



# Applied Artificial Intelligence

## An International Journal

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/uaai20>

## DR-CIML: Few-shot Object Detection via Base Data Resampling and Cross-iteration Metric Learning

Guoping Cao, Wei Zhou, Xudong Yang, Feijia Zhu & Lin Chai

To cite this article: Guoping Cao, Wei Zhou, Xudong Yang, Feijia Zhu & Lin Chai (2023) DR-CIML: Few-shot Object Detection via Base Data Resampling and Cross-iteration Metric Learning, Applied Artificial Intelligence, 37:1, 2175116, DOI: [10.1080/08839514.2023.2175116](https://doi.org/10.1080/08839514.2023.2175116)

To link to this article: <https://doi.org/10.1080/08839514.2023.2175116>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 09 Feb 2023.



Submit your article to this journal [↗](#)



Article views: 506



View related articles [↗](#)



View Crossmark data [↗](#)

# DR-CIML: Few-shot Object Detection via Base Data Resampling and Cross-iteration Metric Learning

Guoping Cao<sup>a</sup>, Wei Zhou<sup>b</sup>, Xudong Yang<sup>c</sup>, Feijia Zhu<sup>c</sup>, and Lin Chai<sup>a</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of Complex Systems of Engineering and School of Automation, Southeast University, Nanjing, P. R. China; <sup>b</sup>The School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China; <sup>c</sup>Shenzhen R&D Center, Walvision Intelligent Technology Co. Ltd, Nanjing, China

## ABSTRACT



Aiming at the problems of difficult data collection and labor-intensive manual annotation, few-shot object detection (FSOD) has gained wide attention. Although current transfer-learning-based detection methods outperform meta-learning-based methods, their data organization fails to fully utilize the diversity of the source domain data. In view of this, Data Resampling (DR) organization is proposed to fine-tune the network, which can be employed as a component of any model and dataset without additional inference overhead. In addition, in order to improve the generalization of the model, a Cross-Iteration Metric-Learning (CIML) branch is embedded in the few-shot object detector, thus actively improving intra-category feature propinquity and inter-category feature discrimination. Our generic DR-CIML approach obtained competitive scores in extensive comparative experiments. The nAP50 performance on PASCAL VOC improved by up to 6.3 points, and the bAP50 performance reached 81.0, surpassing the base stage model (80.8) for the first time. The nAP75 performance on MS COCO improved by up to 1.6 points.

## ARTICLE HISTORY

Received 20 October 2022  
Revised 4 January 2023  
Accepted 27 January 2023

## Introduction

With their forceful feature learning and visual perception capability, deep convolutional neural networks (CNNs) show performance beyond human level, achieving great success in computer vision fields such as image classification, object detection, and segmentation (Bochkovskiy, Wang, and Liao 2020; He et al. 2016; Ren et al. 2015; Simonyan and Zisserman 2014; Tan and Le 2019; Tian et al. 2019). However, general object detection algorithms require a large amount of labeled data to generalize well and obtain an effective model, which imposes a heavy workload and is costly (Deng et al. 2009; Everingham et al. 2010, 2015; Lin et al. 2014). In cases such as medical applications (Katzmann et al. 2021) or rare species (Mannocci et al. 2022), it is unrealistic to obtain a large amount of data, while ordinary people quickly

**CONTACT** Lin Chai  [chailin1@seu.edu.cn](mailto:chailin1@seu.edu.cn)  Key Laboratory of Measurement and Control of Complex Systems of Engineering and School of Automation, Southeast University, Nanjing 210096, P. R. China

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

learn new concepts from only a few observations, even at early ages (Samuelson and Smith 2005). In view of this, it is of great significance to study how to obtain a vision system with good generalization performance using a small number of samples.

Scholars have researched deep learning-based few-shot object detection (Lu et al. 2020), which is complicated because it must accurately locate different categories of objects (Köhler, Eisenbach, and Gross 2021). If only a small number of novel categories are trained, the detector is prone to nonconvergence or overfitting, resulting in poor generalization (Chen et al. 2018), and it cannot correctly detect instances from novel categories.

Some meta-learning-based methods (Finn, Abbeel, and Levine 2017; Nichol, Achiam, and Schulman 2018; Vinyals et al. 2016; Yan et al. 2019) effectively learn prior knowledge from multiple subtasks and even learn to learn, so as to learn new tasks with few training examples. The performance of transfer-learning-based FSOD methods (Sun et al. 2021; Wang et al. 2020) exceeds that of meta-learning-based FSOD methods. Among them, TFA (Wang et al. 2020) has a simple and effective two-stage, single-branch structure. By freezing all model parameters except the last layer, the problem of losing source domain knowledge in transfer-learning (Zhuang et al. 2020) was solved. In addition, TFA has established a new evaluation protocol and new benchmarks by repeating runs to obtain a stable evaluation. FSCE (Sun et al. 2021) adopts the same data division as TFA, where the major cause affecting the AP on novel class (nAP) is the misclassification of novel categories rather than inaccurate positioning. Therefore, a CPE branch is embedded in the RoI feature extractor to improve classification performance, inspired by the successful application of comparative learning in image recognition (Schroff, Kalenichenko, and Philbin 2015; Sun et al. 2014) and self-supervised representation learning (Khosla et al. 2020). However, CPE is slow to increase the differences of feature embeddings of inter-class objects, and a large amount of data is needed for self-supervised learning. Therefore, we apply triple loss, which is considered more appropriate (Schroff, Kalenichenko, and Philbin 2015), to actively minimize the distance between the anchor and the positive examples of the same category and to maximize the distance between the anchor and negative examples of different categories. Cross-iteration metric learning is added to increase the feature diversity of metric learning, so as to better solve the problem of misclassification in FSOD.

In addition, it is obvious that current methods based on transfer learning train on the fixed base class and novel class samples, i.e., the training data of each epoch are exactly the same. Although this settles the problem of data imbalance, the diversity of the abundant base data is not fully utilized. As a result, the model still has room to improve its detection performance. In view of the above problems, the main contributions of this paper include the following:

- Base data resampling. To fully utilize the diversity of the abundant base data, instead of fixing the base categories samples in the fine-tuning phase, we randomly sample  $K$ -shot base category instances for each epoch when maintaining the same partition and fixed novel categories instances as TFA (Wang et al. 2020). We believe this is the first application of data resampling organization to transfer-learning-based FSOD;
- Cross-iteration metric-learning branch. To deal with the misclassification of novel category instances, we employ a cross-iteration metric-learning branch with triplet loss (Weinberger and Saul 2009) in FSOD for supervised learning. We retain object feature embeddings from adjacent iterations to increase the feature diversity of metric learning, so as to actively promote intra-category feature propinquity and inter-category feature discrimination.

Figure 1 shows the modifications and improvements of transfer-learning-based FSOD via our proposed DR-CIML.

