Meta-tuning Loss Functions and Data Augmentation for Few-shot Object Detection

Berkan Demirel[1,2] Orhun Buğra Baran[1] Ramazan Gokberk Cinbis[1]
[1]Middle East Technical University [2]HAVELSAN Inc.
bdemirel@havelsan.com.tr bugra@ceng.metu.edu.tr gcinbis@ceng.metu.edu.tr

Abstract

Few-shot object detection, the problem of modelling novel object detection categories with few training instances, is an emerging topic in the area of few-shot learning and object detection. Contemporary techniques can be divided into two groups: fine-tuning based and meta-learning based approaches. While meta-learning approaches aim to learn dedicated meta-models for mapping samples to novel class models, fine-tuning approaches tackle few-shot detection in a simpler manner, by adapting the detection model to novel classes through gradient based optimization. Despite their simplicity, fine-tuning based approaches typically yield competitive detection results. Based on this observation, we focus on the role of loss functions and augmentations as the force driving the fine-tuning process, and propose to tune their dynamics through meta-learning principles. The proposed training scheme, therefore, allows learning inductive biases that can boost few-shot detection, while keeping the advantages of fine-tuning based approaches. In addition, the proposed approach yields interpretable loss functions, as opposed to highly parametric and complex few-shot meta-models. The experimental results highlight the merits of the proposed scheme, with significant improvements over the strong fine-tuning based few-shot detection baselines on benchmark Pascal VOC and MS-COCO datasets, in terms of both standard and generalized few-shot performance metrics.

1. Introduction

Object detection is one of the computer vision problems that has greatly benefited from the advances in supervised deep learning approaches. However, similar to the case in many other problems, state-of-the-art in object detection relies on the availability of large-scale fully-annotated datasets, which is particularly problematic due to the difficulty of collecting accurate bounding box annotations [18, 46]. This practical burden has lead to a great interest in the approaches

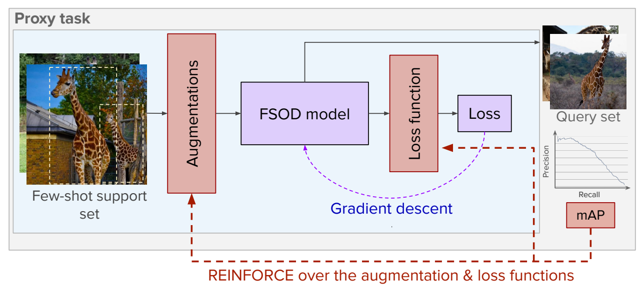


Figure 1. The overall architecture of the meta-tuning approach.

that can potentially reduce the annotation cost, such as weakly-supervised learning [29, 57], learning from point annotations [7], and mixed supervised learning [45]. A more recently emerging paradigm in this direction is *few-shot object detection* (FSOD). In the FSOD problem, the goal is to build detection models for the *novel* classes with few labeled training images by transferring knowledge from the *base* classes with a large set of training images. In the closely related Generalized-FSOD (G-FSOD) problem, the goal is to build few-shot detection models that perform well on both base and novel classes.

FSOD methods can be categorized into meta-learning and fine-tuning approaches. Although meta-learning based methods are predominantly used in the literature in FSOD research [8,22,31,36,52,75,76,79,81,83], several fine-tuning based works have recently reported competitive results [6, 15, 32, 53, 61, 65, 72, 84]. The main premise of meta-learning approaches is to design and train dedicated meta-models that map given few train samples to novel class detection models, *e.g.* [73] or learn easy-to-adapt models [30] in a MAML [16] fashion. In contrast, however, fine-tuning based methods tackle the problem as a typical transfer learning problem and apply the general purpose supervised training techniques, *i.e.* regularized loss minimization via gradient-based optimization, to adapt a pre-trained model to few-shot classes. It is also worth noting that the recent results on fine-tuning based FSOD are aligned with related observations on few-shot classification [9, 12, 63] and segmentation [4].

While some of the FSOD meta-learning approaches are attractive for being able to learn dedicated parametric training mechanisms, they also come with two important shortcomings: (i) the risk of overfitting to the base classes used for training the meta-model due to model complexity, and (ii) the difficulty of interpreting what is actually learned; both of which can be crucially important for real-world, in-the-wild utilization of a meta-learned model. From this point of view, the simplicity and generality of a fine-tuning based FSOD approach can be seen as major advantages. In fact, one can find a large machine learning literature on the components (optimization techniques, loss functions, data augmentation, and architectures) of an FT approach, as opposed to the unique and typically unknown nature of a meta-learned inference model, especially when the model aims to replace standard training procedures for modeling the novel few-shot classes. While MAML [16] like meta-learning for quick adaptation is closer in nature to fine-tuning based approaches, the vanishing gradient problems and the overall complexity of the meta-learning task practically limits the approach to target only one or few model update steps, whereas an FT approach has no such computational difficulty.

