the 10-5-5 setting on VOC 2007 dataset

(b) Visualization of some typical training samples in IOD

**Fig. 1:** Challenges faced by the IOD task. Classes from previous, current, and potential future stages may appear arbitrarily in the current training stage, while only the annotations of current classes are available to train the detector. Best viewed in color.

to the present or when data privacy concerns are encountered. To this end, incremental object detection (IOD) [5, 26, 41, 65, 67, 69] has caught progressive attention, which aims to continually detect new objects without forgetting the previously learned ones.

Unlike incremental learning in classification, the incremental object detection task faces a dual challenge caused by the concurrence of different classes: (1) Foreground objects identified in previous stages are **unlabeled** in the current stage, and thus, the model currently trained may regard them as background. (2) The current stage's background may also contain objects that will be recognized as the foreground in future stages. Taking the 10-5-5 setting on the VOC dataset [12] as an example, as shown in Figure 1, a large proportion of instances belonging to classes of the previous and future stages occur concurrently in the training images in the current stage. This asymmetry in information between the past and current stages causes the current model to erroneously classify objects of old categories as background, aggravating the catastrophic forgetting problem. Similarly, the asymmetry between the current and future stages leads the current model to classify objects of potential future categories as background erroneously, which requires models trained in future stages to rectify their misperception, thus hindering the learning of new classes. As a result, the optimization objectives across different learning stages are inconsistent, significantly limiting the generalization of current incremental detectors. Previous works [5, 65, 67] concentrate on introducing strong regularization to prevent catastrophic forgetting but ignore the impact of concurrence of different classes, resulting in limited performance on both old and new classes.

This work aims to tackle the critical information asymmetry challenge in IOD by utilizing the abundant concurrence information within an image. We introduce a novel approach named Bridge Past and Future (BPF), which is designed to connect the past and future stages with the current stage, ensuring the model adheres to consistent optimization objectives throughout the entire incremental learning process. For the past stages, we utilize the previous model as a reliable labeler to construct some pseudo labels for old classes and combine them with current annotations for training the current model symmetrically.

Simultaneously, we re-identify some salient regions that may contain potential future objects out of the background and exclude them from negative samples in the current stage to avoid classifying them into the background at this stage and disturbing future learning, which is expected to achieve symmetry with future stages. These two novel designs alleviate the impact of inconsistent training objectives across different stages from the aspects of the past and the future, thus easing the training difficulty of incremental detectors.

In addition, we propose a distillation technique to augment the detector in the current stage, where we take a step back by viewing the current stage as a clear future stage for previous stages and propose a Distillation with Future (DwF) loss. The distillation is carried out by two teachers, *i.e.*, the detector trained on the previous stage and an expert detector trained with only current data. The student detector absorbs the knowledge of old and current classes from the old and expert models, respectively, in a class-by-class manner, preventing the detector from catastrophic forgetting and facilitating learning of current classes.

We conduct extensive experiments on the PASCAL VOC [12] and COCO [37] datasets to evaluate the effectiveness of our BPF, which outperforms other state-of-the-art (SOTA) methods under multiple settings in a memory-free way. Moreover, we also conduct comprehensive ablation studies and visualizations to help better understand how each component of the BPF works.

## 2 Related Works

**Object Detection.** Traditional object detectors can be primarily classified into two streams: one-stage [36,52,58,61] and two-stage detectors [18,22,35,53]. Two-stage detectors first predict several coarse candidate proposals via region proposal extractors [18,53] and then adopt a region of interest (RoI) head to refine these proposals and output final predictions. Unlike two-stage detectors, one-stage counterparts directly generate final outputs without predicting candidate proposals. Despite performing well under the standard training setup, both of them fail to generalize to the incremental training setup due to lacking previous training data. Without losing generality, this paper concentrates on enhancing two-stage detectors, *i.e.*, Faster R-CNN [53], enabling it to learn new classes incrementally while retaining previously acquired knowledge.

**Incremental Learning.** Incremental learning methods can be mainly divided into three categories: memory-based [4,24,32,33,48,51,56,59,62], regularization-based [6,8,30,34,66], and structure-based [42,45,46,54,60]. Memory-based methods store a handful of samples for replaying [24, 40, 41] or suggest generating samples [29,62] for data compensation in new stages. Some regularization-based methods try to eliminate the effect of new tasks on previously learned knowledge by identifying the most crucial parameters [30, 43, 62], while some other regularization-based methods propose to distill knowledge [23] from old to current models by transiting knowledge on logits, final features or intermediate features [8, 10, 11, 24, 34]. Structure-based methods dedicate specific parameters to each task, freezing them to mitigate forgetting. In this study, we focus on the

regularization-based knowledge distillation approach for object detection and introduce a novel distillation with future information.

**Incremental Object Detection.** Most previous works [9,28,31,40,50,55] focus on distilling old classes' knowledge of intermediate features [5,21,39,41,49,65,67], region proposal network [5, 49, 67], or RoI head [13] to prevent forgetting. A stream of methods also pays attention to classification loss. PPAS [67] introduces a pseudo-positive-aware sampling algorithm to identify regions corresponding to old classes and prevents them from being sampled as background. MMA [5] proposes unbiased classification loss, consolidating the background and all old classes into one entity, which aims to minimize the optimization objective conflict between the current stage's background and the past stages' foreground. However, it diminishes the ability to distinguish old classes from the background. Other methods [1, 20, 25, 26, 41, 64] focus on rehearsal to maintain the previous stage knowledge, either performing replay of the intermediate features [1], the images [25, 26, 40], or the instances [41]. Incdet [38] presents a parameter isolation strategy that builds upon EWC [30]. In this study, in addition to preserving knowledge for old classes, we also take future classes into account, aligning optimization objectives consistently across all stages.

## 3 Method

### 3.1 Preliminaries

**Problem Formulation.** In incremental object detection, the training is performed over multiple learning stages, each one introducing a new set of classes to be detected. Let $\mathcal{C}$ denote the set of classes that are incrementally introduced to the object detector $\mathcal{M}$. In the $t$-th training stage, a grouping of classes $\mathcal{C}_t$ are introduced to the detector: $\mathcal{C}_t \subset \mathcal{C}$, such that $\mathcal{C}_i \cap \mathcal{C}_j = \varnothing$, for any $i \neq j$ and $i, j \leq t$. Let $\mathcal{D}_t$ denote the images containing annotated objects of classes in $\mathcal{C}_t$. Each image can contain multiple objects of different classes, but annotations $\mathcal{Y}_t$ are available only for those object instances that belong to classes in $\mathcal{C}_t$. The challenge in class incremental object detection is to continually update $\mathcal{M}_t$ to $\mathcal{M}_{t+1}$ by learning $\mathcal{C}_{t+1}$, without access to $\{\mathcal{D}_0, \cdots, \mathcal{D}_t\}$ while maintaining original performance on $\{\mathcal{C}_0, \cdots, \mathcal{C}_t\}$.