In extensive experiments, our generic training scheme obtained the highest novel-categories AP50 (nAP50) almost in three different splits under  $K$ -shot settings with  $K = 1, 2, 3, 5,$  and  $10$  on PASCAL VOC (Everingham et al. 2010, 2015), and the nAP50 performance improved by up to 6.3 points. Furthermore, the proposed method is the first to achieve  $>80$  base-categories

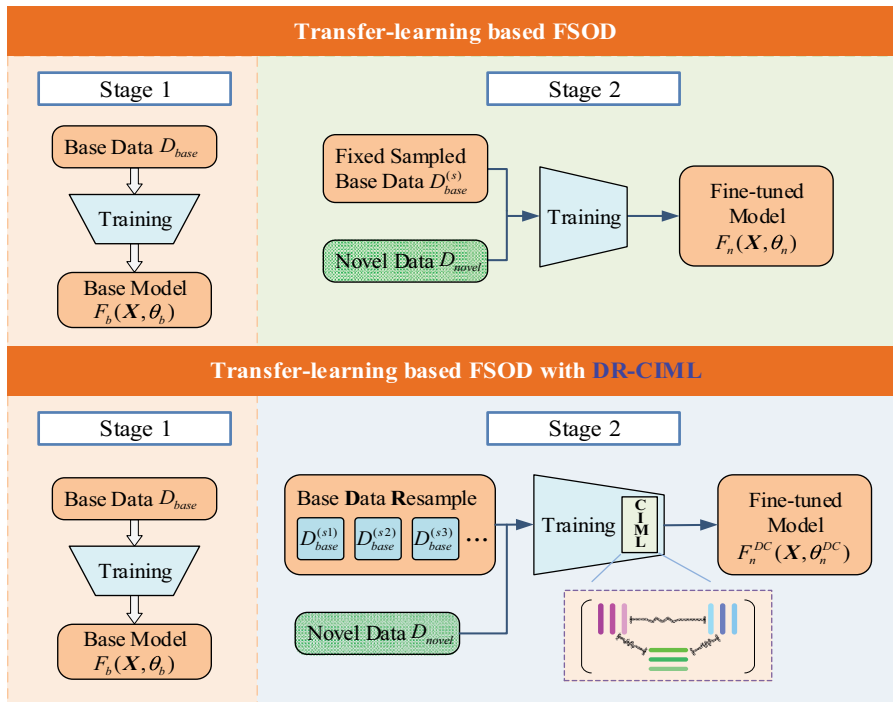


Figure 1. Modifications and improvements of DR-CIML applied to transfer-learning-based FSOD.

AP50 (bAP50) on all shots after fine-tuning on PASCAL VOC, even exceeding the bAP50 performance of the base training stage (80.8) when  $K > 3$ . Competitive scores were also achieved on MS COCO (Lin et al. 2014) under the  $K$ -shot setting with  $K = 10, 30$ . The nAP75 performance improved by up to 1.6 points.

The remainder of this paper is organized as follows. Section 2 introduces work related to few-shot object detection. Section 3 introduces our proposed data resampling and cross-iteration metric-learning methods. Section 4 discusses our experimental results and provides a comparative analysis. Section 5 summarizes the paper.

## Related Work

Most FSOD methods were developed in the context of few-shot classification (Sun et al. 2021). However, it is more difficult than the classification task because it must accurately and simultaneously locate and classify objects (Zhuang et al. 2020). Many FSOD approaches are based on meta-learning and feature re-weighting to avoid overfitting and learn to learn (Finn, Abbeel, and Levine 2017; Han et al. 2022; Karlinsky et al. 2019; Li et al. 2020; Michaelis et al. 2018; Nichol, Achiam, and Schulman 2018; Wang, Ramanan, and Hebert 2019). Recent transfer-learning-based FSOD methods have shown strong generalization capability (Fan et al. 2021; Li et al. 2021; Sun et al. 2021; Wang et al. 2020; Wu et al. 2020; Zhang, Wang, and Forsyth 2020), surpassing many methods based on meta-learning.

## Meta-Learning

Meta-learning aims to learn meta-knowledge through episodic training, so as to quickly learn new concepts through small amounts of labeled data.

In the MAML method (Finn, Abbeel, and Levine 2017), the meta-network assigns parameters to each  $n$ -way  $k$ -shot subtask. The sub-network performs one-step learning and parameter updating on the supply set of the subtask. The query set of the subtask is used to calculate the sub-network loss, and the gradient is calculated to update the meta-network parameters.

Reptile (Nichol, Achiam, and Schulman 2018) improves MAML by updating meta-network parameters through the difference between them and sub-network parameters instead of the gradient of subtasks.

With an episodic training scheme, meta-learning inadequately trains existing data, but the model can quickly learn new concepts through a small amount of annotated data. Therefore, meta-learning pays more attention to the future potential of initialization parameters than pre-training, and not the current performance on multitasking

## **Metric Learning**

Metric learning learns feature embedding, where inputs with similar content are encoded in features with small metric distances, while coded features from different types of inputs should be far from each other (Kaya and Bilge 2019), so as to obtain better feature representation ability for more accurate classification prediction (Weinberger, Blitzer, and Saul 2005; Xing et al. 2002). Basic metric distance calculation methods include Euclidean, Mahalanobis, Matusita (Matusita et al. 1955), Bhattacharyya (Aherne, Thacker, and Rockett 1998), and Kullback Leibler (Elgammal, Duraiswami, and Davis 2003). A metric-learning loss function such as triplet loss (Weinberger and Saul 2009) or its variant form (Aganian et al. 2021) can shorten the distance between the anchor and positive examples, and increase the distance between the anchor and negative examples, and is suitable for few-shot learning tasks. Because the learned feature-embedding network usually has good generalization, the model can make metric-based decisions without further training for unseen objects (Köhler, Eisenbach, and Gross 2021). For example, in the inference stage of the classification task, the feature embedding of the test image is compared with those of the novel categories, and the class corresponding to the nearest feature embeddings is the recognized class.

Therefore, metric learning is conducive to the alleviation of the misclassification of novel categories in few-shot object detection.

## **Meta-Learning-Based Few-Shot Object Detection**

Meta-learning-based FSOD includes dual-branch (Han et al. 2022; Li et al. 2020; Michaelis et al. 2018; Yan et al. 2019) and single-branch (Karlinsky et al. 2019; Wang, Ramanan, and Hebert 2019) methods, both of which utilize episodic training.

Dual-branch methods consist of a query branch Q and support branch S, which share the backbone. Q extracts the query RoI features through a region proposal network (RPN) and RoI Align, and S extracts the representative support feature vector of each category. Therefore, the query RoI features and support feature vectors can be aggregated, and these are input to the RoI head for bounding box regression and binary classification. Dual-branch methods vary most in the means of aggregation between RoIs, and support feature vectors are employed. Meta R-CNN (Yan et al. 2019) takes channel-wise soft-attention on RoI features to remodel the predictor head when more complicated aggregation approaches are adopted by OSWF (Li et al. 2020), OSIS (Michaelis et al. 2018), and Meta Faster R-CNN (Han et al. 2022). Although dual-branch methods allow the quick learning of new categories without fine-tuning in meta-testing, they demand complex episodic training.