Perhaps the biggest advantage of a fine-tuning based FSOD approach, however, can also be its biggest disadvantage: its generality may lack the inductive biases needed for effective learning with few novel class samples while preserving the knowledge of base classes. To this end, such approaches focus on the design of fine-tuning details, *e.g.* whether to freeze the representation parameters [65], use contrastive fine-tuning losses [61], increase the novel class variances [84], introduce the using additional detection heads and branches [15, 72]. However, optimizing such details specifically for few-shot classes in a hand-crafted manner is clearly difficult, and likely to be sub-optimal.

To address this problem, we focus on applying meta-learning principles to tune the loss functions and augmentations to be used in the fine-tuning stage for FSOD, which we call *meta-tuning* (Figure 1). More specifically, much like the meta-learning of a meta-model, we define an episodic training procedure that aims to progressively discover the optimal loss function and augmentation details for FSOD purposes in a data-driven manner. Using reinforcement learning (RL) techniques, we aim to tune the loss function and augmentation details such that they maximize the expected detection quality of an FSOD model obtained by fine-tuning to a set of novel classes. By defining meta-tuning over well-designed loss terms and an augmentation list, we restrict the search process to effective function families, reducing the computational costs compared to AutoML methods that aim to discover loss terms from scratch for fully-supervised learning [20, 42]. The resulting meta-tuned loss functions and augmentations, therefore, inject the learned FSOD-specific inductive biases into a fine-tuning based approach.

To explore the potential of the meta-tuning scheme for FSOD, we focus on the details of classification loss functions, based on the observations that FSOD prediction mistakes tend to be in classification rather than localization details [61]. In particular, we first focus on the softmax temperature parameter, for which we define two versions: (i) a simple constant temperature, and (ii) time (fine-tuning iteration index) varying dynamic temperature, parameterized as an exponentiated polynomial. In all cases, the parameters learned via meta-tuning yield an interpretable loss function that has a negligible risk of over-fitting to the base classes, in contrast to a complex meta-model. We also model augmentation magnitudes during meta-tuning for improving the data loading pipeline for few-shot learning purposes. Additionally, we incorporate a score scaling coefficient for learning to balance base versus novel class scores.

We provide an experimental analysis on the Pascal VOC [13] and MS-COCO [40] benchmarks for FSOD, using the state-of-the-art fine-tuning based baselines MPSR [72] and DeFRCN [53]. Our experimental results show that the proposed meta-tuning approach provides significant performance gains in both FSOD and Generalized FSOD settings, suggesting that meta-tuning loss functions and data augmentation can be a promising direction in FSOD research.

2. Related Work

This section provides an overview of recent developments on few-shot image classification, few-shot object detection, automated loss function and data augmentation discovery.

Few-shot classification. Most of the meta-learning approaches for few-shot learning (FSL) of classification models can be grouped as *adaptation-based* and *mapping-based* approaches. Adaptation-based (also called *gradient-based*) approaches aim to learn model parameters that can easily be adapted to new unseen few-shot tasks within a few model update steps, *e.g.* [17, 38, 47, 48, 51, 54, 58]. Mapping-based approaches (also called *metric-based*) aim to bypass a gradient-descent based adaptation step, and instead learn a data-to-classifier mapping, *e.g.* [5, 44, 49, 59, 60, 62, 64, 77, 78, 80].

Some of the other notable approaches include learning to generate synthetic data for novel classes [23, 33, 68], using better feature representations [1, 2, 19, 28, 41, 63, 67] or utilizing differentiable convex solvers [3, 34]. Importantly, several works highlight that a carefully trained representation combined with simple fine-tuning or even just shallow classifiers can yield competitive or better performance than meta-learning based approaches, *e.g.* [9, 12, 63].

Few-shot object detection. The FSOD approaches can be summarized as meta-learning and fine-tuning (also called *transfer-learning*) based ones. Most meta-learning based FSOD approaches embrace formulations similar to those used in mapping-based meta-learning approaches for FSL, *e.g.* [8, 22, 31, 36, 52, 75, 76, 79, 81, 83]. Support feature aggregation is one of the main aspects where meta-learning-based methods differ from each other. Xiao and Marlet [75] use both the differences and the channel-wise multiplication of the features in addition to the combination of the features directly for support-query aggregation. Fan *et al.* [14] use attention blocks to make support and query features more distinguishable for base and novel object classes. Zhang *et al.* [81] use inter-class correlations to highlight important support features. Li *et al.* [36] propose to use specialized support and query features for classification and localization.

Recent efforts towards improving meta-learning based FSOD include complimentary techniques, mainly to improve loss functions, feature matching, and novel class sample usage efficiency. [36] uses class margin loss, [26] uses margin-based ranking loss, [82] uses hybrid loss which consist of focal loss, adaptive margin loss and contrastive loss. Hu *et al.* [27] perform feature matching between query and support images to use the information from the support images more effectively. Similarly, Han *et al.* [21] construct a matching network between query and support instances using heterogeneous graph convolutional networks. Li and Li [35] augment novel class samples via adding Gaussian noise. Yin *et al.* [79] decouple classification task from localization by using the proposed class-conditional architecture.