**General Detector Training Pipeline.** This work starts with the representative Faster RCNN-like detectors [53] $f = \{f_b, f_{rpn}, f_{roi}\}$, which generally includes a backbone network $f_b$, Region Proposal Network (RPN) $f_{rpn}$ and Region of Interest Head (RoI Head) $f_{roi}$. During training, all training proposals are categorized into positive, negative, and ignored samples for supervised training based on their IoUs with the ground truth. Positive samples are assigned to predict the class of their matched ground truth, whereas negative samples are designated as background. Thus, the objects without annotations in the current stage will be classified as background, hindering the learning of incremental models.

**Common Distillation Methods for Incremental Object Detection.** To prevent forgetting in IOD, a widely adopted solution involves knowledge distil-

lation in RPN and RoI head [5, 49, 57, 67]:

$$\mathcal{L}_{dist} = \mathcal{L}_{dist}^{rpn} + \mathcal{L}_{dist}^{roi}. \tag{1}$$

The distillation in the RoI head includes the L2 loss between the box coordinates and the Kullback-Leibler divergence for class probabilities. MMA [5] notices the missing old annotations and proposes Unbiased Knowledge Distillation (UKD):

$$\mathcal{L}_{dist,cls}^{roi}(i) = \frac{1}{|\mathcal{C}_{t-1}|+1}(p_i^{b,t-1}\log(p_i^{b,t} + \sum_{c\in\mathcal{C}_t} p_i^{c,t}) + \sum_{c\in\mathcal{C}_{1:t-1}} p_i^{c,t-1}\log(p_i^{c,t})), \tag{2}$$

where $p_i^{c,t-1}$ and $p_i^{c,t}$ indicate the classification output for the proposal $i$ and class $c$ of the teacher (old) model and student (current) model, respectively, and $b$ is the background class. As a common practice, 64 proposals are used for distillation, which are randomly selected out of the top 128 proposals with the highest objectness scores from the RPN network of the old model. However, given the class-agnostic characteristic of the RPN, the proposals for distillation cover objects of both old and current classes. In such cases, the old model, without seeing current classes, cannot provide useful knowledge for the current classes and even hinders their learning. Further, treating the background probability and current class probabilities as a unified entity during distillation inevitably diminishes the ability to differentiate between current classes and background.

### 3.2 Overall Framework

Unlike classification tasks with a single label per input, incremental object detection presents a scenario where an image $\mathcal{I}_t$ can encompass objects from the current class set $\mathcal{C}_t$, previous class sets $\mathcal{C}_{1:t-1}$ and future class sets $\mathcal{C}_{t+1:\infty}$. Nonetheless, the annotations $\mathcal{Y}_t$ are limited to bounding boxes and class labels for objects in $\mathcal{C}_t$, while objects of other classes are regarded as background.

To address the challenge of inconsistent optimization objectives, we introduce a novel strategy termed Bridge Past and Future (BPF). As shown in Figure 2, the method is divided into two parts. From the perspective of supervised learning, we use high-confidence predictions of old classes given by the old model as pseudo labels to bridge the information from the past stage (Section 3.3) and discard some potential objects from the background to bridge the future stage (Section 3.4). Regarding the distillation learning, we propose a novel Distillation with Future (DwF) loss by considering current classes in the distillation. The distillation probabilities are combined from two teacher models: the old model $\mathcal{M}_{t-1}$ for old classes and an expert model for current classes. By utilizing abundant information in background probability, the distillation takes both previous stages and the current stage into consideration (Section 3.5).

### 3.3 Bridging the Past

To address the inconsistency in optimization objectives between the previous and current models, we bridge the past information to the current stage by incorporating pseudo supervision signals from past stages into the current model's

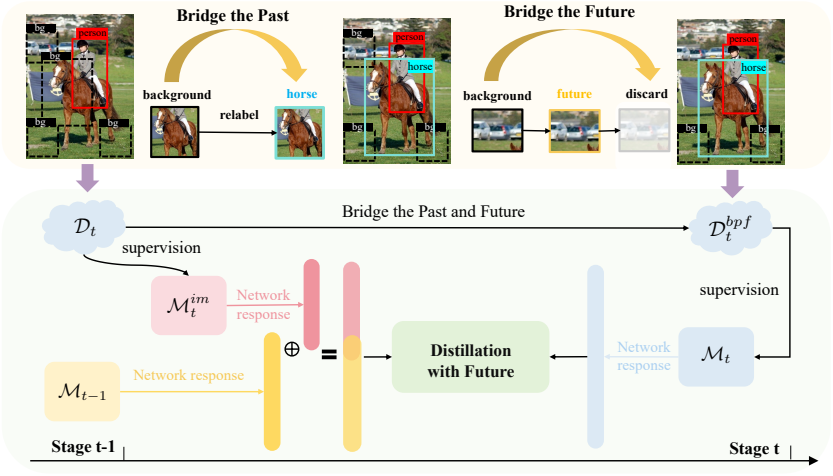



**Fig. 2:** The overall framework of our method. The top side illustrates the Bridge Past and Future (BPF) procedure, which identifies objects of past classes and excludes several potential objects of future classes to ensure consistent optimization during the entire training process. The bottom side shows the Distillation with Future (DwF) process, which employs both the old model $\mathcal{M}_{t-1}$ adept at detecting old categories and the interim model $\mathcal{M}_t^{im}$ trained on $\mathcal{D}_t$ and specialized in new categories, to conduct a comprehensive distillation across all categories for the current model $\mathcal{M}_t$.

optimization process. The old model was trained under the supervision of human annotations, which is a high-quality pseudo labeler for old classes. We take the high-confidence inference results of the old model $\mathcal{M}_{t-1}$ on current training images as pseudo labels and combine them with current annotations to train the current detector, thereby aligning the supervision of old classes with previous stages. The left part of Figure 3 shows the pipeline of bridging the past strategy.

