Few-Shot Object Detection with Model Calibration

Qi Fan[1], Chi-Keung Tang[1], and Yu-Wing Tai[1,2]

[1] The Hong Kong University of Science and Technology, [2] Kuaishou Technology
fanqics@gmail.com, cktang@cs.ust.hk, yuwing@gmail.com

Abstract. Few-shot object detection (FSOD) targets at transferring knowledge from known to unknown classes to detect objects of novel classes. However, previous works ignore the model bias problem inherent in the transfer learning paradigm. Such model bias causes overfitting toward the training classes and destructs the well-learned transferable knowledge. In this paper, we pinpoint and comprehensively investigate the model bias problem in FSOD models and propose a simple yet effective method to address the model bias problem with the facilitation of model calibrations in three levels: 1) Backbone calibration to preserve the well-learned prior knowledge and relieve the model bias toward base classes, 2) RPN calibration to rescue unlabeled objects of novel classes and, 3) Detector calibration to prevent the model bias toward a few training samples for novel classes. Specifically, we leverage the overlooked classification dataset to facilitate our model calibration procedure, which has only been used for pre-training in other related works. We validate the effectiveness of our model calibration method on the popular Pascal VOC and MS COCO datasets, where our method achieves very promising performance. Codes are released at https://github.com/fanq15/FewX.

Keywords: few-shot object detection, model bias, model calibration, uncertainty-aware RPN, detector calibration

1 Introduction

Object detection [26,77,58] is a fundamental and well-studied research problem in computer vision, which is instrumental in many down-stream vision tasks and applications. Current deep learning based methods have achieved significant performance on object detection task by leveraging abundant well-annotated training samples. However, box-level annotation for object detection is very time and labor consuming, and it is impossible to annotate every class in the real world. Typical object detection models [59,75,76,63] degrade when labeled training data are scarce, and fail to detect novel classes unseen in the training set. Few-shot object detection (FSOD) [22,8,41] targets at solving this problem. Given only a few support samples of novel classes, FSOD model can detect objects of the target novel classes in query images.

This research was supported in part by Kuaishou Technology, and the Research Grant Council of the Hong Kong SAR under grant No. 16201420.

Table 1. The bias source of each detection module and the performance improvement from our calibrated model on COCO [60].

Module	Biased Model		Calibrated Model
	Bias Source	mAP	mAP
RPN	Base class	11.3	11.8 (*+0.5*)
Backbone	Base class		14.7 (*+3.4*)
Detector	Novel samples		11.9 (*+0.6*)
Overall	All above		15.1 (*+3.8*)

Few-shot object detection is an emerging task and has received considerable attention very recently. Some works [40,95,94] detect query objects by exploring the relationship between support and query images through a siamese network equipped with meta learning. Other works [91,93,108] adopt a fine-tuning strategy to transfer knowledge priors from base classes to novel classes which have very few annotated samples.

Despite their success, a critical intrinsic issue of few-shot object detection has long been neglected by these previous works: the model bias problem. The problem is caused by the extremely unbalanced training datasets between the base and novel classes. Specifically, to learn general and transferable knowledge, the model is first trained on base classes with numerous annotated samples, where the potential novel classes are labeled as background. Therefore, the model will bias toward only recognizing base classes and reject novel classes in RPN [77]. Second, the backbone feature will bias toward the base classes under the biased training supervision, where the feature distribution learned from the abundant classes from the pretraining dataset [13] will be destroyed with the backbone overfitting to the limited base classes. Third, when the detector is finetuned on novel classes, the very few training samples cannot represent the real class statistics. Therefore, the model will bias toward the limited training samples of novel classes, and thus cannot generalize well to the real data distribution, which can not solved by the class-imbalance learning methods [14,42,89,43,11].

The model bias problem has not been given adequate research attention in previous works. There are only few common practices to prevent model bias. A naive and common practice is to leverage a large-scale classification dataset with numerous classes, *e.g.*, ImageNet [13], to pretrain the model for better and general prior knowledge to alleviate the model bias toward base classes. Granted that the ImageNet pretraining provides better prior knowledge with the fully-supervised [33] or self-supervised learning [31,35,56], the pertinent model still biases to base classes because of the lack of explicit bias constraints. Fan *et al.* [22] proposes another solution to further alleviate this problem, by leveraging an object detection dataset with numerous training classes to prevent overfitting toward base classes. While this work has significantly improved the generalization performance on novel classes, and such improvement has validated the model bias problem in existing FSOD mothods, this approach requires to establish a large-scale dataset for few-shot object detection, which is expensive and hard to generalize to other high-level few-shot tasks.

Previous works [22,91,93] either have limited performance or require extra annotated dataset. In this paper, we pinpoint and thoroughly investigate the model bias problems in each detection module and present a simple but effective method to calibrate the model from three levels to address the model bias problem: RPN calibration, detector calibration and backbone calibration. Table 1 presents the model bias problem and the bias source for different detection modules, and the performance improvement from our model calibration.

Specifically, *RPN calibration* is designed to calibrate the biased RPN by identifying potential objects of the novel classes and rectify their training labels. We propose Uncertainty-Aware RPN (UA-RPN) to evaluate the uncertainty of each proposal, and exploit such uncertainty to mine novel classes. The *detector calibration* is designed to leverage the feature statistics of both base and novel classes to generate proposal features for unbiased detector training. For the *backbone calibration*, we propose to leverage the overlooked classification dataset, *i.e.*, ImageNet dataset to address the model bias problem, which is freely available but is only used for pretraining in other FSOD works. The backbone is jointly trained on both detection and classification datasets with pseudo box annotations to bridge the domain gap. In summary, our paper has three contributions:

- We identify and thoroughly inspect the model bias problem of each detection module in existing few-shot object detection methods.
- We propose to address the model bias problem from three levels via backbone calibration, RPN calibration and detector calibration. We leverage the overlooked classification dataset to further facilitate our model calibration procedure.
- We verify the effectiveness of our method on two datasets and show that our method achieves very promising performance.