Fine-tuning-based methods typically freeze parts of a pre-trained detection network, add auxiliary detection heads, increase the novel class variances and then apply gradient descent based model update steps, unlike meta-learning-based methods that use complex episodic learning [15, 32, 53, 61, 65, 71, 72]. Wang *et al.* [65] propose a Faster-RCNN [56] based approach, where the class-agnostic region proposal network (RPN) component is kept frozen during fine-tuning. Sun *et al.* [61] use a similar approach and differently include FPN and RPN layers to the learnable parameter set in the same architecture. These learnable layers allow using contrastive proposal encodings that facilitate the more accurate classification of novel objects. Wu *et al.* [72] show that the scale distribution of support set tends to be imbalanced, and proposes a multi-scale positive sample refinement (MPSR) branch as an addition to the main model. Fan *et al.* [15] propose Retentive R-CNN architecture to prevent forgetting during fine-tuning for base classes. The obtained object proposals are fed into two ROI detectors responsible for base class and novel class instances. Qiao *et al.* [53] focus on decoupling network modules, and introduce a gradient decoupling layer and prototypical calibration block. Kaul *et al.* [32] extend the novel class annotations in the training set. In this context, the proposed method obtains object candidates from the base detector, and applies the box refinement step.

While our approach is based on fine-tuning based FSOD, we embrace meta-learn principles to optimize the loss function and augmentations to improve the fine-tuning process for FSOD, without learning a complex and over-fitting-prone meta-model. The resulting loss function and data augmentations are then utilized within the fine-tuning steps.

Automated loss function discovery. Loss function discovery is an emerging AutoML topic towards improving the learning systems in a data-driven manner. Existing methods are mainly based on either (i) constructing the loss function directly from the basic operators [20, 42, 55] or (ii) optimizing parameterized loss functions [37, 66]. For loss construction, [42] proposes a genetic algorithm that consists of loss function verification and quality filtering modules. In this approach, the predefined proxy task eliminates divergent and poor candidate loss functions and survives the promising loss functions for other steps. [20] uses a genetic algorithm to select candidate loss functions from a tree of simple mathematical operations, and the successful loss functions pass to other stages to mutate. [55] suggests a method to learn not only the loss function but also the whole machine learning algorithm from scratch. For loss optimization, [37] re-analyzes the existing loss functions and presents them in a combined formula. [66] observes that the search space used in [37] can be too complex, and propose to simplify the search space via heuristics. In contrast to these works targeting supervised training scenarios, we aim to adapt loss function learning principles to the FSOD problem.

AutoML for data augmentation. A variety of automated data augmentation techniques have recently been proposed [10, 11, 25, 39]. Cubuk *et al.* [10] generate augmentation policies using reinforcement learning and a controller RNN. Ho *et al.* [25] propose a method that reduces the computational costs compared to [10] by using a population-based framework. Similarly, Lim *et al.* [39] propose a direct Bayesian method to reduce costs. Cubuk *et al.* [11] show that the optimal augmentation magnitudes tend to be similar across transformations, and the search process can greatly be simplified by using a shared value. We follow this suggestion and use a shared magnitude across the transforms in our formulation. In contrast to these works on supervised learning, however, we focus on learning detectors with few-samples.

In summary, while loss function and augmentation discovery topics increasingly attract attention towards improving supervised training pipelines, ours is the first work on learning few-sample specific inductive biases for fine-tuning based few-shot object detection based on meta-learning and AutoML principles, to the best of our knowledge.

3. Method

This section provides a brief summary of the FSOD problem definition and the baseline model we utilize. We then present our definition and instantiation of meta-tuning.

Problem definition. We follow the FSOD setup of [31], where a relatively large set of training images for the set C_b of *base* classes is made available. Each training image corresponds to a tuple (x, y) consisting of image x and annotations $y = \{y_0, ..., y_M\}$. Each object annotation $y_i = \{c_i, b_i\}$ contains a category label (c_i) and a bounding box $(b_i = \{x_i, y_i, w_i, h_i\})$. Once the FSOD model training is complete, the evaluation is carried out based on a limited number (k) of training images made available for the set C_n of distinct *novel* (*i.e.* few-shot) classes.

Base model. We use the MPSR FSOD method [72] as the infrastructure for our loss function and data augmentation search methods. MPSR adapts the Faster-RCNN to be suitable for fine-tuning-based FSOD and uses an auxiliary multi-scale positive sample refinement (MPSR) branch to handle the scale scarcity problems. This branch expands the scale space of positive samples without increasing improper negative instances, unlike feature pyramid networks and image pyramids that do not change data distribution, hence the scale sparsity problem. In this context, objects in the images are cropped and resized in multiple sizes to create scale pyramids. The MPSR uses two groups of loss functions for the region proposal network (RPN) and detection heads, and feeds differently scaled positive samples to these loss functions together with the main detection branch. Finally, we note that the proposed approach can in principle be applied to virtually any fine-tuning based FSOD model.