Formally, given predictions $\hat{y}^{old} \in \mathcal{P}$ of the old model $\mathcal{M}_{t-1}$ on current images, we first select some high-confidence regions of previous classes $\mathcal{C}_{1:t-1}$ with a confidence threshold $\eta$, followed by the Non-Maximum Suppression (NMS) operation to reduce duplication:

$$\mathcal{U} = \{j \in \mathcal{P} : \max_{c \in \mathcal{C}_{1:t-1}} \hat{p}_j^{\text{old}}(c) > \eta\}, \quad \mathcal{E} = \text{NMS}(\mathcal{U}). \tag{3}$$

Then, we further narrow down the predictions to a subset $\mathcal{W} \subset \mathcal{E}$ that do not overlap with the ground-truth labels of the new categories via an IoU threshold $\lambda_1$, ensuring a clear distinction between past and present classes:

$$\mathcal{W} = \{j \in \mathcal{E} : \forall i \in \mathcal{Y}_t, \text{IoU}(\hat{\boldsymbol{b}}_j^{\text{old}}, \boldsymbol{b}_i) \leq \lambda_1\}. \tag{4}$$

By modeling the previous supervision signals from the old model, we bridge the past stages to the current one, ensuring the current model's optimization direction encompasses the objectives of earlier stages, significantly mitigating

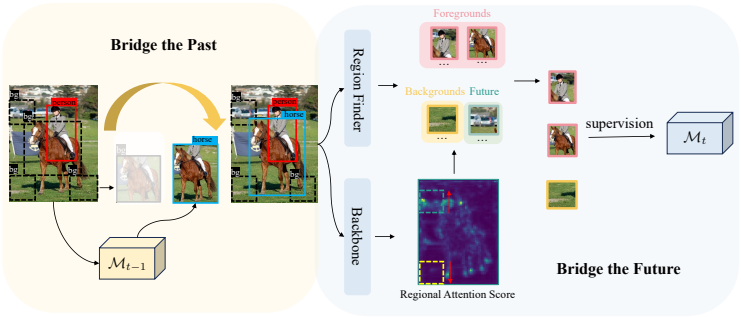



**Fig. 3:** Overview of Bridge Past and Future. We adopt the previous model $\mathcal{M}_{t-1}$ to predict some pseudo labels for past classes to complement their missing supervision in the current stage. Additionally, we exclude several proposals that are likely to be an object but are not included in the current ground truth and pseudo labels from the background to avoid classifying them into background mistakenly.

the forgetting problem. MMA [5] treats the background probability and old class probabilities as a unified entity in the classification loss, making old classes hard to separate from the background. While our method explicitly models the old classes in the current stage out of the background.

### 3.4 Bridging the Future

Bridging the future aims to rectify the misalignment between the optimization goals of the current and upcoming stages, as objects from future class sets $\mathcal{C}_{t+1:\infty}$ in the current dataset $\mathcal{D}_t$ are classified as background. The core design is to find some salient objects in the background and exclude them from the negative samples during the training. By separating pure background from regions likely to contain future category objects in the current stage, we align the optimization objectives for background treatment across both current and future models.

Specifically, we find that the activation in feature maps is a good indicator for salient objects. As illustrated in Figure 3, in the absence of annotations for old and future classes, foreground and background features demonstrate notable differences in spatial attention. We produce the attention map $A_i \in \mathbb{R}^{H\times W}$ from the backbone feature map $F_i = f_b(I_i) \in \mathbb{R}^{H\times W\times C}$:

$$A_i = \text{Softmax}\left(\sum_{c=1}^{C} |F_i|^p\right), \tag{5}$$

where $H$, $W$, and $C$ denote the feature's height, width, and channel. We compute the attention score for each region, indicating the likelihood of containing foreground objects. The attention score $a_{i,j}^{roi}$ for region $r_j$ is calculated as follows:

$$a_{i,j}^{roi} = \text{Avg}(\text{RoIPool}(A_i, r_j)) \in \mathbb{R}. \tag{6}$$

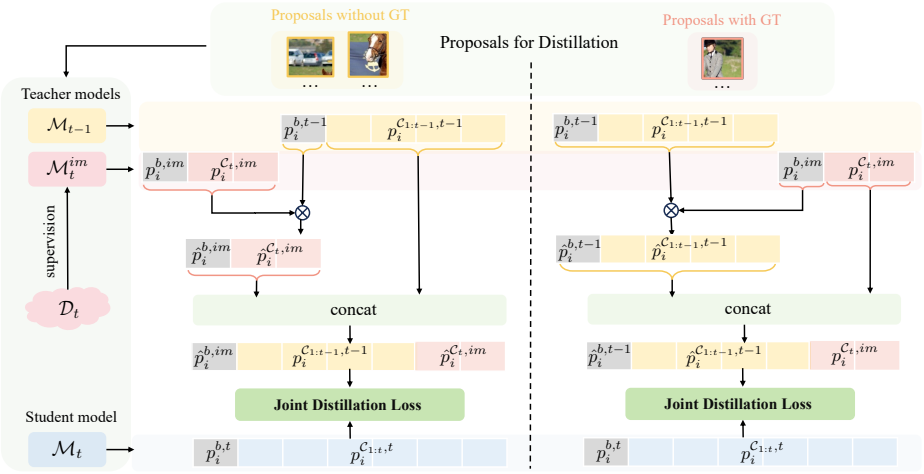



**Fig. 4:** The illustration of Distillation with Future strategy. An intermediate teacher model trained on the current dataset is used to compensate for the lack of current class information in the old model. For proposals overlapping ground truth of the current stage, since the intermediate teacher is specialized in detecting current classes, we directly inherit its probabilities on current classes and use the old model to enrich its background probability with old class knowledge. On the contrary, for proposals that do not overlap with GT, the old model is preferred, and the intermediate model is used as compensation. Combining two teachers makes the distillation class by class.

Regions with high attention scores $a_{i,j}^{roi}$ from feature maps and objectness scores $o_j$ from class-agnostic RPN imply a greater chance of being future category objects. We discard these RoIs when sampling negative samples when training the RoI head, thus maintaining the model's consistency with future stage background definitions. Regions with lower scores or having a considerable IoU with ground truth are considered reliable backgrounds.

### 3.5 Combining Current Classes into Distillation

Distillation is an effective way to prevent forgetting in incremental object detection. However, distilling from the teacher model, which is biased toward old classes, inevitably hinders the learning for current classes. To address this challenge, we introduce the Distillation with Future (DwF) loss to distill from the teacher model in a more fine-grained and adaptive way. Different from MMA [5], which aligns the current classes and the background as a whole with the old model's background, we distill each class one by one, preserving the distinction between current classes and the background. However, the model $\mathcal{M}_{t-1}$ from the previous stage has not been trained on current classes, thus the distillation on current classes can not be performed. We introduce another intermediate teacher model $\mathcal{M}_t^{im}$, which is trained using the dataset from the current stage $\mathcal{D}_t$ in a fully supervised way, as a supplement. Taking the intermediate model $\mathcal{M}_t^{im}$ as a

complementary teacher, we explicitly consider the current stage $t$ as the future stage of stage $t-1$, making the distillation future-aware.