As the support category increases, to aggregate separately for each category requires more RAM.

Without Q and S branches, single-branch methods obtain more discriminative features through metric learning or diminish learnable parameters when training novel data. (Karlinsky et al. 2019) calculated the similarity between the embedded feature vector of RoI and category-representative vectors, exploiting extra embedding loss to learn discriminative feature embeddings.

### ***Transfer-Learning-Based Few-Shot Object Detection***

Compared with meta-learning-based FSOD methods, which require complex episodic training, transfer-learning-based FSOD methods utilize a relatively simple two-stage approach on a single-branch architecture. In the first stage, the detector is trained on all base categories. In the second stage, unfrozen layers on the balanced base and novel categories are fine-tuned, while freezing the other components of the model. There are many modifications based on this.

#### ***Modifications of RPN***

CoRPN (Zhang, Wang, and Forsyth 2020) replaces the single binary classifier in the original RPN with  $N$  binary classifiers to avoid missing the foreground RoI from RPN. FSCE (Sun et al. 2021) doubles the maximum number of proposals kept after Non-Maximum Suppression (NMS) to avoid abandoning the foreground RoI. RPN parameters are learnable in the fine-tuning stage to benefit novel detection results.

#### ***Modifications of FPN***

FSCE is based on the assumption that fine-tuning FPN parameters in the second stage performs better than freezing them. MPSR (Wu et al. 2020) implements the FPN processing of multiscale positive sample refinement through object pyramids, so as to expand the scale distribution of novel categories and reduce improper negative samples containing a large proportion of positive instances.

#### ***Modifications of Loss Function***

MPSR applies a refinement branch, adding the extra classification loss of the extracted multiscale positive samples to the RPN loss function and ROI loss function of Faster R-CNN. FSCE employs contrastive proposal encoding (CPE) loss to promote the compactness of intra-class instances. CGDP +FSCN (Li et al. 2021) applies additional semi-supervised loss to exploit unlabeled instances, thereby promoting the learning of sparse novel category objects.



### **Maintaining Performance on Base Categories**

Retentive R-CNN (Fan et al. 2021) utilizes separate classification heads for novel and base categories to avoid the catastrophic forgetting of base categories.

To sum up, transfer-learning-based FSOD does not need complex episodic training; it achieves state-of-the-art (SOTA) performance by appropriately freezing network components or modifying the loss function.

Nonetheless, existing transfer-learning-based FSOD approaches are trained on fixed base class objects and novel class objects, and it is obvious that the diversity of the abundant annotated base data is not fully utilized. These mean that the model still has room to improve the detection performance. Therefore, we expect that our transfer-learning-based DR-CIML can make full use of base class data.

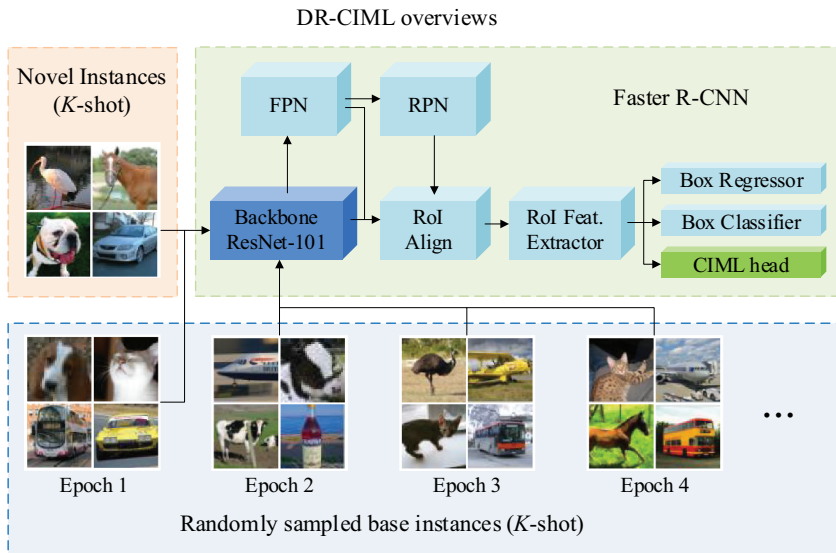
### **Method**

Our proposed method follows the standard transfer-learning-based FSOD methods (Sun et al. 2021; Wang et al. 2020). We explore some neglected properties of the abundant base data for fine-tuning. The training scheme has two stages. The first stage is training Faster R-CNN with abundant base data. In the second stage, the metric-learning branch is embedded in the RoI feature extractor, and the base model is fine-tuned with sufficient base data and sparse novel data. In addition, we freeze different components of the model according to particular K-shot tasks. We optimize the model by jointly optimizing the RPN, regression, and classification loss of the standard Faster R-CNN, as well as the cross-iteration metric-learning loss added in the fine-tuning stage. Figure 2 shows the structure of the DR-CIML method.