3.1. Meta-tuning loss functions

Our main goal is to improve few-shot detector fine-tuning based on meta-learning principles. For meta-tuning the FSOD loss, we specifically focus on the classification loss term, as the FSOD errors tend to be primarily caused by misclassifications [61]. The MPSR classification loss term can be expressed as follows:

$$\ell_{cls}(x, y) = -\frac{1}{N_{ROI}} \sum_{i}^{N_{ROI}} \log \left(\frac{e^{f(x_i, y_i)}}{\sum_y e^{f(x_i, y)}} \right) \quad (1)$$

where N_{ROI} is the number of ROIs (*i.e.* candidate regions) in an image, y_i is the groundtruth class label for the i-th ROI, and $f(x_i, y)$ is the corresponding class y prediction score. To add more flexibility into the loss function, we re-define it as a parametric function $\ell_{cls}(x, y; \rho)$, where ρ represents the loss function parameters. First, we introduce a temperature scalar ρ_τ, *i.e.* $\rho = (\rho_\tau)$:

$$\ell_{cls}(x, y; \rho) = -\frac{1}{N_{ROI}} \sum_{i}^{N_{ROI}} \log \left(\frac{e^{f(x_i, y_i)/\rho_\tau}}{\sum_{y'} e^{f(x_i, y')/\rho_\tau}} \right) \quad (2)$$

Our motivation comes from the observations on the importance of temperature scaling in log loss on various other problems, such as knowledge distillation [24], few-shot classification [49, 78], and zero-shot learning [43]. While temperature is typically tuned in a manual manner, here we aim to meta-learn it specifically for fine-tuning based FSOD purposes, giving a chance to observe the behavior of meta-tuning in a simple case. We also define a more sophisticated variant of the loss function by defining the *dynamic temperature* function f_ρ and *novel class scaling* α:

$$\ell_{cls}(x, y; \rho) = \frac{-1}{N_{ROI}} \sum_{i}^{N_{ROI}} \log \left(\frac{e^{\alpha(y_i) f(x_i, y_i)/f_\rho(t)}}{\sum_{y'} e^{\alpha(y') f(x_i, y')/f_\rho(t)}} \right) \quad (3)$$

where $f_\rho(t) = \exp(\rho_a t^2 + \rho_b t + \rho_c)$. Here, $\rho = (\rho_a, \rho_b, \rho_c)$ is a 3-tuple of polynomial coefficients, and $t \in [0, 1]$ is the normalized fine-tuning iteration index. The temperature can increase or decrease over time, making the predicted class distributions smoother or sharper. $\alpha(y)$ is set to 1 for $y \in C_b$, and otherwise the novel class score scaling coefficient ρ_α, as a way to learn base and novel score balancing.

3.2. Meta-tuning augmentations

For meta-tuning augmentations, we focus on the photometric augmentations that are likely to be transferable from base to novel classes. In this context, we model the brightness, saturation, contrast, and hue transforms, with a shared magnitude parameter (ρ_{aug}), which is known to be effective for supervised training [11].

3.3. Meta-tuning procedure

In our work, we utilize a REINFORCE [70] based reinforcement learning (RL) approach to search for the optimal loss function and augmentations, where we use the AutoML approach of Wang *et al.* [66] on loss function search for fully-supervised face recognition as our starting point.

In order to meta-tune the loss function and augmentations to maximize FSOD generalization abilities, we generate *proxy tasks* over base class training data to imitate real FSOD tasks over the novel classes. For this purpose, we divide base classes into two subsets, proxy-base $C_{\text{p-base}}$ and proxy-novel $C_{\text{p-novel}}$. We then construct three non-overlapping data set splits using the base class training set: (i) $D_{\text{p-pretrain}}$ containing $C_{\text{p-base}}$-only samples, used for training a temporary object detection model for meta-tuning purposes; (ii) $D_{\text{p-support}}$ containing samples of $C_{\text{p-base}} \cup C_{\text{p-novel}}$ classes to be used as fine-tuning images during meta-tuning; (iii) $D_{\text{p-query}}$ containing samples of $C_{\text{p-base}} \cup C_{\text{p-novel}}$ classes to be used for evaluating the generalized FSOD performance of a fine-tuned model during meta-tuning.

We generate a series of FSOD proxy tasks for meta-tuning, similar to episodic meta-learning: at each proxy task T, we sample a few-shot training set from $D_{\text{p-support}}$. We also sample a loss function/augmentation magnitude parameter combination ρ, where each $\rho_j \in \rho$ is modeled in terms of a Gaussian distribution: $\rho_j \sim \mathcal{N}(\mu_j, \sigma^2)$. Using the loss function or augmentations corresponding to the sampled ρ, we fine-tune the initial model on the support images using gradient-based optimization, and compute the mean