2 Related Works

Object Detection. One-stage detectors [59,75,76,63,62,102] directly predict classes and locations of anchors densely on the extracted backbone features in a single-shot manner. These methods [46,80,109,105,103,90,70,85] usually have fast inference at the expense of detection performance. But the recently proposed anchor-free algorithms [47,66,107,18,99,87,64,45] have significantly boosted the performance with competitive running speed at the same time. Two-stage detectors, which are pioneered and represented by R-CNN methods [26,77,58], first generate candidate proposals likely to contain objects, using traditional techniques [88] or a jointly optimized region proposal network (RPN) [77]. Then these proposals are refined for accurate locations and classified into different classes. These methods [32,83,5,6,12,55,3,81,74,69,82,34] usually have higher detection performance but are slower with its two-stage pipeline. Overall, even these methods perform excellently on multiple object detection datasets, they can only detect objects of training classes and cannot generalize to detecting novel classes.

Few-Shot Learning. Recent few-shot classification methods can be roughly classified into two approaches depending on the prior knowledge learning and adaptation methods. The first approaches few-shot classification as a meta-learning problem by metric-based or optimization-based methods. The metric-based approaches [1,16,36,44,52,53] leverage a siamese network [44] to learn feature embedding of both support and query images and evaluate their relevance using a general distance metric regardless of their categories. The optimized-based approaches [24,4,49,27,48,2,28,79] meta-learn the learning procedures to rapidly update models online with few examples. The second is the recently proposed transfer-learning approach [10,25,15,71] which consists of two separate training stages. These methods first pretrain model on base classes to obtain transferable backbone features at the first stage, and then finetune high-level layers to adapt to novel classes at the second stage. This simple transfer-learning procedure obtains strong performance as validated by multiple recent works [61,86,110]. Our work is inspired by the transfer learning approach, with the idea applied in the few-shot object detection task.

Few-Shot Object Detection. Until now, few-shot learning has achieved impressive progress on multiple important computer vision tasks [65,54,20,21], *e.g.*, semantic segmentation [17,68,38], human motion prediction [29] and object detection [8]. The FSOD methods can be classified into two approaches: meta-learning and transfer-learning methods. The meta-learning methods adopt a siamese network to detect novel classes in query images based on the similarity with given support images. Fan *et al.* [22] proposes attention-rpn and multi-relation detector for better similarity measurement. FR [40] proposes a meta feature learner and a reweighting module to quickly adapt to novel classes. Other methods improve the meta-learning based FSOD methods with different modules, *e.g.*, joint feature embedding [94], support-query mutual guidance [101], dense relation distillation [37] and others [51,95,94]. Transfer-learning methods first train a model on base classes followed by finetuning the model on novel classes to gain good generalization ability. TFA [91] introduces a cosine similarity classifier to detect novel classes. MSPR [93] proposes a multi-scale positive sample refinement module to enrich FSOD object scales. Techniques have been proposed to improve the transfer-learning FSOD methods, *e.g.*, semantic relation reasoning [108], hallucinator network [104], contrastive proposal encoding [84], and others [57,50,23,98,8,41,7,72,30,92]. Our work belongs to the transfer-learning approach. Notably, we identify the model bias problem in existing FSOD models, and propose corresponding model calibration modules to address the bias problems so as to make the trained model generalize better on novel classes. It is also promising to apply our method to the zero-shot task [100].

3 FSOD with Model Calibration

The few-shot object detection (FSOD) task is formally defined as following: given two disjoint classes, base class and novel class, where the base class dataset D_b contains massive training samples for each class, whereas the novel class dataset

Table 2. Model bias problem.

	Stage 1	Stage 2
Training data	Base Classes	Novel Classes
Data Volume	Massive	Few
Biased Module	Backbone & RPN	Detector
Bias Toward	Base classes	Novel samples

D_n has very few (usually no more than 10) annotated instances per class. The base class dataset D_b has been available for model training. The novel class dataset D_n has however been unavailable until now. FSOD targets at detecting all objects of the novel classes for any given input images by transferring knowledge learned from the base class dataset. The performance is measured by average precision (AP) [60] of the novel classes.

Current methods usually take a two-stage training scheme:the base training stage (stage-1) is conducted on the base dataset D_b to extract transferable knowledge. The novel finetuning stage (stage-2) is conducted on the novel dataset D_n for generalization on novel classes. Because D_n contains very few training samples, the backbone weights are frozen, and only the detector and RPN are trained in the novel finetuning stage. This two-stage training scheme results in a model bias problem, as shown in Table 2. In this paper, we propose to calibrate the FSOD model to solve the model bias problem from three levels: backbone calibration, RPN calibration, and detector calibration, as shown in Figure 1.

3.1 Backbone Calibration

For few-shot object detection task, a common practice is to pretrain the backbone on a large-scale *classification* dataset, *e.g.*, ImageNet [13], to obtain good feature initialization for faster training convergence, while simultaneously providing general prior data distribution for good generalization on novel classes. However, the separate, two-stage training of FSOD impedes the generalization gain for novel classes, which is only finetuned in the second training stage. The key reason is that the base classes with massive training samples significantly change the backbone feature distribution in the first training stage. The model is trained to fit the data distribution of the limited base classes (less than 100 classes) at the expense of losing the general distribution learned from massive classes (at least 1000 classes) in the pretraining stage. The backbone biases toward base classes, thus impeding the generalization on novel classes.

Although the base class training stage can destroy the well-learned class distribution from pretraining, it does provide model location supervision by enabling novel *object detection*. In practice, the detection performance on novel classes will significantly degrade if we discard the base class training stage. This paradox motivates us to find a good solution to simultaneously preserve the well-learned class distribution while enabling location supervision. In this paper, we propose to achieve this by providing the backbone with an implicit feature constraint in the base training stage, so as to keep the well-learned data distribution

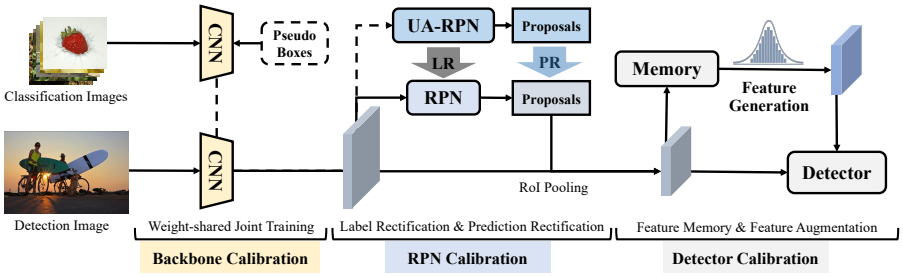

Fig. 1. Overall network architecture. The detection image is fed into the convolutional backbone which is jointly trained with classification images alongside with pseudo boxes. The feature of detection images is then processed by RPN and UA-RPN (uncertainty-aware RPN) to generate proposals. The UA-RPN provides the regular RPN with label rectification (LR) and prediction rectification (PR). The proposal features generated by RoI pooling are fed to the feature memory module to generate features for novel classes. The detector processes both the original and generated proposal features and outputs object detection predictions.

covering massive classes, while enabling the object detection training under the location supervision from base classes.