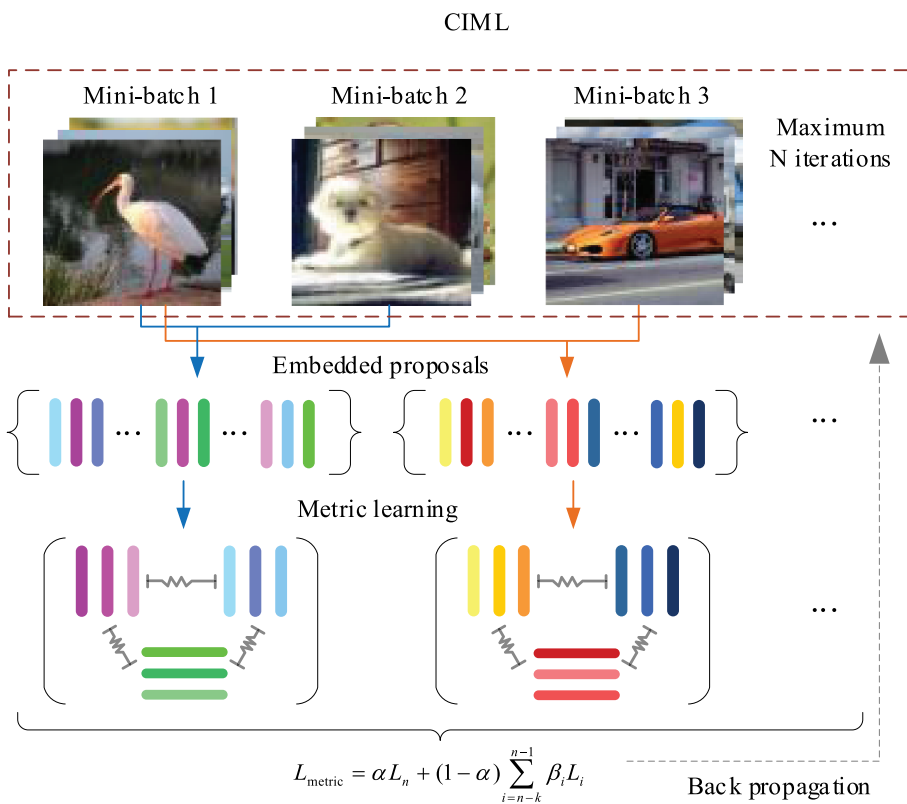
### **Data Resampling**

Standard FSOD methods adopt a unified dataset organization. In the basic training phase, the model is trained on all base data. In the fine-tuning phase, we utilize the fixed-base and novel class samples under the  $K$ -shot setting. This dataset organization has the advantage of avoiding the imbalance of base and novel categories, as well as constructing practical few-shot application scenarios. The disadvantage is that a large amount of base-class data is not fully exploited. Hence the data resampling organization technique is proposed. The basic training stage is the same as with the standard FSOD. However, in the fine-tuning stage, novel class data  $D_{novel}$  are organized in the same way as Wang et al. (2020), Sun et al. (2021), and Kang et al. (2019), and base-class data  $D_{base}$  consist of all data that do not contain  $D_{novel}$ , i.e.,  $D_{base} \cap D_{novel} = \emptyset$ . Because the data volume of  $D_{base}$  is much larger than that of  $D_{novel}$ , at the beginning of each epoch,  $D_{base}$  is randomly sampled to construct a balanced





**Figure 2.** Structure of DR-CIML method.



**Figure 3.** Overview of cross-iteration metric-learning (CIML) algorithm.

sub-dataset under the  $K$ -shot setting. Data resampling has two advantages: 1) We have constructed a new balanced sub-dataset that conforms to the FSOD application scenario, so that we can make full use of the base-class data and reduce performance degradation on the base categories; 2) It improves the diversity of sample features, so as to promote the inter-class distance between novel and base categories (Figure 3).

### Cross-Iteration Metric-Learning Branch

Standard Faster R-CNN extracts features with its backbone. RPN regresses the bounding box of the predefined anchor and decides whether it is foreground or background. Then region proposals are given by RoI Align. Finally, the regressor of the RoI head fine-tunes the location of region proposals again, and the classifier of the RoI head classifies objects contained in region proposals. Among them, the optimization goal of the classifier is a one-hot vector. In the scenario with large-scale training data, the optimization goal of the classifier is applicable, but it will decrease the robustness of the model while it lacks training data. From this point of view, we embed the metric-learning branch in the RoI feature extractor and calculate the similarity of object features generated by RPN. Specifically, the triple loss (Schroff, Kalenichenko, and Philbin 2015) function is applied to increase intra-category feature propinquity and inter-category feature discrimination, so as to reduce misclassification. Our metric loss function is

$$f_{metric}(x_i^a, x_i^p, x_i^n) = \max \left\{ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha_m, 0 \right\} \quad (1)$$

where  $f(x) \in R^d$  embeds a proposal  $x$  into a  $d$ -dimensional Euclidean space,  $x_i^a$  is a proposal of a specific object (anchor),  $x_i^p$  is a proposal of the same class (positive),  $x_i^n$  is a proposal of any other class (negative), and  $\alpha_m$  is the margin enforced between positive and negative pairs.

However, the quality of region proposals from RPN is uneven, including both high-quality proposals with high IoU with ground truth, and low-quality ones with low IoU with ground truth. Although low-quality proposals with low IoU can be filtered by the IoU threshold  $T_{IoU}$ , to select an appropriate  $T_{IoU}$  is usually problematic. When  $T_{IoU}$  is undersized, numerous low-quality proposals containing superfluous backgrounds will be kept, which leads to negative optimization of the model, and when  $T_{IoU}$  is too large, few high-quality proposals are preserved, which will make it hard to optimize the model. Therefore, to improve the diversity of features when metric-learning, we consider that adjacent iterations contain features that can be fed into the metric-learning branch to optimize the model. Specifically, we employ a cross-iteration metric-learning branch. Our cross-iteration metric loss function is

$$L_{metric} = \alpha L_n + (1 - \alpha) \sum_{i=n-k}^{n-1} \beta_i L_i \quad (2)$$

$$L_i = f_{metric}(F_n \cup F_i) \quad (3)$$

$$L_n = f_{metric}(F_n) \quad (4)$$

$$\beta_i = (0.9)^{n-i} \quad (5)$$

where  $n$  is the current iteration;  $F_i$  denotes the input features of the metric branch of the  $i$ th iteration, i.e., the filtered RPN output;  $\alpha$  is the weight coefficient of the metric loss of the current iteration; and  $\beta_i$  is the weight coefficient of the metric loss of all feature embeddings of the current and the  $i$ th iterations.

Therefore, the joint optimization objectives of the model are

$$L = \lambda_{rpn} L_{rpn} + \lambda_{cls} L_{cls} + \lambda_{reg} L_{reg} + \lambda_{metric} L_{metric} \quad (6)$$

where  $L_{rpn}$  utilizes binary cross-entropy loss to generate foreground proposals;  $L_{cls}$  utilizes cross-entropy loss for bounding box classifiers;  $L_{reg}$  utilizes smooth-L1 loss for bounding box regression deltas;  $L_{metric}$  is the metric loss; and  $\lambda_{rpn}$ ,  $\lambda_{cls}$ ,  $\lambda_{reg}$ , and  $\lambda_{metric}$  are the weight coefficients of  $L_{rpn}$ ,  $L_{cls}$ ,  $L_{reg}$ , and  $L_{metric}$ , respectively. Our revised joint loss functions are improved based on the standard Faster R-CNN loss (Ren et al. 2015).

## Experiments

Comprehensive experiments were conducted on PASCAL VOC and MS COCO, and our proposed method showed competitive scores. We followed the dataset division method of Wang et al. (2020), Sun et al. (2021), and Kang et al. (2019) in order to provide reliable comparative evaluation results. We provide implementation details and results of comparative and ablation experiments, as well as visualization outcomes.

## Implementation

Faster R-CNN with ResNet-101 and FPN were employed as our few-shot object detector, and a single Nvidia GeForce RTX 3080 Ti was used to accelerate graphic calculation while loading two images per iteration. Because it would lead to gradient oscillation and non-convergence if the batch were undersized, we adopted the accumulate gradients approach which updates parameters once every  $n$  batches trained to update the parameters every eight iterations, so as to increase the batch size to 16. The

maximum number of iterations was 48,000. The optimizer was SGD, with momentum 0.9 and weight decay  $1e-4$ . The learning rate scheduler adopted linear preheating and cosine attenuation. Multiscale training, random flipping, image mosaic, and other data-enhancement methods were adopted.