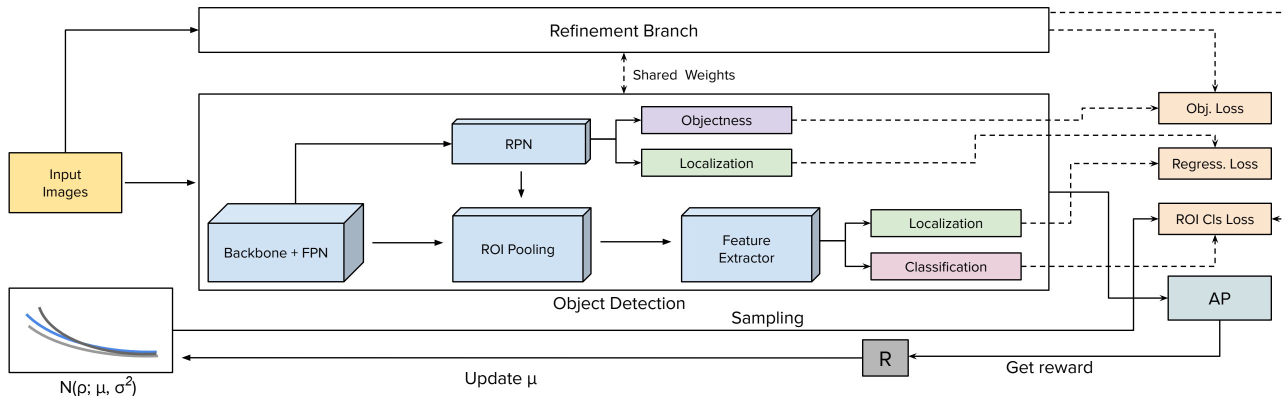


Figure 2. The meta-tuning approach. At each RL iteration over a proxy task, the distribution parameters modeling the loss function and augmentations are updated as a function of the obtained mAP scores, towards improved training with few-samples.

average precision (mAP) scores on $D_{\text{p-query}}$. We get multiple mAP scores by repeating this process multiple times over multiple proxy support samples. Meta-tuning is then carried over by updating μ values via the REINFORCE rule after each episode, towards finding μ values centered around well-performing ρ combinations.

$$\mu_j' \leftarrow \mu_j + \eta R(\rho) \nabla_\mu \log\left(p(\rho_j; \mu_j, \sigma)\right) \tag{4}$$

where $p(\rho; \mu, \sigma)$ is the Gaussian probability density function, η is the RL learning rate.

We apply the REINFORCE update rule using the ρ with the highest reward per episode. $R(\rho)$ is the *normalized* reward function obtained by whitening the mAP scores. We empirically observe that normalization improves the results (Section 4) since without reward normalization, the RL updates are scaled with respect to the inherent difficulty of the proxy task, which greatly varies depending on the sampled support examples. Reward normalization approximately removes the *average* reward, enabling better performing ρ samples to influence based on their *relative* success.

Finally, similar to [50], starting with $\sigma = 0.1$, we diminish σ over the RL iterations to progressively reduce explorations by sampling more conservatively, which improves converge. The final scheme is illustrated in Figure 2.

4. Experiments

Metrics. We use mAP to evaluate the base and novel class detection results separately. To evaluate the generalized FSOD performance, we use the Harmonic Mean (HM) metric to compute a balanced aggregation of base and novel class performance scores. Adapted from generalized zero-shot learning [74], HM is defined as the harmonic mean of mAP_{base} and $\text{mAP}_{\text{novel}}$ scores.

Datasets. We use Pascal VOC [13] and MS COCO [40] with the same splits defined in FSOD benchmarks [65, 72]. On Pascal VOC, three separate base/novel class splits exist, where each one consists of 15 base and 5 novel classes. In each split, we select 5 base classes to mimic novel classes during meta-tuning. On MS-COCO, we select 15 base classes to mimic novel classes in each proxy task, and evaluate the models for the 10-shot and 30-shot settings.

Baselines. We primarily use the MPSR [72] and DeFRCN [53] as our baselines, which are among the best performing fine-tuning based FSOD methods on Pascal VOC. For the DeFRCN experiments, we transfer the meta-tuned loss functions and augmentation magnitudes from MPSR to the DeFRCN method, which are both based on Faster-RCNN. We take the results for FRCN [76], Ret. R-CNN [15], Meta-RCNN [76], FSRW [31], MetaDet [69], FsDetView [75] and ONCE [52] from [15] for a fair comparison. For the MPSR, DeFRCN (*seed* is set to 0) and FSCE [61], we report the results we obtain experimentally. We take the results for TFA+Hal [84], CME [36], TIP [35], DCNet [27], QA-FewDet [21] FADI [6], LVC [32], KFSOD [83] and FCT [22] from the original papers. Finally, while it is difficult to fairly compare fine-tuning versus meta-learning based approaches, we provide a discussion in the supplementary material.

Implementation details. We use 200 RL episodes for loss function meta-tuning, with REINFORCE learning rate set to 0.0005. The meta-tuning for augmentation parameter is carried out using the trained and frozen the loss function parameters. We keep the fine-tuning implementation details of MPSR unchanged, which uses 4000 and 8000 gradient descent iterations for 10-shot and 30-shot experiments on MS-COCO, and 2000 iterations on Pascal VOC. We will publish the full source code upon publication; a preliminary version is provided as supplementary material.