A naive solution is to jointly train the backbone on both the large-scale classification dataset and object detection dataset to keep both of their data properties. But we find that the classification and detection tasks are not well compatible, where the classification task always dominates the training procedure while significantly degrading the object detection performance. Therefore, we propose to equip the classification images with *pseudo box* annotations to transform the classification dataset into object detection dataset.

The classification images have a desirable property: most images are dominated by the salient objects of the labeled target class. We make use this attribute to generate pseudo boxes for target objects. We utilize a pretrained salient object detection (SOD) model [73] to detect salient objects in classification images. Then we remove SOD mask outliers and only keep the largest contiguous region as the salient region, and generate corresponding pseudo box.

We jointly train the model with a weight-shared backbone on both the detection dataset and classification dataset equipped with pseudo boxes. Specifically, for the detection branch, we keep the training setting as other FSOD methods. For the backbone trained on the classification dataset, we use a smaller input size to fit its original image size. In this way, the model can be trained to perform object detection task while simultaneously keeping the data distribution learned from massive classes of the classification dataset. Note that our backbone calibration is scalable to the number of classes in the classification dataset, where more classes will produce better data distribution and thus better detection performance.

Our backbone calibration method significantly alleviates the backbone bias toward limited detection classes. This is achieved through the distribution calibration from abundantly available classes of the jointly trained classification

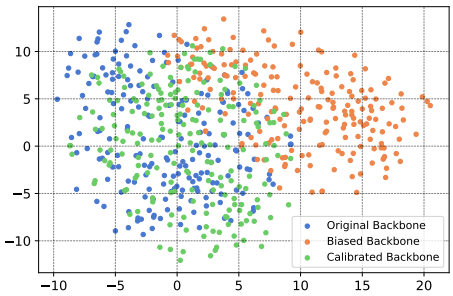

Fig. 2. The t-SNE visualization for feature distributions of original, biased and calibrated backbones.

task. To show the backbone bias problem and the effect of our backbone calibration, we use t-SNE [67] to plot the respective feature distribution of the original, biased and calibrated backbone feature. As shown in Figure 2, for the same input images, the feature distribution of the biased backbone suffer obvious drift from the well-learned distribution of the original backbone. After our model calibration, the backbone feature distribution is well aligned to the original distribution which is closer to the real data distribution.

3.2 RPN Calibration

The region proposal network (RPN) is designed to generate proposals for potential objects, which is trained in a class-agnostic manner. Thus RPN is widely regarded as a general object detection module capable of detecting arbitrary known or unknown classes [39]. However, we find that RPN is general only for the known training classes, with its performance significantly degrading on unknown novel classes for the following two intrinsic reasons.

First, there are no training samples available for novel classes and thus the RPN cannot be trained with the supervision signals of novel classes. Notwithstanding, the RPN generalization ability on novel classes (RPN still can detect some objects of novel classes) is derived from supervision signals of the similar training classes, *e.g.*, RPN trained on horse and sheep can detect zebra, but the performance is not as satisfactory as the training classes. Second, negative supervision signals on novel classes may in fact be present. Note that the training images from RPN probably contain objects of novel classes, but they are not labeled and therefore are regarded as background during training. This inadvertent ignorance can further reduce the RPN generalization ability on novel classes. To make things worse, FSOD models further suffer from the biased RPN problem because of its two-stage training scheme, where the novel classes are totally unlabeled in first base training stage.

We propose to calibrate the biased RPN by rectifying the labels and predictions of potential objects of novel classes. The challenge lies on the object discovery of novel classes. We find an interesting attribute for the RPN proposals of novel classes: their object scores drastically fluctuate across different

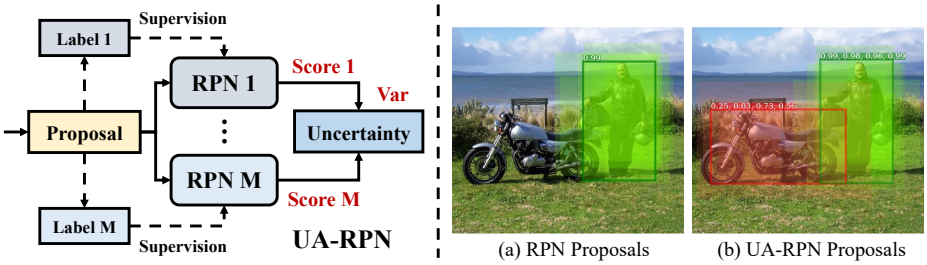

Fig. 3. Uncertainty-aware RPN (UA-RPN) and visualization for proposals of the biased RPN and our UA-RPN. The proposals are visualized with semi-transparent insets, with the most confident proposals highlighted with solid rectangular outlines. The base and novel classes are represented by green and red proposals respectively.

images even for objects of the same novel class, while the base classes almost *always* have stable high object scores. This means proposals of novel classes has higher prediction uncertainty compared to pure background and base classes. On the one hand, the RPN is trained to identify class-agnostic objects irrespective of classes. On the other hand, some objects of novel classes are regarded as background during training. This observation motivates us to leverage the prediction *uncertainty* to identify objects of novel classes.

We propose uncertainty-aware RPN (**UA-RPN**) to rectify the RPN training labels and predictions for RPN calibration. Our method is inspired by a widely used uncertainty estimation method M-head [78], which models deterministic features with a shared backbone, as well as *stochastic* features with multiple different predictors, where the multiple predicted hypotheses are utilized to represent the uncertainty. Our UA-RPN extends the M-head idea to estimate the uncertainty of proposals, as shown in Figure 3. Specifically, we first build M separate RPN heads f_m with the same architecture. Each RPN shares the same input backbone feature x and generates one object score prediction for each proposal. Therefore, we can obtain M object scores for each proposal.

$$f(x) = (f_1(x), f_2(x), ..., f_M(x)) \tag{1}$$

Then we compute the variance Var

$$Var(f(x)) = \frac{\sum_{m=1}^{M}(f_m(x) - \mu)^2}{M - 1} \tag{2}$$

of these multiple predictions for each proposal to represent the prediction uncertainty, where μ is the mean of $f(x)$.