## **Few-Shot Object Detection Benchmarks**

### **Pascal Voc**

We used the dataset division of Wang et al. (2020), Sun et al. (2021), and Kang et al. (2019), randomly dividing PASCAL VOC (Everingham et al. 2010, 2015) into split 1, split 2, and split 3, each containing 15 base categories with abundant instances and five novel categories sampled from training data under the  $K$ -shot setting with  $K = 1, 2, 3, 5, \text{ and } 10$ . In the base training stage, the detector was trained on all annotated base categories. In the fine-tuning stage, balanced base category instances and novel category instances were utilized with  $K$ -shot, with the modification that the training scheme of data resampling was adopted to fully utilize the diversity of the abundant base data. We evaluated AP50 for novel categories (nAP50) and base categories (bAP50) on the PASCAL VOC2007 test set.

### **Ms Coco**

There are 80 categories in MS COCO (Lin et al. 2014), which were divided into 60 base categories and 20 novel categories with  $K = 10, 30$ . We report novel AP50–95 and novel AP75 on 5,000 images of COCO2014val.

## **Few-Shot Object Detection Comparison Results**

### **Results on PASCAL VOC**

Table 1 compares nAP50 between our proposed method and existing methods on PASCAL VOC with three novel splits. Our proposed method reaches the highest nAP50 in different splits under  $K$ -shot settings with  $K = 1, 2, 3, 5, 10$ , and nAP50 improves by up to 6.3 points, which fully demonstrates the effectiveness of our method. Moreover, for further demonstration the generality of our DR-CIML, we implement them on multiple baselines, as shown in Table 2. Both nAP50 and bAP50 have been improved which fully verifies the generalization ability of the DR-CIML for different baselines.

The bAP50 performance on three base splits is shown in Table 3. Obviously, the proposed method is the first to achieve  $>80$  bAP50 on all shots after fine-tuning, even exceeding bAP50 of the base training stage (80.8) when  $K > 3$ . Besides, its score slightly lower than that of the base training stage when  $K = 3$  mainly because the sampled base class data is insufficient. However, the bAP50 score of the other methods decreased significantly, FSCE reduced by 6.7% and TFA reduced by 2.4%. This shows the strong capacity of DR-CIML to retain



**Table 1.** nAP50 performance of existing FSOD methods on three PASCAL VOC novel splits with  $K = 1, 2, 3, 5$ , and  $10$ . “●” represents meta-learning-based methods. “◇” represents transfer-learning-based methods. “-” represents unreported results of other methods.

Method/Shot	Backbone	Novel Split 1					Novel Split 2					Novel Split 3				
		1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
● MetaYOLO (Kang et al. 2019)	YOLOv2	14.8	15.5	26.7	33.9	47.2	15.7	15.3	22.7	30.1	40.5	21.3	25.6	28.4	42.8	45.9
● RepMet (Karlinsky et al. 2019)	ResNet-101	26.1	32.9	34.4	38.6	41.3	17.2	22.1	23.4	28.3	35.8	27.5	31.1	31.5	34.4	37.2
● Meta R-CNN (Yan et al. 2019)	ResNet-101	19.9	25.5	35.0	45.7	51.5	10.4	19.4	29.6	34.8	45.4	14.3	18.2	27.5	41.2	48.1
◇ TFA w/cos (Wang et al. 2020)	ResNet-101	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
◇ CoRPN w/cos (Zhang, Wang, and Forsyth 2020)	ResNet-101	44.4	38.5	46.4	54.1	55.7	25.7	29.5	37.3	36.2	41.3	35.8	41.8	44.6	51.6	49.6
◇ MPSR (Wu et al. 2020)	ResNet-101	41.7	-	51.4	55.2	61.8	24.4	-	39.2	39.9	47.8	35.6	-	42.3	48.0	49.7
● DCNet (Hu et al. 2021)	ResNet-101	33.9	37.4	43.7	51.1	59.6	23.2	24.8	30.6	36.7	46.6	32.3	34.9	39.7	42.6	50.7
● TIP (Li and Li 2021)	ResNet-101	27.7	36.5	43.3	50.2	59.6	22.7	30.1	33.8	40.9	46.9	21.7	30.6	38.1	44.5	50.9
◇ Halluc. (TFA) (Zhang and Wang 2021)	ResNet-101	45.1	44.0	44.7	55.0	55.9	23.2	27.5	35.1	34.9	39.0	30.5	35.1	41.4	49.0	49.3
◇ CGDP+FSFN (Li et al. 2021)	ResNet-50	40.7	45.1	46.5	57.4	62.4	27.3	31.4	40.8	42.7	46.3	31.2	36.4	43.7	50.1	55.6
◇ SRR-FSD (Zhu et al. 2021)	ResNet-101	<b>47.8</b>	<b>50.5</b>	51.3	55.2	56.8	<b>32.5</b>	35.3	39.1	40.8	43.8	<b>40.1</b>	41.5	44.3	46.9	46.4
◇ FSOD-UP (Wu et al. 2021)	ResNet-101	43.8	47.8	50.3	55.4	61.7	31.2	30.5	41.2	42.2	48.3	35.5	39.7	43.9	50.6	53.5
◇ FSCE (Sun et al. 2021)	ResNet-101	44.2	43.8	51.4	61.9	63.4	27.1	29.5	43.5	44.2	50.2	37.2	41.9	47.5	54.6	<b>58.5</b>
◇ FSCE + DR-CIML (ours)	ResNet-101	47.4	48.7	<b>57.3</b>	<b>62.7</b>	<b>65.3</b>	30.7	<b>35.8</b>	<b>47.9</b>	<b>49.8</b>	<b>53.1</b>	37.2	<b>42.2</b>	<b>51.2</b>	<b>55.8</b>	57.4

**Table 2.** Apply DR-CIML to different baselines on PASCAL VOC split 1 with  $K=3, 5, 10$ . “ $\diamond$ ” represents transfer-learning-based methods. “-” represents unreported results of other methods.

Method/Shot	Novel Split 1			Base Split 1		
	3	5	10	3	5	10
$\diamond$ TFA w/cos (Wang et al. 2020)	44.7	55.7	56.0	79.1	-	78.4
$\diamond$ TFA + DR-CIML (ours)	53.2 (+8.5%)	62.7 (+7.0%)	64.0 (+8.0%)	79.2 (+0.1%)	78.3	79.0 (+0.6%)
$\diamond$ FSCE (Sun et al. 2021)	51.4	61.9	63.4	74.1	76.6	-
$\diamond$ FSCE + DR-CIML (ours)	57.3 (+5.9%)	62.7 (+0.8%)	65.3 (+1.9%)	80.5 (+6.4%)	80.9 (+4.3%)	81.0

**Table 3.** bAP50 of existing FSOD methods on three PASCAL VOC base splits. “-” represents unreported results of other methods.