4.1. Main results

We first compare the meta-tuning results against the corresponding MPSR baseline in Table 1. In the table, *Meta-*

Method/Shot	Pascal VOC										MS-COCO			
	Novel Classes					All Classes (HM)					Novel Classes		All Classes (HM)	
	1	2	3	5	10	1	2	3	5	10	10	30	10	30
MPSR [72]	33.1	37.2	44.3	47.1	52.1	43.1	47.4	54.5	57.2	60.8	9.1	13.7	11.5	15.0
MPSR+Meta-Static	33.4	39.4	45.1	47.3	52.6	43.7	50.4	55.4	57.5	61.4	10.1	14.8	12.7	16.4
MPSR+Meta-Dynamic	34.5	39.8	45.0	48.2	52.5	45.0	51.0	55.5	58.3	61.6	11.9	14.9	14.3	16.6
MPSR+Meta-ScaledDynamic	35.2	40.3	45.8	48.4	52.9	45.6	51.2	55.9	58.3	61.8	12.3	15.0	14.4	16.7
MPSR+Aug	34.6	38.6	46.0	48.3	52.7	45.1	49.5	56.2	58.4	62.0	9.9	14.9	12.5	16.3
MPSR+Meta-Static+Aug	35.3	39.1	46.1	48.4	52.7	45.9	49.9	56.2	58.3	61.8	10.2	15.2	12.8	16.7
MPSR+Meta-Dynamic+Aug	35.4	39.6	46.5	48.9	53.3	46.0	50.5	56.8	58.9	62.5	12.1	15.3	14.5	16.8
MPSR+Meta-ScaledDynamic+Aug	**35.8**	**40.6**	**46.8**	**49.2**	**53.7**	**46.3**	**51.5**	**57.0**	**59.2**	**62.7**	**12.5**	**15.4**	**14.7**	**16.9**

Table 1. FSOD (mAP) and G-FSOD (HM of the base and novel class mAPs) results on Pascal VOC and MS-COCO datasets for MPSR baseline method. HM stands for harmonic mean.

Static, *Meta-Dynamic*, *Meta-ScaledDynamic* refer to meta-tuning a single temperature, dynamic temperature, and novel class scaled dynamic temperature functions, respectively. Similarly, *Aug*, *Meta-Static+Aug*, *Meta-Dynamic+Aug*, and *Meta-ScaledDynamic+Aug* refer to meta-tuning only augmentation, single temperature and augmentation, dynamic temperature and augmentation, and novel class scaled dynamic temperature and augmentation functions, respectively. We observe that meta-tuning consistently improves the FSOD and G-FSOD results of the MPSR model. We also observe steady improvements gradually from the baseline to Meta-Static, to Meta-Dynamic, and finally to Meta-ScaledDynamic. In addition, the meta-tuned augmentation magnitude parameter also contributes positively to the few-shot object detection performance. The overall consistency of the improvements provides positive evidence for the value of loss and augmentation meta-tuning.

Pascal VOC results. In Table 2, we report the Pascal VOC results for our MPSR and DeFRCN based Meta-ScaledDynamic+Aug approach and compare them against the state-of-the-art fine-tuning based FSOD methods. While we present the scores averaged over the three splits in this table, additional per-split FSOD and G-FSOD results can be found in the supplementary material. The left side of Table 2 presents the FSOD results for the varying number of support images. We observe that DeFRCN combined with Meta-ScaledDynamic+Aug, *i.e.* meta-tuning of the score coefficient, dynamic temperature and the augmentation parameter, yields the best mAP scores in all k-shot settings among all methods.

The right side of Table 2 presents the G-FSOD results on Pascal VOC. We observe that the best-performing Meta-ScaledDynamic+Aug method improves the HM scores further above the state-of-the-art in all k-shot settings. Overall, these results suggest that the proposed framework is an effective way for meta-learning inductive biases to be used in fine-tuning-based FSOD.

Figure 3 presents visual detection examples without and with meta-tuned scaled dynamic temperature and augmentations in the first and second rows, respectively. We observe various improvements, such as reductions in false positives, improved recall, and more precise boxes, most likely due to the improved model fitting in the low-data regime.

MS-COCO results. In Table 3, we compare the MPSR and DeFRCN based Meta-ScaleDynamic+Aug results against other fine-tuning based FSOD methods that report 10-shot and 30-shot results on the MS-COCO dataset. We observe that with meta-tuning, the FSOD scores of MPSR improve from 9.1 to 12.5 (10-shot mAP), and from 13.7 to 15.4 (30-shot mAP). We also observe that the scores of DeFRCN improve from 18.5 to 18.8 (10-shot mAP), and from 21.9 to 23.4 (30-shot mAP), obtaining the best and second best results against all other models. Similarly, in the case of G-FSOD, with meta-tuning, the 10-shot HM score of DeFRCN improves from 24.0 to 24.4, outperforming all other models. In addition, the 30-shot HM score of DeFRCN improves from 26.8 to 28.0, which is slightly below the 28.1 score of LVC-PL [32].