As the old model $\mathcal{M}_{t-1}$ performs well on $\mathcal{C}_{1:t-1}$ while the intermediate model $\mathcal{M}_t^{im}$ is the expert in $\mathcal{C}_t$, we distill different regions with different combination of teacher models, as shown in Figure 4. Specifically, we divide the regions for distillation $\mathcal{R}$ into two subsets $\mathcal{R}_1, \mathcal{R}_2 \subset \mathcal{R}$ that based on their intersection over union with the ground truth labels for the new categories $\mathcal{C}_t$:

$$\begin{aligned} \mathcal{R}_1 &= \{j \in \mathcal{R} : \forall i \in \mathcal{Y}_t, \mathrm{IoU}(b_j, b_i) \leq \lambda_2.\}, \\ \mathcal{R}_2 &= \{j \in \mathcal{R} : \forall i \in \mathcal{Y}_t, \mathrm{IoU}(b_j, b_i) > \lambda_2.\}. \end{aligned} \tag{7}$$

For regions $r_i \in \mathcal{R}_1$, which are likely to be the regions for old classes, we take $\mathcal{M}_{t-1}$ as the primary model for distillation and reconstruct its background representation with the model $\mathcal{M}_t^{im}$:

$$\hat{p}_i^{c,im} = p_i^{c,im} \times p_i^{b,t-1}, \quad r_i \in \mathcal{R}_1, \tag{8}$$

where $p_i^{b,t-1}$ and $p_i^{c,im}$ are the classification probabilities of background in the model $\mathcal{M}_{t-1}$ and the classification probabilities for current classes and background $c \in \mathcal{C}_t \cup \mathcal{B}$ in the model $\mathcal{M}_t^{im}$ for the region $r_i$ respectively. After weighting, $\sum_{c=1}^{\mathcal{C}_t \cup \mathcal{B}} \hat{p}_i^{c,im} = p_i^{b,t-1}$. The final distillation probabilities for regions in $\mathcal{R}_1$ are $[p_i^{\mathcal{C}_{1:t-1},t-1}, \hat{p}_i^{\mathcal{C}_t \cup \mathcal{B},im}] \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$.

On the contrary, for regions $r_i \in \mathcal{R}_2$, they are regions for current classes; thus, the intermediate model $\mathcal{M}_t^{im}$ is taken as the primary model for distillation, and its background representation is reconstructed by the model $\mathcal{M}_{t-1}$:

$$\hat{p}_i^{c,t-1} = p_i^{c,t-1} \times p_i^{b,im}, \quad r_i \in \mathcal{R}_2, \tag{9}$$

where $\sum_{c=1}^{\mathcal{C}_{1:t-1} \cup \mathcal{B}} \hat{p}_i^{c,t-1} = p_i^{b,im}$. The distillation probabilities for regions in $\mathcal{R}_2$ are $[\hat{p}_i^{\mathcal{C}_{1:t-1},t-1}, p_i^{\mathcal{C}_t,im}, \hat{p}_i^{\mathcal{B},t-1}] \in \mathbb{R}^{|\mathcal{C}_{1:t}|+1}$.

With the expanded probabilities from the complimentary teacher models, the distillation on the classification head can be performed using a conventional Kullback-Leibler divergence. Regarding the box distillation, we use the output boxes from the old model $\mathcal{M}_{t-1}$ for regions in $\mathcal{R}_1$ and the intermediate model $\mathcal{M}_t^{im}$ for $\mathcal{R}_2$. As a result, the complementary knowledge from two teachers, the expansion of background probability, and the combination of adaptive probability for different regions not only prevent the student from catastrophic forgetting but also facilitate the learning for current classes.

## 4 Experiments

### 4.1 Experiment Settings

**Datasets and Evaluation Metrics.** Following previous works [5,20,25,26,41, 49,57,65,67], we evaluate our method on PASCAL VOC 2007 [12] and MS COCO

**Table 1:** mAP@0.5 results on single incremental step on PASCAL VOC 2007. The best performance in each is presented with **bold**, and the second best is presented with underlined. Methods with $*$ use exemplars.

| Method | 19-1 | | | | 15-5 | | | | 10-10 | | | | 5-15 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1-19 | 20 | 1-20 | Avg | 1-15 | 16-20 | 1-20 | Avg | 1-10 | 11-20 | 1-20 | Avg | 1-5 | 5-15 | 1-20 | Avg |
| Joint Training | 76.0 | 76.7 | 76.1 | 76.4 | 78.0 | 70.4 | 76.1 | 74.2 | 75.9 | 76.3 | 76.1 | 76.1 | 72.4 | 77.3 | 76.1 | 74.9 |
| Fine-tuning | 12.0 | 62.8 | 14.5 | 37.4 | 14.2 | 59.2 | 25.4 | 36.7 | 9.5 | 62.5 | 36.0 | 36.0 | 6.9 | 63.1 | 49.1 | 35.0 |
| ORE* [25] | 69.4 | 60.1 | 68.9 | 64.7 | 71.8 | 58.7 | 68.5 | 65.2 | 60.4 | 68.8 | 64.6 | 64.6 | - | - | - | - |
| OW-DETR* [20] | 70.2 | 62.0 | 69.8 | 66.1 | 72.2 | 59.8 | 69.1 | 66.0 | 63.5 | 67.9 | 65.7 | 65.7 | - | - | - | - |
| ILOD-Meta* [26] | 70.9 | 57.6 | 70.2 | 64.2 | 71.7 | 55.9 | 67.8 | 63.8 | 68.4 | 64.3 | 66.3 | 66.3 | - | - | - | - |
| ABR* [41] | 71.0 | **69.7** | 70.9 | **70.4** | 73.0 | **65.1** | 71.0 | 69.1 | 71.2 | 72.8 | 72.0 | 72.0 | 64.7 | 71.0 | 69.4 | 67.9 |
| Faster ILOD [53] | 68.9 | 61.1 | 68.5 | 65.0 | 71.6 | 56.9 | 67.9 | 64.3 | 69.8 | 54.5 | 62.1 | 62.1 | 62.0 | 37.1 | 43.3 | 49.6 |
| PPAS [67] | 70.5 | 53.0 | 69.2 | 61.8 | - | - | - | - | 63.5 | 60.0 | 61.8 | 61.8 | - | - | - | - |
| MVC [65] | 70.2 | 60.6 | 69.7 | 65.4 | 69.4 | 57.9 | 66.5 | 63.7 | 66.2 | 66.0 | 66.1 | 66.1 | - | - | - | - |
| PROB [69] | 73.9 | 48.5 | 72.6 | 61.5 | 73.5 | 60.8 | 70.1 | 67.0 | 66.0 | 67.2 | 66.5 | 66.5 | - | - | - | - |
| PseudoRM [64] | 72.9 | 67.3 | 72.6 | 70.1 | 73.4 | 60.9 | 70.3 | 66.9 | 69.1 | 68.6 | 68.9 | 68.9 | - | - | - | - |
| MMA [5] | 71.1 | 63.4 | 70.7 | 67.2 | 73.0 | 60.5 | 69.9 | 66.7 | 69.3 | 63.9 | 66.6 | 66.6 | **66.8** | 57.2 | 59.6 | 62.0 |
| **BPF (Ours)** | **74.5** | 65.3 | **74.1** | 69.9 | **75.9** | 63.0 | **72.7** | **69.5** | **71.7** | **74.0** | **72.9** | **72.9** | 66.4 | **75.3** | **73.0** | **70.9** |

2017 [37] datasets. PASCAL VOC 2007 dataset comprises 9,963 images across 20 categories. The COCO 2017 dataset encompasses objects from 80 categories, with around 118k images for training and 5,000 images for validation. The mean average precision at the 0.5 IoU threshold (mAP@0.5) is used as the primary evaluation metric for the VOC dataset, and the mean average precision ranging from 0.5 to 0.95 is the main evaluation metric for the COCO dataset.