Method/Shot	Base Split 1		
	3	5	10
● Meta YOLO (Kang et al. 2019)	64.8	-	69.7
● Meta R-CNN (Yan et al. 2019)	64.8	-	67.9
$\diamond$ MPSR (Wu et al. 2020)	67.8	-	71.8
● PNPDet (Zhang et al. 2021)	75.5	-	75.5
$\diamond$ TFA w/cos (Wang et al. 2020)	79.1	-	78.4
$\diamond$ FSCE (Sun et al. 2021)	74.1	76.6	-
Train $D_{base}$ only (Wang et al. 2020)	<b>80.8</b>	80.8	80.8
$\diamond$ FSCE + DR-CIML (ours)	80.5	<b>80.9</b>	<b>81.0</b>

base category knowledge in the fine-tuning stage. DR-CIML improves accuracy while incurring no extra inference overhead.

### Results on MS COCO

Table 4 compares the results (nAP50–95 and nAP75) of the proposed and existing methods with  $K=10, 30$ . Our method surpasses many methods, with nAP50–95 and nAP75 improved by up to 1.6 points, which fully verifies the generalization ability of the proposed method for different datasets.

### Ablation Research and Visualization

Modifications were implemented in the fine-tuning stage, and the baseline was the standard FSCE (Sun et al. 2021) which employs contrastive proposal encoding (CPE) loss to promote instance-level intra-class compactness and inter-class variance. DR-CIML increases instance diversity through base data resampling (DR) to fully utilize source domain data and increases feature

**Table 4.** Evaluation results of existing FSOD methods on two MS COCO novel splits.

Method/Shot	Novel AP50–95		Novel AP75	
	10	30	10	30
● MetaYOLO (Kang et al. 2019)	5.6	9.1	4.6	7.6
◇ CoRPN w/cos (Zhang, Wang, and Forsyth 2020)	9.0	13.9	8.3	13.9
● Meta-RCNN (Yan et al. 2019)	8.7	12.4	6.6	10.8
◇ MPSR (Wu et al. 2020)	9.8	14.1	9.7	14.2
◇ TFA w/cos (Wang et al. 2020)	10.0	13.7	9.3	13.4
● QA-FewDet (Han et al. 2021)	10.2	16.5	9.0	15.5
◇ FSCE (Sun et al. 2021)	11.9	16.4	10.5	16.2
◇ SVD (FSCE) (Wu et al. 2021)	12.0	16.2	10.4	15.9
◇ FSCE + DR-CIML (ours)	<b>13.6</b>	<b>17.2</b>	<b>12.0</b>	<b>16.6</b>

diversity of comparative learning through CIML. We performed ablation experiments combining DR, CPE, or CIML components when fine-tuning the model. The ablation results obtained from PASCAL VOC split 1 are shown in Table 5.

#### *Ablation for Base Data Resampling*

FSCE fine-tunes on the fixed base categories, so that the diversity of base data is not fully utilized. We promote this by adopting the training scheme of data resampling (DR). The results of (Exp1, Exp3) and (Exp2, Exp4) show that under different metric-learning methods, the DR strategy can improve both nAP50 and bAP50, and even bAP50 surpasses that of the base model (80.8) obtained in the basic training stage, which indicates that DR has a strong ability to maintain base category knowledge.

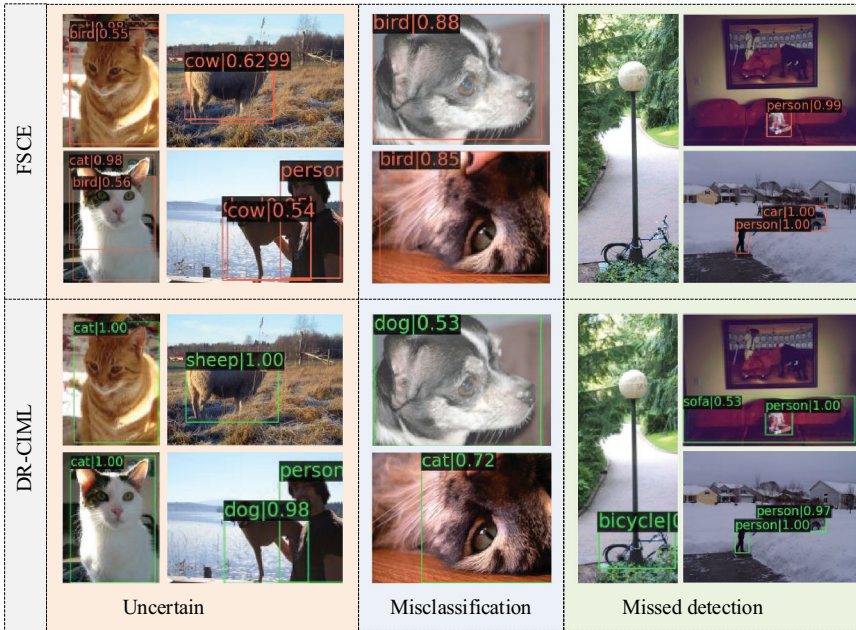
#### *Ablation for Metric-Learning Branch*

We also explored the effect of the cross-iteration metric-learning (CIML) branch. The experimental results of (Exp1, Exp2) show that nAP50 is increased by up to 4.7 points with  $K = 2$ , as compared to CPE. (Exp3, Exp4) show that both nAP50 and bAP50 are improved through DR-CIML, which

**Table 5.** Ablation for data resample organization and cross-iteration metric learning; results gained from PASCAL VOC split 1. “–” represents unreported results of other methods.

Model	Exp	Data Resample	Metric Learning	Base AP50					Novel AP50				
				1	2	3	5	10	1	2	3	5	10
ResNet-101	Exp1	×	CPE	78.9	–	74.1	76.6	–	44.2	43.8	51.4	61.9	63.4
	Exp2	×	CIML	77.5	77.9	74.4	75.1	76.8	45.2	48.5	55.1	<b>63.4</b>	65.0
	Exp3	√	CPE	80.4	80.4	80.1	80.4	<b>81.0</b>	45.2	48.1	57.1	62.7	65.2
	Exp4	√	CIML	<b>80.6</b>	<b>80.6</b>	<b>80.5</b>	<b>80.9</b>	<b>81.0</b>	<b>47.4</b>	<b>48.7</b>	<b>57.3</b>	62.7	<b>65.3</b>





**Figure 4.** Visual detection results of our method and standard FSCE.

demonstrates that our CIML branch can improve intra-category feature proximity and inter-category feature discrimination.

To sum up, the experimental results show that the DR and CIML branches can both promote model performance, and only the combined DR-CIML can maximize this. Furthermore, DR-CIML does not lead to extra inference cost, so the inference speed is the same as that of Faster R-CNN.