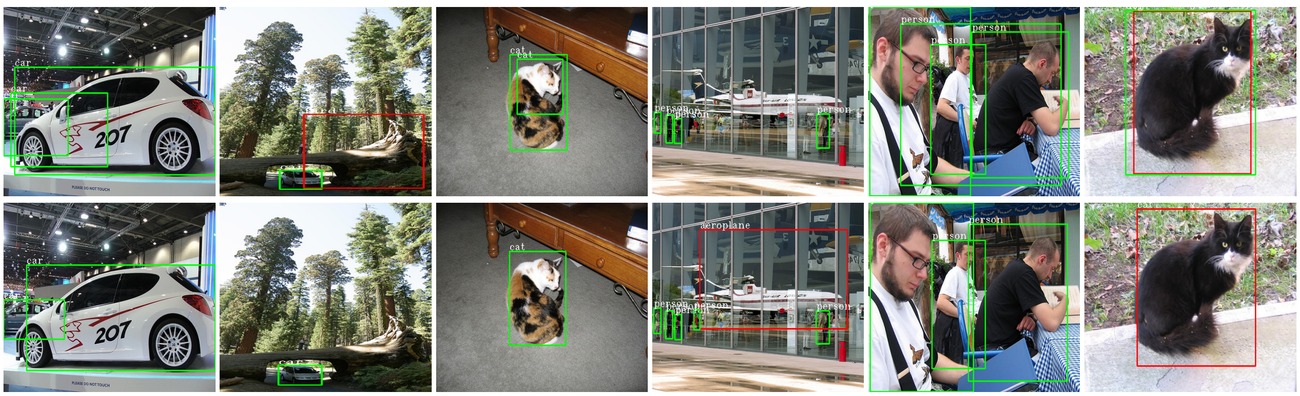
4.2. Ablation studies

Meta-tuning details. The proposed meta-tuning approach involves three important technical details: *Proxy-novel imitation*, *model re-initialization*, and *reward normalization*. Proxy-novel imitation refers to reinforcement learning over the sampled proxy-novel tasks, instead of the whole training set, to mimic the test-time FSOD challenges. Model re-initialization is the re-initialization of the base model for each task. Without re-initialization, not only the sampled loss/augmentation parameters and tasks but also the accumulated model updates undesirably affect the rewards. Reward normalization further reduces the effect of task difficulty variance by normalizing the rewards obtained within a single episode, allowing a more isolated assessment of the sampled loss functions and augmentations.

We evaluate the contributions of these three important details in terms of G-FSOD HM scores using the 5-shot setting of Pascal VOC Split-1 with MPSR+Meta-Dynamic. The results averaged over 5 runs are given in Table 4. We observe that each component progressively improves the

Method/Shot	Novel Classes 1	2	3	5	10	All Classes (HM) 1	2	3	5	10
FRCN [76] (ICCV'19)	16.1	20.6	28.8	33.4	36.5	25.9	31.7	40.0	44.3	46.7
TFA-fc [65] (ICML'20)	27.6	30.6	39.8	46.6	48.7	40.5	44.1	52.9	58.3	59.9
TFA-cos [65] (ICML'20)	31.4	32.6	40.5	46.8	48.3	44.6	46.0	53.5	58.4	59.6
FSCE [61] (CVPR'21)	29.2	36.3	42.5	47.1	52.2	41.8	48.8	54.2	57.7	61.0
Ret. R-CNN [15] (CVPR'21)	31.4	37.1	41.4	46.8	48.8	44.7	50.5	54.7	59.1	60.8
TFA+Hal [84] (CVPR'21)	32.9	35.5	40.4	46.3	48.1	-	-	-	-	-
FADI [6] (NeurIPS'21)	42.2	46.5	47.9	52.4	56.9	-	-	-	-	-
LVC [32] (CVPR'22)	30.9	35.4	43.6	51.1	54.1	-	-	-	-	-
LVC-PL [32] (CVPR'22)	45.2	45.0	54.8	57.5	58.6	-	-	-	-	-
MPSR [72] (ECCV'20)	33.1	37.2	44.3	47.1	52.1	43.1	47.4	54.5	57.2	60.8
DeFRCN [53] (ICCV'21)	**46.5**	**52.6**	**55.9**	**60.0**	**60.8**	**57.6**	**62.5**	**64.7**	**67.6**	**67.8**
MPSR+Meta-ScaledDynamic+Aug	35.8	40.6	46.8	49.2	53.7	46.3	51.5	57.0	59.2	62.7
DeFRCN+Meta-ScaledDynamic+Aug	**49.2**	**54.0**	**57.2**	**61.3**	**61.8**	**59.8**	**63.7**	**65.9**	**68.6**	**68.7**

Table 2. FSOD (mAP) and G-FSOD (HM of the base and novel class mAPs) results on Pascal VOC. The best and the second-best results are marked with red and blue. HM stands for harmonic mean.



Split-1/5-shot Split-1/10-shot Split-2/5-shot Split-2/10-shot Split-3/5-shot Split-3/10-shot

Figure 3. Qualitative results using MPSR without (first row) and with (second row) meta-tuning, over multiple Pascal VOC splits. Base and novel class detections are shown with green and red boxes, respectively. (Best viewed in color.)