With the aforementioned M prediction heads, we want their predictions to be different for the samples of novel classes, so that the prediction variance can be used to evaluate uncertainty. Thus, we propose the following two designs to make UA-RPN more uncertainty-aware.

The first is training UA-RPN with diverse label assignments, where different RPN heads have different supervision labels for the same proposal. Because we

target at discovering novel classes from cluttered background, we only change the labels of background proposals, where some background proposals are selectively ignored under different label assignments. For the regular RPN, the proposals are labeled as background when the overlap OV with the groundtruth bounding boxes satisfies $OV \in [0.0, 0.3]$ and those proposals with $OV \in (0.3, 0.7)$ are ignored during training. For our UA-RPN, we set different OVs for the background label assignment of different RPN heads to improve the supervision ambiguity and diversity. Those original background proposals which are not in accordance to the label assignment are ignored during training.

The second design is to detach UA-RPN from the backbone gradients so that the backbone and UA-RPN are separately optimized. This separate optimization not only avoids the backbone feature from the ambiguity supervision of UA-RPN, but also ensures the independent training for different heads of UA-RPN.

During training, those background proposals with large uncertainty $Var > \alpha$ are regarded as potential objects of novel classes, for which we rectify their labels to foreground. During testing, the predicted object scores of UA-RPN are used to rectify the RPN predictions through an average operation. The above rectifications are performed in a module ensemble manner with only a few extra computation, thanks to the light-weight design of UA-RPN.

We analyze proposal scores of both base and novel classes to illustrate the RPN bias problem and our RPN calibration. As shown in Figure 3, the regular RPN fails to detect the novel class caused by its training bias, while our UA-RPN successfully detects the object of this novel "motorbike". Note that the proposal of novel classes has diverse object scores from multiple heads of UA-RPN and therefore their uncertainty (variance) is very high, while the base class proposal has consistently high object scores with low uncertainty.

3.3 Detector Calibration

The biased detector is mainly caused by the limited training samples in the novel finetuning stage. With numerous labeled samples in the base training stage to represent the real data distribution of base classes, the detector can be optimized to the unbiased base classes and generalize well in inference. However, in the novel finetuning stage, there are only a few training samples for each novel class, which is insufficient to represent the real data distribution of novel classes. Thus the model can easily overfit to the biased distribution underrepresented by these few samples.

We propose to calibrate the biased detector by providing rectified proposal features with the help of base classes. This idea is motivated by the fact that different classes share some commonalities, *e.g.*, horse and cow share similar legs and bodies. We can accurately estimate the data distribution of base classes under the supervision of numerous training samples. Then we can retrieve the similar base classes for novel classes, and leverage the accurate distribution estimation to rectify the data distribution. With the rectified distribution estimation, we can generate unbiased proposal features to train the detector for detector calibration. This idea has been validated in other few-shot learning tasks [96,97].

We assume the class feature distribution is Gaussian. Then we leverage a feature memory module to accumulate the feature prototype statistics of both base and novel classes. The global prototype of class c is denoted as P_c. For each image i during training, we select all proposal features F_c of the target class and accumulate them into the global prototype: $P_c \leftarrow mP_c + (1-m)F_c^i$, where m is the update momentum and set as 0.999 in our experiments. In this way, the class prototype is iteratively updated during training in an exponential moving average manner. The feature memory encodes various objects in the dataset for each class, and thus the representative prototype can capture transferable commonalities shared among base and novel classes with higher chance.

The class prototypes effectively represent the feature statistics of each class under the Gaussian distribution assumption. For a novel class n, we compute the cosine similarity S between its prototype P_n and all base class prototypes P_b. Then we select the prototype of the most similar base class P_{bs}, where $bs = \text{argmax}(S)$. Then we calibrate the novel class prototype as $\hat{P_n} = \gamma P_n + (1-\gamma)P_{bs}$, where γ is the adjustable weight and we adopt $\gamma = 0.5$ in our experiments. Then we utilize a Gaussian distribution $F_G \sim N(\hat{P_n}, 1)$ to generate features F_G for this novel class.

We perform detector calibration in the novel finetuning stage, where we fix the backbone and only finetune the detector and RPN. We use the generated feature F_G with the original proposal feature F_O to jointly train the detector. There are at most 32 generated features for each foreground class.

4 Experiments

In this section, we conduct extensive experiments to validate the model bias problem in current FSOD models, and demonstrates the effectiveness of our proposed model calibration modules.

4.1 Experimental Settings

Datasets. Our experiments are conducted on MS COCO [60] and Pascal VOC [19] datasets. MS COCO dataset contains 80 classes and we follow previous works [22,91,40] to split them into two separate sets, where the 20 classes overlapping with Pascal VOC dataset are treated as novel classes, with the remaining 60 classes regarded as base classes. We utilize the 5K images of the val2017 set for evaluation, and train2017 set with around 80K images for training. We use the support samples of novel classes in FSOD [22] to finetune our model on MS COCO dataset. As for Pascal VOC dataset which contains 20 classes, random split is done to produce 5 novel classes and 15 base classes. Specifically, there are three random class split groups, and we follow the split setting of previous works [84,91] for a fair comparison. The VOC 2007 and VOC 2012 trainval sets are utilized for training, and VOC 2007 test set are used for evaluation.

Training and Evaluation. The model is first trained on base classes and then finetuned on novel classes. Specifically, the instance number K of novel

Table 3. Ablation studies on different calibration (Cal.) cooperations.

RPN	Detector	Backbone	AP	AP_{50}	AP_{75}
			11.3	20.9	11.0
Cal			11.8	21.5	11.8
	Cal.		12.0	22.1	11.9
		Cal.	14.7	25.8	14.6
Cal.	Cal.		12.2	23.0	11.6
Cal.	Cal.	Cal.	15.1	27.2	14.6

classes for finetuning is $K = 5, 10$ for MS COCO dataset, and $K = 1, 2, 3, 5, 10$ for Pascal VOC dataset. The model is evaluated multiple times on novel classes with average precision (AP) as the evaluation metric. We report the COCO-style mAP on COCO dataset and AP_{50} on Pascal VOC dataset, which are the common practice to fit the dataset characteristics.