For each incremental setting (A-B), the first number A denotes the number of classes in the first stage and the second number is the number of classes newly introduced in each new stage. Note that the columns with gray background in the table of experimental results represent the average AP among all classes.

**Implementation Details.** Similar to [5,25,26,41,49], we build our incremental object detector based on Faster R-CNN [53] with R50. Our method can easily be adapted to transformer-based detectors [3,17,68]. We conduct the experiments under a strict **rehearsal-free** setting, where no memory is used. We set $\eta = 0.75$, $\lambda_1 = 0.7$, $\lambda_2 = 0.5$. For $\mathcal{W}$ in Equation (4), we use an IOU threshold of 0.3 to divide it into two sets. For the set with IoU $< 0.3$, the supervision signal weights are set to 1.0, while the other set is assigned a weight of 0.3.

### 4.2 Quantitative Evaluation

Following previous work [5,7,41,49,57,65,67], our method is evaluated on settings with a range of initial classes and incorporating one or more incremental tasks. We benchmark our method against two baselines: Fine-Tuning, where the model is incrementally trained on new data without any regularization strategy or data replay, and Joint Training, which involves training the model on the complete dataset using all annotations.

**Table 2:** mAP@0.5 results on multiple incremental steps on PASCAL VOC 2007. The best performance in each is presented with **bold**, and the second best is presented with underlined. Methods with $*$ use exemplars.

| Method | 10-5 (3 tasks) | | | 5-5 (4 tasks) | | | 10-2 (6 tasks) | | | 15-1 (6 tasks) | | | 10-1 (10 tasks) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1-10 | 11-20 | 1-20 | 1-5 | 6-20 | 1-20 | 1-10 | 11-20 | 1-20 | 1-15 | 16-20 | 1-20 | 1-10 | 11-20 | 1-20 |
| Joint Training | 75.9 | 76.3 | 76.1 | 72.4 | 77.3 | 76.1 | 75.9 | 76.3 | 76.1 | 78.0 | 70.4 | 76.1 | 75.9 | 76.3 | 76.1 |
| Fine-tuning [41] | 5.3 | 30.6 | 18.0 | 0.5 | 18.3 | 13.8 | 3.8 | 13.6 | 8.7 | 0.0 | 10.5 | 5.3 | 0.0 | 5.1 | 2.6 |
| ABR* [41] | 68.7 | 67.1 | 67.9 | **64.7** | 56.4 | 58.4 | 67.0 | **58.1** | **62.6** | 68.7 | **56.7** | 65.7 | 62.0 | **55.7** | **58.9** |
| Faster ILOD [53] | 68.3 | 57.9 | 63.1 | 55.7 | 16.0 | 25.9 | 64.2 | 48.6 | 56.4 | 66.9 | 44.5 | 61.3 | 52.9 | 41.5 | 47.2 |
| MMA [5] | 66.7 | 61.8 | 64.2 | 62.3 | 31.2 | 38.9 | 65.0 | 53.1 | 59.1 | 68.3 | 54.3 | 64.1 | 59.2 | 48.3 | 53.8 |
| **BPF (Ours)** | **69.1** | **68.2** | **68.7** | 60.6 | **63.1** | **62.5** | **68.7** | 56.3 | 62.5 | **71.5** | 53.1 | **66.9** | **62.2** | 48.3 | 55.2 |

**Table 3:** mAP results on COCO2017. Methods with $*$ use exemplars.

| Method | 40-40 | | | 70-10 | | |
|---|---|---|---|---|---|---|
| | $AP$ | $AP_{50}$ | $AP_{75}$ | $AP$ | $AP_{50}$ | $AP_{75}$ |
| Joint Training | 36.7 | 57.8 | 39.8 | 36.7 | 57.8 | 39.8 |
| Fine-tuning [41] | 19.0 | 31.2 | 20.4 | 5.6 | 8.6 | 6.2 |
| ILOD-Meta* [26] | 23.8 | 40.5 | 24.4 | - | - | - |
| ABR* [41] | **34.5** | **57.8** | 35.2 | 31.1 | 52.9 | 32.7 |
| Faster ILOD [49] | 20.6 | 40.1 | - | 21.3 | 39.9 | - |
| PseudoRM [64] | 25.3 | 44.4 | - | - | - | - |
| MMA [5] | 33.0 | 56.6 | 34.6 | 30.2 | 52.1 | 31.5 |
| **BPF (Ours)** | 34.4 | 54.3 | **37.3** | **36.2** | **56.8** | **38.9** |

**Table 4:** Ablation study of various combinations of teacher models.

| Distillation | | VOC(10-10) | | |
|---|---|---|---|---|
| $\mathcal{L}^{roi}_{dist,cls}$ | $\mathcal{L}^{roi}_{dist,bbox}$ | 1-10 | 11-20 | 1-20 |
| $\lambda_2 = 1.0$ part boxes | | 71.5 | 73.3 | 72.4 |
| $\lambda_2 = 0.5$ part boxes | | **71.7** | 74.0 | **72.9** |
| $\lambda_2 = 0.5$ all boxes | | 71.3 | **74.4** | **72.9** |

**Table 5:** Effect of BF.

| BF | VOC(5-15) | | | VOC(10-10) | | | VOC(15-5) | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1-5 | 6-20 | 1-20 | 1-10 | 11-20 | 1-20 | 1-15 | 16-20 | 1-20 |
| × | 66.3 | 74.4 | 72.4 | 71.2 | 73.3 | 72.3 | 75.6 | 62.8 | 72.4 |
| ✓ | **66.4** | **75.3** | **73.0** | **71.7** | **74.0** | **72.9** | **75.9** | **63.0** | **72.7** |

**PASCAL VOC 2007.** For PASCAL VOC 2007, we order the classes alphabetically and evaluate our method with one or multiple training steps. We perform our experiments by adding 1 (19-1), 5 (15-5), 10 (10-10), or 15 (5-15) classes in a single incremental step. For multi-step incremental settings, we evaluate 10-5, 5-5, 10-2, 15-1, and 10-1 settings, where we add 5, 5, 2, 1, and 1 classes respectively at every step until all 20 classes are seen.