### *Visualization for Analysis*

Figure 4 shows the visualization results of our method and the standard FSCE. It is found that our proposed DR-CIML method improves misclassification, uncertain recognition, and missed detection. For example, our method would unlikely to recognize dogs and cats as birds or cows, it can also detect the bicycle in Figure 4 while FSCE ignores it. Therefore, DR-CIML learns superior semantic and spatial information.

## **Conclusion**

We explored the deficiencies of transfer-learning-based FSOD methods in data utilization. We are the first to apply data resampling organization and cross-iteration metric learning (DR-CIML) in the transfer-learning-based detection method, so as to make full use of the diversity of base-class data and increase the feature diversity of metric learning. Extensive experiments in PASCAL VOC and MS COCO fully verified the effectiveness of the data resampling method

applied to transfer-learning-based FSOD. Our proposed method is independent of models and datasets; hence it can be readily embedded in any object detector without extra inference overhead. FSOD is a challenging task, and we hope our work can inspire more research on FSOD regarding data resampling and visual feature metric-learning. In the future, we will study the effectiveness of data resampling in few-shot segmentation.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China, Grant/Award Number: 61973075; Special Funds of the Jiangsu Provincial Key Research and Development Projects, Grant/Award Number: BE2019612; and Jiangsu Provincial Cadre Health Research Projects, Grant/Award Number: BJ17006. We thank LetPub ([www.letpub.com](http://www.letpub.com)) for its linguistic assistance during the preparation of this manuscript.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The work was supported by the National Natural Science Foundation of China [61973075]; Jiangsu Provincial Cadre Health Research Projects [BJ17006]; Special Funds of the Jiangsu Provincial Key Research and Development Projects [BE2019612]

## ORCID

Wei Zhou  <http://orcid.org/0000-0002-5794-7567>

Xudong Yang  <http://orcid.org/0000-0003-3062-2153>

Lin Chai  <http://orcid.org/0000-0002-3960-8828>

## References

- Aganian, D., M. Eisenbach, J. Wagner, D. Seichter, and H. M. Gross. 2021. Revisiting loss functions for person re-identification. In *International Conference on Artificial Neural Networks* (pp. 30–42). Springer, Cham. doi: [10.1007/978-3-030-86383-8\\_3](https://doi.org/10.1007/978-3-030-86383-8_3).
- Aherne, F. J., N. A. Thacker, and P. I. Rockett. 1998. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika* 34 (4):363–68.
- Bochkovskiy, A., C. Y. Wang, and H. Y. M. Liao. 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. doi:[10.48550/arXiv.2004.10934](https://doi.org/10.48550/arXiv.2004.10934).
- Chen, H., Y. Wang, G. Wang, and Y. Qiao. 2018. Lstd: A low-shot transfer detector for object detection. *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1). doi:[10.1609/aaai.v32i1.11716](https://doi.org/10.1609/aaai.v32i1.11716).

- Deng, J., W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–55). Ieee. doi: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- Elgammal, A., R. Duraiswami, and L. S. Davis 2003. Probabilistic tracking in joint feature-spatial spaces. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings*. (Vol. 1, pp. I–I). IEEE. doi: [10.1109/CVPR.2003.1211432](https://doi.org/10.1109/CVPR.2003.1211432).
- Everingham, M., S. M. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. 2015. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision* 111 (1):98–136. doi:[10.1007/s11263-014-0733-5](https://doi.org/10.1007/s11263-014-0733-5).
- Everingham, M., L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. 2010. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision* 88 (2):303–38. doi:[10.1007/s11263-009-0275-4](https://doi.org/10.1007/s11263-009-0275-4).
- Fan, Z., Y. Ma, Z. Li, and J. Sun 2021. Generalized few-shot object detection without forgetting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4527–36). doi: [10.1109/CVPR46437.2021.00450](https://doi.org/10.1109/CVPR46437.2021.00450).
- Finn, C., P. Abbeel, and S. Levine 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International conference on machine learning* (pp. 1126–35). PMLR.
- Han, G., Y. He, S. Huang, J. Ma, and S. F. Chang 2021. Query adaptive few-shot object detection with heterogeneous graph convolutional networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3263–72). doi: [10.1109/ICCV48922.2021.00325](https://doi.org/10.1109/ICCV48922.2021.00325).
- Han, G., S. Huang, J. Ma, Y. He, and S. F. Chang 2022. Meta faster r-cnn: Towards accurate few-shot object detection with attentive feature alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 1, pp. 780–89). doi: [10.1609/aaai.v36i1.19959](https://doi.org/10.1609/aaai.v36i1.19959).
- He, K., X. Zhang, S. Ren, and J. Sun 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–78). doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- Hu, H., S. Bai, A. Li, J. Cui, and L. Wang 2021. Dense relation distillation with context-aware aggregation for few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10185–94). doi: [10.1109/CVPR46437.2021.01005](https://doi.org/10.1109/CVPR46437.2021.01005).
- Kang, B., Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell 2019. Few-shot object detection via feature reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8420–29). doi: [10.1109/ICCV.2019.00851](https://doi.org/10.1109/ICCV.2019.00851).
- Karlinisky, L., J. Shtok, S. Harary, E. Schwartz, A. Aides, R. Feris, and A. M. Bronstein 2019. Repmet: Representative-based metric learning for classification and few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5197–206). doi: [10.1109/CVPR.2019.00534](https://doi.org/10.1109/CVPR.2019.00534).
- Katzmann, A., O. Taubmann, S. Ahmad, A. Mühlberg, M. Sühling, and H. M. Groß. 2021. Explaining clinical decision support systems in medical imaging using cycle-consistent activation maximization. *Neurocomputing* 458:141–56. doi:[10.1016/j.neucom.2021.05.081](https://doi.org/10.1016/j.neucom.2021.05.081).
- Kaya, M., and H. Ş. Bilge. 2019. Deep metric learning: A survey. *Symmetry* 11 (9):1066. doi:[10.3390/sym11091066](https://doi.org/10.3390/sym11091066).
- Khosla, P., P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, and D. Krishnan. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems* 33:18661–73. doi:[10.48550/arXiv.2004.11362](https://doi.org/10.48550/arXiv.2004.11362).
- Köhler, M., M. Eisenbach, and H. M. Gross. 2021. Few-shot object detection: A Survey. *arXiv preprint arXiv:2112.11699*. doi:[10.48550/arXiv.2112.11699](https://doi.org/10.48550/arXiv.2112.11699).