Model Details. We adopt the pretrained U^2-Net [73] salient object detection network to generate pseudo boxes. We reuse ImageNet [13] dataset for backbone calibration, which is only used for model pretraining in other works. UA-RPN has $M = 4$ RPNs with different $OV \in [0.01, 0.3], [0.1, 0.3], [0.15, 0.3], [0.2, 0.3]$[1]. We dynamically set α by selecting top-1,000 uncertain proposals.

Implementation Details. We adopt Faster R-CNN [77] with Feature Pyramid Network [58] (FPN) as our basic detection framework, with the ResNet-50 [33] backbone pretrained on ImageNet [13] dataset. We use SGD to optimize our model with weight decay of $5e^{-5}$ and momentum of 0.9. The model is trained 50,000 iterations at the base training stage. The learning rate is set as 0.02 in the first 30,000 iterations, which decays by $10\times$ for every 10,000 iterations. For the novel class finetuning stage, the model is trained 3,000 iterations with 0.01 initial learning rate, which decays by $10\times$ upon reaching the 2,000-th iteration. The object detection images are resized with the fixed height/width ratio, where the shorter image side is resized to 600 while the longer side is capped at 1,000.

4.2 Ablation Studies

Table 3 shows that each model calibration module improves the detection performance, and their combination promotes the overall performance from 11.3 to 15.1 AP. We further conduct extensive experiments on MS COCO dataset to investigate the efficacy of our model calibration on handling model bias in different modules.

Backbone calibration The backbone bias is introduced at the base training stage, where the generalized feature distribution learned from massive classes can be scrapped. This problem is effectively relieved by our backbone calibration

[1] Proposals with $OV \in \{[0, 0.01), [0, 0.1), [0, 0.15), [0, 0.2)\}$ are respectively ignored during training for each UA-RPN head.

Table 4. Ablation studies on backbone calibrations. "CAM" denotes class activation map, PB denotes pseudo boxes, and † means with careful hyperparameter tuning.

Backbone	AP	AP_{50}	AP_{75}
Baseline	11.3	20.9	11.0
Cal. Backbone	14.7	25.8	14.6
Cal. w/ 725 Cls	14.2	25.4	13.9
Cal. w/ Rand. 500 Cls	13.9	24.7	13.8
Cal. w/ Rand. 300 Cls	13.3	25.3	12.7
Cal. w/ Rand. 100 Cls	12.3	22.2	12.1
Cal. w/ CAM	14.4	27.5	13.3
Cal. w/o PB	7.5	16.2	5.8
Cal. w/o PB†	14.2	27.1	13.2

module, where the generalization performance on novel classes is significantly improved by 3.4 AP. (Table 4)

Discussions. Concerns about backbone calibration include:

Does the performance gain come from the overlapped classes in the ImageNet dataset? The performance gain mainly comes from the well-learned feature distribution covering the massive classes. To remove any effect of the overlapped classes, we use a purified ImageNet [13] dataset to calibrate the backbone, which contains 725 classes by removing all classes similar to the novel classes in MS COCO. Compared to the backbone calibrated on the full ImageNet dataset, the performance slightly degrades by 0.5 AP. With 500 randomly selected classes, the performance with model calibration only degrades by 0.8 AP. These results validate that the improvement of backbone calibration is mainly derived from the massive classes, rather than from the overlapped classes. We also present the performance with 100 and 300 randomly selected classes to further show the impact of the class diversity on backbone calibration.

Does the performance gain come from the pretrained SOD model? To address this concern, we utilize class activated map [106] (CAM) to generate pseudo boxes, which is an unsupervised method and the pseudo mask can be directly generated from the pretrained classification model. The performance only degrades by 0.3 AP with the inaccurate CAM generated pseudo boxes. We also directly and jointly train the model on both detection and classification datasets without pseudo boxes. The performance dramatically degrades to 7.5 AP because of the dominating classification branch. But with careful hyperparameter tuning by reducing the loss weights from the classification branch, the performance can reach 14.2 AP. These results validates that the backbone calibration can be *only* slightly affected by the pseudo box quality.

Does the backbone calibration introduce extra data? Our method does not introduce any extra data. We only reuse ImageNet dataset, which is a free resource, to perform backbone calibration. Other FSOD methods however only use ImageNet for model pretraining.

Table 5. Ablation studies for UA-RPN. "DT" denotes detaching UA-RPN gradient from backbone, "LA" denotes label assignment for background proposals, "LR" denotes label rectification and "PR" denotes prediction rectification.

Method	Stage1 AR @100	Stage1 AR @1000	Stage2 AR @100	Stage2 AR @1000	AP
RPN	17.1	35.3	26.1	38.6	11.3
Cal. RPN	20.1	37.4	28.5	39.6	11.8
Cal. w/o DT	19.2	37.0	26.9	39.2	10.2
Cal. w/o LA	17.5	35.5	26.5	38.7	11.5
Cal. w/o LR	19.1	37.1	27.9	39.0	11.6
Cal. w/o PR	18.5	36.7	27.0	39.4	11.4

Table 6. Experimental results on Pascal VOC dataset. The **best** and second best results are highlighted with **bold** and underline, respectively.