**- Single-step Incremental Settings:** Table 1 shows our BPF methods against the existing methods using rehearsal or not. Rehearsal-based methods are not compared fairly with our BPF since we do not store old samples and use replay memory. As shown in Table 1, BPF consistently outperforms all previous methods, including those designed to combat forgetting using exemplars, validating the superiority of our approach. In particular, in the 19-1, 15-5, 10-10, and 5-15 settings, BPF significantly improved over MMA [5] by 3.4%, 2.8%, 6.3%, and 13.4% on mAP@0.5 across all classes. Similarly, BPF outperforms the best rehearsal-based method ABR [41] by 3.2%, 1.7%, 0.9%, and 3.6%. The *Avg* metric equally averages old and new classes mAP, which straightly reports the incremental ability without the influence of the number of classes. BPF also outperforms most methods on the *Avg* metric, demonstrating BPF's adaptiveness in learning new classes and preserving the knowledge of old classes.

**- Multi-step Incremental Settings:** The issues of inconsistent optimization objectives across multiple stages and catastrophic forgetting are more crucial under the longer incremental settings. As shown in Table 2, BPF consistently outperforms MMA [5] across all the settings. Specifically, BPF improves over MMA by 1.4%, 2.8%, 3.4%, 4.5%, and even 23.6% at **1-20** mAP@0.5 under

**Table 6:** Ablation study on each component.

| Model | Bridge the Past | Bridge the future | Distillation with Future | VOC(10-10) | | | VOC(10-5) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1-10 | 11-20 | 1-20 | 1-10 | 11-15 | 16-20 | 1-20 |
| (a) | | | | 58.1 | 72.4 | 65.3 | 54.4 | 69.9 | 59.3 | 59.5 |
| (b) | ✓ | | | 71.2 | 72.1 | 71.7 | 69.1 | 73.5 | 60.0 | 67.9 |
| (c) | | ✓ | | 61.0 | 73.3 | 67.1 | 54.8 | 70.7 | 58.7 | 59.8 |
| (d) | ✓ | ✓ | | **71.9** | 72.7 | 72.3 | **70.4** | 73.7 | 58.7 | 68.3 |
| (e) | ✓ | ✓ | ✓ | 71.7 | **74.0** | **72.9** | 69.1 | **75.2** | **61.2** | **68.7** |

the 10-1, 15-1, 10-2, 10-5, and 5-5 settings and enjoy improvement across all learning stages. Moreover, we find that even without storing memory, BPF still outperforms ABR [41] by 0.8% and 4.1% on overall mAP@0.5, 1.1% and 6.7% mAP@0.5 on new classes under 10-5 and 5-5 settings. In the settings of 10-2 and 10-1, limited by the small incremental data, BPF is inferior to ABR, but considering we do not require memory, these losses are acceptable.

**MS COCO 2017.** On the COCO2017 dataset, we perform experiments on 40-40 and 70-10 settings, adding 40 and 10 classes, respectively, following [41]. As illustrated in Table 3, our method improves over MMA on average $AP$ by 1.4% on 40-40 settings and by 6.0% on 70-10 settings. These results once again confirm the effectiveness of our method.

### 4.3 Analysis and Ablation Study

We examine the contributions of the "Bridge the Past", "Bridge the Future", and "Distillation with Future" in Table 6 within the VOC 10-10 and 10-5 settings. We take the unbiased knowledge distillation proposed by [5] as the baseline model. By bridging the past, our model (b) aligns its optimization objectives with earlier ones, effectively reducing catastrophic forgetting of old classes, greatly enhancing the performance on detecting old classes. Compared to the baseline (a), it significantly improved by 13.1% in the old classes and by 6.4% for all classes on the 10-10 setting. Owing to bridging the future, our model (c) maintains consistent optimization objectives with future models regarding the background, making it easier to incrementally learn new classes. Compared to the baseline (a) on the 10-10 setting, it improves by 2.9% and 0.9% on the old classes and the new classes, respectively. Combining Past and Future (d), the model outperforms consistently on each stage. In the Distillation with Future strategy, we leverage the intermediate model to aid the old model in modeling new classes, conducting distillation across all categories for the current model. Table 6 (e) shows that a comprehensive joint teacher model facilitates improved learning of all categories through knowledge distillation, with a significant improvement (+1.3% AP) in new classes. Similar results can be found in the 10-5 setting.

We further verify the "Bridge the future" (BF) in Table 5.There is a clear trend that as the number of considered future classes increases, the BF shows increasing improvement in future classes (+0.9 % in 5-15) without degrading the old classes' performance.

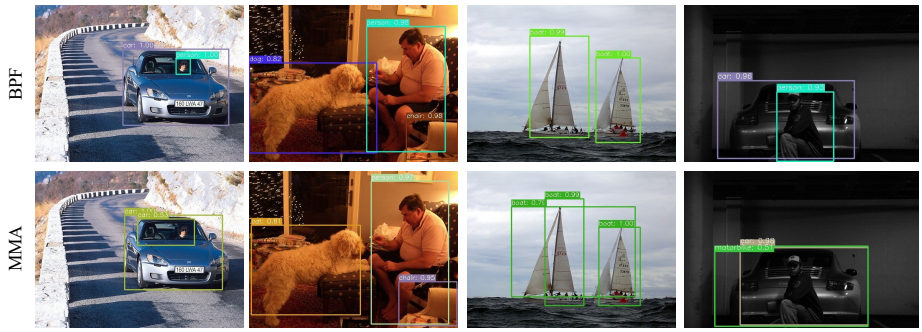



**Fig. 5:** Qualitative results for the model trained under the 10+10 setting on the VOC 2007 test set. 'boat', 'cat', 'chair', and 'car' are old classes from the first stage, and 'person' and 'dog' are the classes from the second stage. Compared with MMA (bottom row), our BPF (top row) can produce reliable predictions on both old and new classes.

We also conducted a quantitative analysis of Bridge Past and Future. Under the second stage of the VOC 10-5-5 (3 tasks), the original background boxes have a 92.6% and 83.9% Recall50 rate for old and future classes, while the Recall50 of pseudo labels for old classes and discarded boxes for future classes are 66.1% and 17.6% respectively, demonstrating the necessity and effectiveness of Bridge the Past and Bridge the Future.