- Li, A., and Z. Li 2021. Transformation invariant few-shot object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3094–102). doi: [10.1109/CVPR46437.2021.00311](https://doi.org/10.1109/CVPR46437.2021.00311).
- Lin, T. Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, and C. L. Zitnick 2014. Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–55). Springer, Cham. doi: [10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48).
- Li, X., L. Zhang, Y. P. Chen, Y. W. Tai, and C. K. Tang. 2020. One-shot object detection without fine-tuning. *arXiv preprint arXiv:2005.03819*. doi:[10.48550/arXiv.2005.03819](https://doi.org/10.48550/arXiv.2005.03819).
- Li, Y., H. Zhu, Y. Cheng, W. Wang, C. S. Teo, C. Xiang, and T. H. Lee 2021. Few-shot object detection via classification refinement and distractor retreatment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 15395–403). doi: [10.1109/CVPR46437.2021.01514](https://doi.org/10.1109/CVPR46437.2021.01514).
- Lu, J., P. Gong, J. Ye, and C. Zhang. 2020. Learning from very few samples: A survey. *arXiv preprint arXiv:2009.02653*. doi:[10.48550/arXiv.2009.02653](https://doi.org/10.48550/arXiv.2009.02653).
- Mannocci, L., S. Villon, M. Chaumont, N. Guellati, N. Mouquet, C. Iovan, and D. Mouillot. 2022. Leveraging social media and deep learning to detect rare megafauna in video surveys. *Conservation Biology* 36 (1):e13798. doi:[10.1111/cobi.13798](https://doi.org/10.1111/cobi.13798).
- Matusita, K. 1955. Decision rules, based on the distance, for problems of fit, two samples, and estimation. *The Annals of Mathematical Statistics* 26 (4):631–40. doi:[10.1214/aoms/1177728422](https://doi.org/10.1214/aoms/1177728422).
- Michaelis, C., I. Ustyuzhaninov, M. Bethge, and A. S. Ecker. 2018. One-shot instance segmentation. *arXiv preprint arXiv:1811.11507*. doi:[10.48550/arXiv.1811.11507](https://doi.org/10.48550/arXiv.1811.11507).
- Nichol, A., J. Achiam, and J. Schulman. 2018. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*. doi:[10.48550/arXiv.1803.02999](https://doi.org/10.48550/arXiv.1803.02999).
- Ren, S., K. He, R. Girshick, and J. Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems* 28. doi:[10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- Samuelson, L. K., and L. B. Smith. 2005. They call it like they see it: Spontaneous naming and attention to shape. *Developmental science* 8 (2):182–98. doi:[10.1111/j.1467-7687.2005.00405.x](https://doi.org/10.1111/j.1467-7687.2005.00405.x).
- Schroff, F., D. Kalenichenko, and J. Philbin 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–23). doi: [10.1109/CVPR.2015.7298682](https://doi.org/10.1109/CVPR.2015.7298682).
- Simonyan, K., and A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. doi:[10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
- Sun, Y., Y. Chen, X. Wang, and X. Tang. 2014. Deep learning face representation by joint identification-verification. *Advances in Neural Information Processing Systems* 27.
- Sun, B., B. Li, S. Cai, Y. Yuan, and C. Zhang 2021. Fscf: Few-shot object detection via contrastive proposal encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7352–62). doi: [10.1109/CVPR46437.2021.00727](https://doi.org/10.1109/CVPR46437.2021.00727).
- Tan, M., and Q. Le 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–14). PMLR. doi: [10.48550/arXiv.1905.11946](https://doi.org/10.48550/arXiv.1905.11946).
- Tian, Z., C. Shen, H. Chen, and T. He 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9627–36). doi: [10.1109/ICCV.2019.00972](https://doi.org/10.1109/ICCV.2019.00972).
- Vinyals, O., C. Blundell, T. Lillicrap, and D. Wierstra. 2016. Matching networks for one shot learning. *Advances in Neural Information Processing Systems* 29.
- Wang, X., T. E. Huang, T. Darrell, J. E. Gonzalez, and F. Yu. 2020. Frustratingly simple few-shot object detection. *arXiv preprint arXiv:2003.06957*. doi:[10.48550/arXiv.2003.06957](https://doi.org/10.48550/arXiv.2003.06957).

- Wang, Y. X., D. Ramanan, and M. Hebert 2019. Meta-learning to detect rare objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9925–34). doi: [10.1109/ICCV.2019.01002](https://doi.org/10.1109/ICCV.2019.01002).
- Weinberger, K. Q., J. Blitzer, and L. Saul. 2005. An information maximization model of eye movements. *Advances in Neural Information Processing Systems* 17:18.
- Weinberger, K. Q., and L. K. Saul. 2009. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research* 10 (2).
- Wu, A., Y. Han, L. Zhu, and Y. Yang 2021. Universal-prototype enhancing for few-shot object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9567–76). doi: [10.1109/ICCV48922.2021.00943](https://doi.org/10.1109/ICCV48922.2021.00943).
- Wu, J., S. Liu, D. Huang, and Y. Wang 2020. Multi-scale positive sample refinement for few-shot object detection. In *European conference on computer vision* (pp. 456–72). Springer, Cham. doi: [10.1007/978-3-030-58517-4\\_27](https://doi.org/10.1007/978-3-030-58517-4_27).
- Wu, A., S. Zhao, C. Deng, and W. Liu. 2021. Generalized and discriminative few-shot object detection via SVD-dictionary enhancement. *Advances in Neural Information Processing Systems* 34:6353–64.
- Xing, E., M. Jordan, S. J. Russell, and A. Ng. 2002. Distance metric learning with application to clustering with side-information. *Advances in Neural Information Processing Systems* 15.
- Yan, X., Z. Chen, A. Xu, X. Wang, X. Liang, and L. Lin 2019. Meta r-cnn: Towards general solver for instance-level low-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9577–86). doi: [10.1109/ICCV.2019.00967](https://doi.org/10.1109/ICCV.2019.00967).
- Zhang, G., K. Cui, R. Wu, S. Lu, and Y. Tian 2021. Pnpdet: Efficient few-shot detection without forgetting via plug-and-play sub-networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3823–32). doi: [10.1109/WACV48630.2021.00387](https://doi.org/10.1109/WACV48630.2021.00387).
- Zhang, W., and Y. X. Wang 2021. Hallucination improves few-shot object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13008–17). doi: [10.1109/CVPR46437.2021.01281](https://doi.org/10.1109/CVPR46437.2021.01281).
- Zhang, W., Y. X. Wang, and D. A. Forsyth. 2020. Cooperating RPN's improve few-shot object detection. *arXiv preprint arXiv:2011.10142*. doi:[10.48550/arXiv.2011.10142](https://doi.org/10.48550/arXiv.2011.10142).
- Zhuang, F., Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, and Q. He. 2020. A comprehensive survey on transfer learning. *Proceedings of the IEEE* 109 (1):43–76. doi:[10.1109/JPROC.2020.3004555](https://doi.org/10.1109/JPROC.2020.3004555).
- Zhu, C., F. Chen, U. Ahmed, Z. Shen, and M. Savvides 2021. Semantic relation reasoning for shot-stable few-shot object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8782–91). doi: [10.1109/CVPR46437.2021.00867](https://doi.org/10.1109/CVPR46437.2021.00867).