Method	Novel Set 1 1	2	3	5	10	Novel Set 2 1	2	3	5	10	Novel Set 3 1	2	3	5	10
LSTD [9]	8.2	1.0	12.4	29.1	38.5	11.4	3.8	5.0	15.7	31.0	12.6	8.5	15.0	27.3	36.3
FSRW [40]	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
MetaRCNN [95]	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
FsDetView [94]	24.2	35.3	42.2	49.1	57.4	21.6	24.6	31.9	37.0	45.7	21.2	30.0	37.2	43.8	49.6
MSPR [93]	41.7	–	**51.4**	55.2	61.8	24.4	–	39.2	39.9	47.8	35.6	–	42.3	48.0	49.7
TFA [91]	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
FSCE [84]	44.2	43.8	**51.4**	61.9	**63.4**	27.3	29.5	**43.5**	44.2	50.2	37.2	41.9	**47.5**	**54.6**	**58.5**
DCNet [37]	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
SRR-FSD [108]	**47.8**	**50.5**	51.3	55.2	56.8	32.5	**35.3**	39.1	40.8	43.8	**40.1**	41.5	44.3	46.9	46.4
Ours	40.1	44.2	51.2	**62.0**	63.0	**33.3**	33.1	42.3	**46.3**	**52.3**	36.1	**43.1**	43.5	52.0	56.0

RPN calibration Table 5 validates the effectiveness of our proposed UA-RPN, where the baseline RPN only has 17.1/35.3 AR@100/1000 on novel classes in the base training stage (stage 1), while the performance on base classes can reach 42.1/49.8 AR@100/1000. The performance of the former on novel classes is only improved to 26.1/38.6 AR@100/1000 after the novel finetuning stage. These results indicate the serious RPN bias toward the base classes, which adversely affects the generalization ability on the novel classes. With our RPN calibration, the recall of the novel classes on both stage 1 and stage 2 is significantly improved, and the overall detection AP performance is also improved by 0.5 AP. We further validate the effectiveness of separate modules in UA-RPN. The gradient detaching is essential for keeping the detection performance by separating UA-RPN gradients from backbone. The diverse label assignment affects the uncertainty of UA-RPN and therefore is essential for the recall of novel classes. Both the label and prediction rectifications are beneficial to the proposal recall and detection performance of novel classes.

Detector calibration We evaluate the models on the training samples of novel classes to demonstrate the detector bias. The detection performance on the training samples reaches 61.7 AP, while the generalization performance on

Table 7. Experimental 5-shot results on MS COCO dataset.

Backbone	Publication	AP	AP_{50}	AP_{75}
FSRW [40]	ICCV'19	5.6	12.3	4.6
MetaRCNN [95]	ICCV'19	8.7	19.1	6.6
FSOD [22]	CVPR'20	11.1	20.4	10.6
MSPR [93]	ECCV'20	9.8	17.9	9.7
FsDetView [94]	ECCV'20	12.5	27.3	9.8
TFA [91]	ICML'20	10.0	–	9.3
SRR-FSD [108]	CVPR'21	11.3	23.0	9.8
FSCE [84]	CVPR'21	11.9	-	10.5
DCNet [37]	CVPR'21	12.8	23.4	11.2
Ours	-	**15.1**	**27.2**	**14.6**

testing samples is only 11.3 AP. The large performance gap between training and testing samples indicates the serious detector bias toward the training samples. Equipped with our detector calibration, the performance on testing samples can be improved from 11.3 to 11.9 AP, with the performance gap also reduced by relieving the overfitting on training samples.

4.3 Comparison with SOTAs

We conduct comparison experiments with state-of-the-art methods on Pascal VOC and MS COCO datasets. Pascal VOC contains more median-sized and large objects and thus the detection performance is much higher than that on MS COCO dataset. As shown in Table 6, out method performs better or comparable to other methods in all class splits. MS COCO is a challenging dataset even for fully-supervised methods. As shown in Table 7, with the proposed model calibration, our model significantly outperforms other methods by a large margin of 2.5 AP, with the detection performance reaching 15.1 AP.

5 Conclusion

Few-shot object detection (FSOD) has recently achieved remarkable progress. However, previous FSOD works have ignored the intrinsic model bias problem in transfer learning. The model bias problem causes overfitting toward training classes while destructing the well-learned transferable knowledge. In this paper, we identify and perform a comprehensive study on the model bias problem in FSOD, and propose a simple yet effective method to address the problem, making use of the ImageNet dataset not limited to pre-training as done in other works. Specifically, we perform model calibrations in three levels: 1) *backbone calibration* to preserve the well-learned prior knowledge which relieves the model from bias towards base classes; 2) *RPN calibration* to rescue unlabeled objects of novel classes; and 3) *detector calibration* to prevent model bias towards a small number of training samples of the novel classes. Extensive experiments and analysis substantiate the effectiveness of our model calibration method.

References

1. Allen, K., Shelhamer, E., Shin, H., Tenenbaum, J.: Infinite mixture prototypes for few-shot learning. In: ICML (2019) 4
2. Antoniou, A., Edwards, H., Storkey, A.: How to train your maml. In: ICLR (2019) 4
3. Bell, S., Lawrence Zitnick, C., Bala, K., Girshick, R.: Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. In: CVPR (2016) 3
4. Bertinetto, L., Henriques, J.F., Torr, P.H., Vedaldi, A.: Meta-learning with differentiable closed-form solvers. In: ICLR (2019) 4
5. Cai, Z., Fan, Q., Feris, R.S., Vasconcelos, N.: A unified multi-scale deep convolutional neural network for fast object detection. In: ECCV (2016) 3
6. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: CVPR (2018) 3
7. Cao, Y., Wang, J., Jin, Y., Wu, T., Chen, K., Liu, Z., Lin, D.: Few-shot object detection via association and discrimination. NeurIPS (2021) 4
8. Chen, H., Wang, Y., Wang, G., Qiao, Y.: Lstd: A low-shot transfer detector for object detection. In: AAAI (2018) 1, 4
9. Chen, H., Wang, Y., Wang, G., Qiao, Y.: Lstd: A low-shot transfer detector for object detection. In: AAAI (2018) 13
10. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: ICLR (2019) 4
11. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: CVPR (2019) 2
12. Dai, J., Li, Y., He, K., Sun, J.: R-fcn: Object detection via region-based fully convolutional networks. In: NeurIPS (2016) 3
13. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR (2009) 2, 5, 11, 12
14. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: CVPR (2019) 2
15. Dhillon, G.S., Chaudhari, P., Ravichandran, A., Soatto, S.: A baseline for few-shot image classification. In: ICLR (2019) 4
16. Doersch, C., Gupta, A., Zisserman, A.: Crosstransformers: spatially-aware few-shot transfer. In: NeurIPS (2020) 4
17. Dong, N., Xing, E.P.: Few-shot semantic segmentation with prototype learning. In: BMVC (2018) 4
18. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q.: Centernet: Keypoint triplets for object detection. In: ICCV (2019) 3
19. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. IJCV (2010) 10
20. Fan, Q., Ke, L., Pei, W., Tang, C.K., Tai, Y.W.: Commonality-parsing network across shape and appearance for partially supervised instance segmentation. In: ECCV (2020) 4
21. Fan, Q., Tang, C.K., Tai, Y.W.: Few-shot video object detection. arXiv preprint arXiv:2104.14805 (2021) 4
22. Fan, Q., Zhuo, W., Tang, C.K., Tai, Y.W.: Few-shot object detection with attention-rpn and multi-relation detector. In: CVPR (2020) 1, 2, 3, 4, 10, 14
23. Fan, Z., Ma, Y., Li, Z., Sun, J.: Generalized few-shot object detection without forgetting. In: CVPR (2021) 4

24. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML (2017) 4
25. Gidaris, S., Komodakis, N.: Dynamic few-shot visual learning without forgetting. In: CVPR (2018) 4
26. Girshick, R.: Fast r-cnn. In: ICCV (2015) 1, 3
27. Gordon, J., Bronskill, J., Bauer, M., Nowozin, S., Turner, R.: Meta-learning probabilistic inference for prediction. In: ICLR (2019) 4
28. Grant, E., Finn, C., Levine, S., Darrell, T., Griffiths, T.: Recasting gradient-based meta-learning as hierarchical bayes. In: ICLR (2018) 4
29. Gui, L.Y., Wang, Y.X., Ramanan, D., Moura, J.M.F.: Few-shot human motion prediction via meta-learning. In: ECCV (2018) 4
30. Han, G., He, Y., Huang, S., Ma, J., Chang, S.F.: Query adaptive few-shot object detection with heterogeneous graph convolutional networks. In: ICCV (2021) 4
31. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: CVPR (2020) 2
32. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: ICCV (2017) 3
33. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016) 2, 11
34. He, Y., Zhu, C., Wang, J., Savvides, M., Zhang, X.: Bounding box regression with uncertainty for accurate object detection. In: CVPR (2019) 3
35. Hénaff, O.J., Koppula, S., Alayrac, J.B., Van den Oord, A., Vinyals, O., Carreira, J.: Efficient visual pretraining with contrastive detection. In: ICCV (2021) 2
36. Hou, R., Chang, H., Ma, B., Shan, S., Chen, X.: Cross attention network for few-shot classification. In: NeurIPS (2019) 4
37. Hu, H., Bai, S., Li, A., Cui, J., Wang, L.: Dense relation distillation with context-aware aggregation for few-shot object detection. In: CVPR (2021) 4, 13, 14
38. Hu, T., Pengwan, Zhang, C., Yu, G., Mu, Y., Snoek, C.G.M.: Attention-based multi-context guiding for few-shot semantic segmentation. In: AAAI (2019) 4
39. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: CVPR (2021) 7
40. Kang, B., Liu, Z., Wang, X., Yu, F., Feng, J., Darrell, T.: Few-shot object detection via feature reweighting. In: ICCV (2019) 2, 4, 10, 13, 14
41. Karlinsky, L., Shtok, J., Harary, S., Schwartz, E., Aides, A., Feris, R., Giryes, R., Bronstein, A.M.: Repmet: Representative-based metric learning for classification and few-shot object detection. In: CVPR (2019) 1, 4
42. Khan, S., Hayat, M., Zamir, S.W., Shen, J., Shao, L.: Striking the right balance with uncertainty. In: CVPR (2019) 2
43. Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R.: Cost-sensitive learning of deep feature representations from imbalanced data. IEEE TNNLS (2017) 2
44. Koch, G., Zemel, R., Salakhutdinov, R.: Siamese neural networks for one-shot image recognition. In: ICML Workshop (2015) 4
45. Kong, T., Sun, F., Liu, H., Jiang, Y., Li, L., Shi, J.: Foveabox: Beyound anchor-based object detection. IEEE TIP (2020) 3
46. Kong, T., Sun, F., Yao, A., Liu, H., Lu, M., Chen, Y.: Ron: Reverse connection with objectness prior networks for object detection. In: CVPR (2017) 3
47. Law, H., Deng, J.: Cornernet: Detecting objects as paired keypoints. In: ECCV (2018) 3
48. Lee, K., Maji, S., Ravichandran, A., Soatto, S.: Meta-learning with differentiable convex optimization. In: CVPR (2019) 4

49. Lee, Y., Choi, S.: Gradient-based meta-learning with learned layerwise metric and subspace. In: ICML (2018) 4
50. Li, A., Li, Z.: Transformation invariant few-shot object detection. In: CVPR (2021) 4
51. Li, B., Yang, B., Liu, C., Liu, F., Ji, R., Ye, Q.: Beyond max-margin: Class margin equilibrium for few-shot object detection. In: CVPR (2021) 4
52. Li, H., Eigen, D., Dodge, S., Zeiler, M., Wang, X.: Finding task-relevant features for few-shot learning by category traversal. In: CVPR (2019) 4
53. Li, W., Wang, L., Xu, J., Huo, J., Gao, Y., Luo, J.: Revisiting local descriptor based image-to-class measure for few-shot learning. In: CVPR (2019) 4
54. Li, X., Wei, T., Chen, Y.P., Tai, Y.W., Tang, C.K.: Fss-1000: A 1000-class dataset for few-shot segmentation. In: CVPR (2020) 4
55. Li, Y., Chen, Y., Wang, N., Zhang, Z.: Scale-aware trident networks for object detection. In: ICCV (2019) 3
56. Li, Y., Xie, S., Chen, X., Dollar, P., He, K., Girshick, R.: Benchmarking detection transfer learning with vision transformers. arXiv preprint arXiv:2111.11429 (2021) 2
57. Li, Y., Zhu, H., Cheng, Y., Wang, W., Teo, C.S., Xiang, C., Vadakkepat, P., Lee, T.H.: Few-shot object detection via classification refinement and distractor retreatment. In: CVPR (2021) 4
58. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: CVPR (2017) 1, 3, 11
59. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017) 1, 3
60. 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) 2, 5, 10
61. Liu, B., Cao, Y., Lin, Y., Li, Q., Zhang, Z., Long, M., Hu, H.: Negative margin matters: Understanding margin in few-shot classification. In: ECCV (2020) 4
62. Liu, S., Huang, D., et al.: Receptive field block net for accurate and fast object detection. In: ECCV (2018) 3
63. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: ECCV (2016) 1, 3
64. Liu, W., Liao, S., Ren, W., Hu, W., Yu, Y.: High-level semantic feature detection: A new perspective for pedestrian detection. In: CVPR (2019) 3
65. Liu, Y., Zhang, X., Zhang, S., He, X.: Part-aware prototype network for few-shot semantic segmentation. In: ECCV (2020) 4
66. Lu, X., Li, B., Yue, Y., Li, Q., Yan, J.: Grid r-cnn. In: CVPR (2019) 3
67. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research (2008) 7
68. Michaelis, C., Bethge, M., Ecker, A.S.: One-shot segmentation in clutter. In: ICML (2018) 4
69. Najibi, M., Rastegari, M., Davis, L.S.: G-cnn: an iterative grid based object detector. In: CVPR (2016) 3
70. Nie, J., Anwer, R.M., Cholakkal, H., Khan, F.S., Pang, Y., Shao, L.: Enriched feature guided refinement network for object detection. In: ICCV (2019) 3
71. Qi, H., Brown, M., Lowe, D.G.: Low-shot learning with imprinted weights. In: CVPR (2018) 4
72. Qiao, L., Zhao, Y., Li, Z., Qiu, X., Wu, J., Zhang, C.: Defrcn: Decoupled faster r-cnn for few-shot object detection. In: ICCV (2021) 4

73. Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M.: U2-net: Going deeper with nested u-structure for salient object detection. PR (2020) 6, 11
74. Qin, Z., Li, Z., Zhang, Z., Bao, Y., Yu, G., Peng, Y., Sun, J.: Thundernet: Towards real-time generic object detection on mobile devices. In: ICCV (2019) 3
75. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR (2016) 1, 3
76. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: CVPR (2017) 1, 3
77. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: NeurIPS (2015) 1, 2, 3, 11
78. Rupprecht, C., Laina, I., DiPietro, R., Baust, M., Tombari, F., Navab, N., Hager, G.D.: Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In: ICCV (2017) 8
79. Rusu, A.A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., Hadsell, R.: Meta-learning with latent embedding optimization. In: ICLR (2019) 4
80. Shen, Z., Liu, Z., Li, J., Jiang, Y.G., Chen, Y., Xue, X.: Dsod: Learning deeply supervised object detectors from scratch. In: ICCV (2017) 3
81. Shrivastava, A., Gupta, A.: Contextual priming and feedback for faster r-cnn. In: ECCV (2016) 3
82. Shrivastava, A., Gupta, A., Girshick, R.: Training region-based object detectors with online hard example mining. In: CVPR (2016) 3
83. Singh, B., Najibi, M., Davis, L.S.: Sniper: Efficient multi-scale training. In: NeurIPS (2018) 3
84. Sun, B., Li, B., Cai, S., Yuan, Y., Zhang, C.: Fsce: Few-shot object detection via contrastive proposal encoding. In: CVPR (2021) 4, 10, 13, 14
85. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: CVPR (2020) 3
86. Tian, Y., Wang, Y., Krishnan, D., Tenenbaum, J.B., Isola, P.: Rethinking few-shot image classification: a good embedding is all you need? In: ECCV (2020) 4
87. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: ICCV (2019) 3
88. Uijlings, J.R., Van De Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. IJCV (2013) 3
89. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., Liu, W.: Cosface: Large margin cosine loss for deep face recognition. In: CVPR (2018) 2
90. Wang, T., Anwer, R.M., Cholakkal, H., Khan, F.S., Pang, Y., Shao, L.: Learning rich features at high-speed for single-shot object detection. In: ICCV (2019) 3
91. Wang, X., Huang, T.E., Darrell, T., Gonzalez, J.E., Yu, F.: Frustratingly simple few-shot object detection. In: ICML (2020) 2, 3, 4, 10, 13, 14
92. Wu, A., Han, Y., Zhu, L., Yang, Y.: Universal-prototype enhancing for few-shot object detection. In: ICCV (2021) 4
93. Wu, J., Liu, S., Huang, D., Wang, Y.: Multi-scale positive sample refinement for few-shot object detection. In: ECCV (2020) 2, 3, 4, 13, 14
94. Xiao, Y., Marlet, R.: Few-shot object detection and viewpoint estimation for objects in the wild. In: ECCV (2020) 2, 4, 13, 14
95. Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., Lin, L.: Meta r-cnn : Towards general solver for instance-level low-shot learning. In: ICCV (2019) 2, 4, 13, 14
96. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: Mining latent classes for few-shot segmentation. In: ICCV (2021) 9

97. Yang, S., Liu, L., Xu, M.: Free lunch for few-shot learning: Distribution calibration. In: ICLR (2021) 9
98. Yang, Y., Wei, F., Shi, M., Li, G.: Restoring negative information in few-shot object detection. In: NeurIPS (2020) 4
99. Yang, Z., Liu, S., Hu, H., Wang, L., Lin, S.: Reppoints: Point set representation for object detection. In: ICCV (2019) 3
100. Zareian, A., Rosa, K.D., Hu, D.H., Chang, S.F.: Open-vocabulary object detection using captions. In: CVPR (2021) 4
101. Zhang, L., Zhou, S., Guan, J., Zhang, J.: Accurate few-shot object detection with support-query mutual guidance and hybrid loss. In: CVPR (2021) 4
102. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: CVPR (2020) 3
103. Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z.: Single-shot refinement neural network for object detection. In: CVPR (2018) 3
104. Zhang, W., Wang, Y.X.: Hallucination improves few-shot object detection. In: CVPR (2021) 4
105. Zhang, Z., Qiao, S., Xie, C., Shen, W., Wang, B., Yuille, A.L.: Single-shot object detection with enriched semantics. In: CVPR (2018) 3
106. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR (2016) 12
107. Zhou, X., Zhuo, J., Krahenbuhl, P.: Bottom-up object detection by grouping extreme and center points. In: CVPR (2019) 3
108. Zhu, C., Chen, F., Ahmed, U., Savvides, M.: Semantic relation reasoning for shot-stable few-shot object detection. In: CVPR (2021) 2, 4, 13, 14
109. Zhu, R., Zhang, S., Wang, X., Wen, L., Shi, H., Bo, L., Mei, T.: Scratchdet: Training single-shot object detectors from scratch. In: CVPR (2019) 3
110. Ziko, I., Dolz, J., Granger, E., Ayed, I.B.: Laplacian regularized few-shot learning. In: ICML (2020) 4