Table 4 presents experiments on the effect of different combinations of teacher models. $\lambda_2$ in Equation (7) determines using which teacher model as the primary. Using the intermediate expert model as the primary on regions overlapping with gt ($\lambda_2 = 0.5$) outperforms using the old model ($\lambda_2 = 1.0$), showing that distilling new class objects with the old model hinders their learning. For box distillation, we find that distilling boxes on primary classes (part boxes, *i.e.*, only boxes for old classes in $\mathcal{R}_1$ and boxes for new classes in $\mathcal{R}_2$ participate in distillation while others are ignored) performs similarly with all classes (all boxes).

### 4.4 Visualization

**Visualization of Detection Results.** We visualize the detection results in Figure 5 to illustrate the significant improvement compared to the previous methods qualitatively. Our method accurately detects both new and old class objects simultaneously. While MMA fails to detect old classes accurately, suffering from catastrophic forgetting.

**Visualization of Bridge Past and Future.** We visualized the module of Bridge the Past and Bridge the Future separately in Figure 6. We effectively model the missing annotations for old class objects to bridge the past. As demonstrated in Figure 6(b), the attention maps clearly differentiate the foreground from the background, regardless of the presence of annotations. The discarded boxes (the third row) validate the effectiveness of our method.

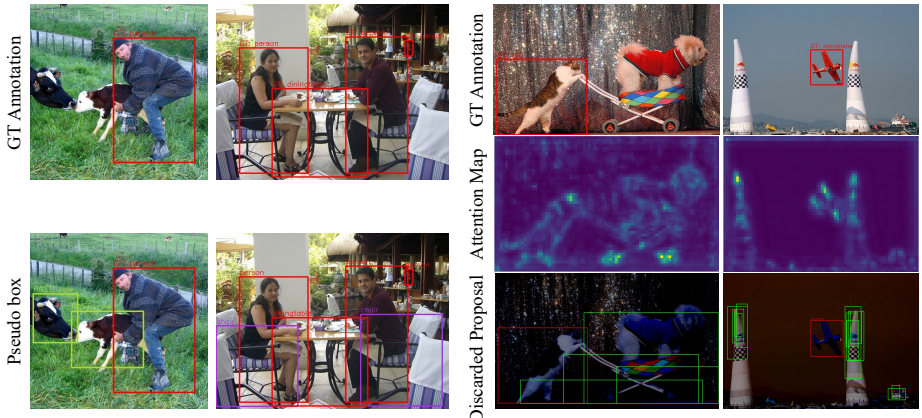



**Fig. 6:** Visualization of Bridge Past and Future. Boxes in red represent the ground truth in the current stage. **(a)** In Bridge the Past, we effectively constructed pseudo labels of past classes. **(b)** In Bridge the Future, salient objects (marked in green boxes) can be easily detected from the attention maps and are excluded from the background regions. Best viewed in color.

## 5 Conclusions

**Limitations.** In our Bridge the Past procedure, we assume several objects of old classes may appear in the current training data. However, when the number of incremental classes is limited, *e.g.*, increasing a single class in each stage, objects of old classes may rarely occur due to limited training images. This is expected to be alleviated by generating samples of old classes via the copy-paste strategy as in ABR [41], while it may introduce little stored samples. Detailed discussion can be found in the Supplement.

**Conclusions.** In this work, we find that the concurrence of classes from different learning stages causes a severe information asymmetry, not only causing catastrophic forgetting for old classes but also hindering the learning of new classes. To tackle the problem, we propose the Bridge Past and Future method, which uses pseudo labels from the old model to fill in the missing annotations and exclude some potential future objects from the background, keeping the learning consistent across all stages. Further, Distillation with Future loss is proposed to solve the problem of the old teacher model's lack of knowledge of new classes. By combining the knowledge from the past to the future, our method consistently outperforms others across different learning stages in most incremental settings. To the best of our knowledge, we are the first to consider future classes during the learning, shedding light on a new aspect of Incremental Object Detection.

**Acknowledgments.** This work was supported partially by the National Key Research and Development Program of China (2023YFA1008503), NSFC(U21A20471), Guangdong NSF Project (No. 2023B1515040025, 2020B1515120085).

## References


1. Acharya, M., Hayes, T.L., Kanan, C.: Rodeo: Replay for online object detection. arXiv preprint arXiv:2008.06439 (2020)
2. Cao, S., Joshi, D., Gui, L.Y., Wang, Y.X.: Contrastive mean teacher for domain adaptive object detectors. In: CVPR (2023)
3. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: ECCV (2020)
4. Castro, F.M., Marín-Jiménez, M.J., Guil, N., Schmid, C., Alahari, K.: End-to-end incremental learning. In: ECCV (2018)
5. Cermelli, F., Geraci, A., Fontanel, D., Caputo, B.: Modeling missing annotations for incremental learning in object detection. In: CVPR (2022)
6. Chaudhry, A., Dokania, P.K., Ajanthan, T., Torr, P.H.: Riemannian walk for incremental learning: Understanding forgetting and intransigence. In: ECCV (2018)
7. Chen, L., Yu, C., Chen, L.: A new knowledge distillation for incremental object detection. In: IJCNN (2019)
8. Dhar, P., Singh, R.V., Peng, K.C., Wu, Z., Chellappa, R.: Learning without memorizing. In: CVPR (2019)
9. Dong, N., Zhang, Y., Ding, M., Bai, Y.: Class-incremental object detection. PR (2023)
10. Douillard, A., Cord, M., Ollion, C., Robert, T., Valle, E.: Podnet: Pooled outputs distillation for small-tasks incremental learning. In: ECCV (2020)
11. Douillard, A., Ramé, A., Couairon, G., Cord, M.: Dytox: Transformers for continual learning with dynamic token expansion. In: CVPR (2022)
12. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. Springer IJCV (2010)
13. Feng, T., Wang, M., Yuan, H.: Overcoming catastrophic forgetting in incremental object detection via elastic response distillation. In: CVPR (2022)
14. Fu, S., Yan, J., Gao, Y., Xie, X., Zheng, W.S.: Asag: Building strong one-decoder-layer sparse detectors via adaptive sparse anchor generation. In: ICCV (2023)
15. Gao, Y., Lin, K.Y., Yan, J., Wang, Y., Zheng, W.S.: Asyfod: An asymmetric adaptation paradigm for few-shot domain adaptive object detection. In: CVPR (2023)
16. Gao, Y., Yang, L., Huang, Y., Xie, S., Li, S., Zheng, W.S.: Acrofod: An adaptive method for cross-domain few-shot object detection. In: ECCV (2022)
17. Gao, Z., Wang, L., Han, B., Guo, S.: Adamixer: A fast-converging query-based object detector. In: CVPR (2022)
18. Girshick, R.: Fast r-cnn. In: ICCV (2015)
19. Gu, X., Lin, T.Y., Kuo, W., Cui, Y.: Open-vocabulary object detection via vision and language knowledge distillation. arXiv preprint arXiv:2104.13921 (2021)
20. Gupta, A., Narayan, S., Joseph, K., Khan, S., Khan, F.S., Shah, M.: Ow-detr: Open-world detection transformer. In: CVPR (2022)
21. Hao, Y., Fu, Y., Jiang, Y.G., Tian, Q.: An end-to-end architecture for class-incremental object detection with knowledge distillation. In: ICME (2019)
22. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: ICCV (2017)
23. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
24. Hou, S., Pan, X., Loy, C.C., Wang, Z., Lin, D.: Learning a unified classifier incrementally via rebalancing. In: CVPR (2019)
25. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: CVPR (2021)

26. Joseph, K., Rajasegaran, J., Khan, S., Khan, F.S., Balasubramanian, V.N.: Incremental object detection via meta-learning. IEEE TPAMI (2021)
27. Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., Darrell, T.: Few-shot object detection via feature reweighting. In: ICCV (2019)
28. Kang, M., Zhang, J., Zhang, J., Wang, X., Chen, Y., Ma, Z., Huang, X.: Alleviating catastrophic forgetting of incremental object detection via within-class and between-class knowledge distillation. In: CVPR (2023)
29. Kemker, R., Kanan, C.: Fearnet: Brain-inspired model for incremental learning. arXiv preprint arXiv:1711.10563 (2017)
30. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences (2017)
31. Li, D., Tasci, S., Ghosh, S., Zhu, J., Zhang, J., Heck, L.: Rilod: Near real-time incremental learning for object detection at the edge. In: Proceedings of the 4th ACM/IEEE Symposium on Edge Computing (2019)
32. Li, M., Cong, Y., Liu, Y., Sun, G.: Class-incremental gesture recognition learning with out-of-distribution detection. In: IROS (2022)
33. Li, Y.M., Zeng, L.A., Meng, J.K., Zheng, W.S.: Continual action assessment via task-consistent score-discriminative feature distribution modeling. TCSVT (2024)
34. Li, Z., Hoiem, D.: Learning without forgetting. IEEE TPAMI (2017)
35. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: CVPR (2017)
36. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017)
37. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: ECCV (2014)
38. Liu, L., Kuang, Z., Chen, Y., Xue, J.H., Yang, W., Zhang, W.: Incdet: In defense of elastic weight consolidation for incremental object detection. IEEE TNNLS (2020)
39. Liu, X., Yang, H., Ravichandran, A., Bhotika, R., Soatto, S.: Multi-task incremental learning for object detection. arXiv preprint arXiv:2002.05347 (2020)
40. Liu, Y., Schiele, B., Vedaldi, A., Rupprecht, C.: Continual detection transformer for incremental object detection. In: CVPR (2023)
41. Liu, Y., Cong, Y., Goswami, D., Liu, X., van de Weijer, J.: Augmented box replay: Overcoming foreground shift for incremental object detection. In: ICCV (2023)
42. Liu, Y., Cong, Y., Sun, G., Zhang, T., Dong, J., Liu, H.: L3doc: Lifelong 3d object classification. IEEE TIP (2021)
43. Lopez-Paz, D., Ranzato, M.: Gradient episodic memory for continual learning. In: NeurIPS (2017)
44. Ma, J., Niu, Y., Xu, J., Huang, S., Han, G., Chang, S.F.: Digeo: Discriminative geometry-aware learning for generalized few-shot object detection. In: CVPR (2023)
45. Mallya, A., Davis, D., Lazebnik, S.: Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In: ECCV (2018)
46. Mallya, A., Lazebnik, S.: Packnet: Adding multiple tasks to a single network by iterative pruning. In: CVPR (2018)
47. Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., et al.: Simple open-vocabulary object detection. In: ECCV (2022)

48. Ostapenko, O., Puscas, M., Klein, T., Jahnichen, P., Nabi, M.: Learning to remember: A synaptic plasticity driven framework for continual learning. In: CVPR (2019)
49. Peng, C., Zhao, K., Lovell, B.C.: Faster ilod: Incremental learning for object detectors based on faster rcnn. PR (2020)
50. Peng, C., Zhao, K., Maksoud, S., Wang, T., Lovell, B.C.: Diode: dilatable incremental object detection. PR (2023)
51. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: icarl: Incremental classifier and representation learning. In: CVPR (2017)
52. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
53. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: NeurIPS (2015)
54. Rusu, A.A., Rabinowitz, N.C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., Hadsell, R.: Progressive neural networks. arXiv preprint arXiv:1606.04671 (2016)
55. Shieh, J.L., Haq, Q.M.u., Haq, M.A., Karam, S., Chondro, P., Gao, D.Q., Ruan, S.J.: Continual learning strategy in one-stage object detection framework based on experience replay for autonomous driving vehicle. Sensors (2020)
56. Shin, H., Lee, J.K., Kim, J., Kim, J.: Continual learning with deep generative replay. In: NeurIPS (2017)
57. Shmelkov, K., Schmid, C., Alahari, K.: Incremental learning of object detectors without catastrophic forgetting. In: ICCV (2017)
58. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: CVPR (2020)
59. Tang, Y.M., Peng, Y.X., Zheng, W.S.: Learning to imagine: Diversify memory for incremental learning using unlabeled data. In: CVPR (2022)
60. Tang, Y.M., Peng, Y.X., Zheng, W.S.: When prompt-based incremental learning does not meet strong pretraining. In: ICCV (2023)
61. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: ICCV (2019)
62. Wu, C., Herranz, L., Liu, X., Van De Weijer, J., Raducanu, B., et al.: Memory replay gans: Learning to generate new categories without forgetting. In: NeurIPS (2018)
63. Yan, J., Yang, L., Gao, Y., Zheng, W.S.: Self-supervised cross-stage regional contrastive learning for object detection. In: ICME (2023)
64. Yang, D., Zhou, Y., Hong, X., Zhang, A., Wei, X., Zeng, L., Qiao, Z., Wang, W.: Pseudo object replay and mining for incremental object detection. In: ACM MM (2023)
65. Yang, D., Zhou, Y., Zhang, A., Sun, X., Wu, D., Wang, W., Ye, Q.: Multi-view correlation distillation for incremental object detection. PR (2022)
66. Zenke, F., Poole, B., Ganguli, S.: Continual learning through synaptic intelligence. In: ICML (2017)
67. Zhou, W., Chang, S., Sosa, N., Hamann, H., Cox, D.: Lifelong object detection. arXiv preprint arXiv:2009.01129 (2020)
68. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. In: ICLR (2021)
69. Zohar, O., Wang, K.C., Yeung, S.: Prob: Probabilistic objectness for open world object detection. In: CVPR (2023)