Method/Shots	Novel Classes 10-shot	30-shot	All Classes (HM) 10-shot	30-shot
FRCN [76] (ICCV'19)	9.2	12.5	12.8	15.6
FRCN-BCE [65] (ICML'20)	6.4	10.3	10.9	16.1
TFA-fc [65] (ICML'20)	10.0	13.4	15.4	19.4
TFA-cos [65] (ICML'20)	10.0	13.7	15.6	19.8
MPSR [72] (ECCV'20)	9.1	13.7	11.5	15.0
FSCE [61] (CVPR'21)	10.5	14.4	16.0	20.2
Ret. R-CNN [15] (CVPR'21)	10.5	13.8	16.6	20.4
FADI [6] (NeurIPS'21)	12.2	16.1	-	-
DeFRCN [53] (ICCV'21)	**18.5**	21.9	**24.0**	26.8
LVC [32] (CVPR'22)	12.1	17.8	17.8	22.8
LVC-PL [32] (CVPR'22)	17.8	**24.5**	22.8	**28.1**
MPSR+Meta-ScaledDynamic+Aug	12.5	15.4	14.7	16.9
DeFRCN+Meta-ScaledDynamic+Aug	**18.8**	**23.4**	**24.4**	**28.0**

Table 3. Comparison of Meta-ScaledDynamic results to the fine-tuning based (G-)FSOD methods on the MS-COCO dataset. The best and the second-best results are marked with red and blue.

Proxy-novel imit.	Model re-init.	Reward norm.	HM
✗	✗	✗	61.5
✓	✗	✗	61.8
✓	✓	✗	62.1
✓	✓	✓	63.3

Table 4. Evaluation of meta-tuning details. *Proxy-novel imitation* is the imitation of novel classes using a subset of base classes. *Model re-initialization* is the re-initialization of the base model at each task. *Reward normalization* is within-episode normalization of the mAP scores during meta-tuning.

HM scores, and the most significant contribution is made by reward normalization, which improves from 62.1 to 63.3. We also observe that reward normalization considerably improves the overall experimental stability. To quantify this observation, we estimate the 95% confidence interval over the runs using $CI = 1.96\frac{s}{\sqrt{n}}$, where s, n, and 1.96 are the standard deviation, number of runs, and Z-value, respectively [65]. According to this estimator, the normalization step narrows the confidence interval from ±0.75 to ±0.13, providing a clear improvement in reliability.

Learned loss functions. In Figure 4, we plot the learned loss functions according to the μ values obtained at the end of the RL process. The upper plot shows the dynamic

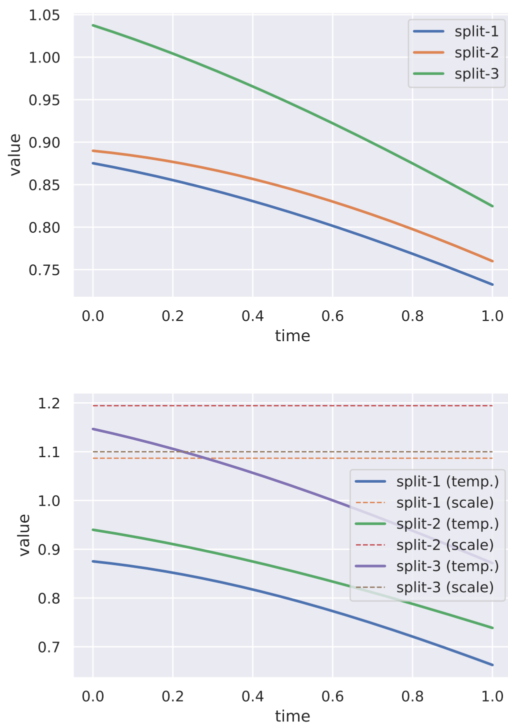


Figure 4. The dynamic temperature functions and score scaling coefficients learned by the meta-tuning process, using Meta-Dynamic (upper) and Meta-ScaledDynamic (lower) formulations. Results for each Pascal VOC split is shown with a separate curve.

temperature functions learned over three different splits. We observe that temporally attenuated temperature values are preferred consistently, sharpening the predictions towards the end of the fine-tuning process. The lower plot shows the learned dynamic temperature functions with novel class score scaling. The learned scaling coefficients, *i.e.* μ_α of the learned ρ_α distribution, are shown as horizontal lines. We observe that similar dynamic temperature functions are learned, and μ_α values vary between 1.09 to 1.2, suggesting that the meta-tuning process learns to boost the novel class scores. The interpretability of these outcomes, we believe, highlights a significant advantage of loss meta-tuning. In the context of interpretability, we observe that as the fine-tuning process continues on the few-shot training set, the predictions are progressively made sharper, *i.e.* the loss becomes more sensitive to classification errors and enforces towards making more confident correct predictions. This is in alignment with one of our original motivations for reducing the dominating classification errors in G-FSOD, as the meta-tuning process automatically learns to enforce more accurate classifications, where the curve steepness and the numerical ranges are learned via RL.

Learned augmentations. The learned photometric augmentation magnitude values learned are 0.29, 0.24, 0.13, and 0.36 for Pascal VOC split-1, split-2, split-3, and MS-COCO datasets, respectively. We observe that the learned augmentation magnitudes positively contribute to the performance. According to the results in Table 1, the average Pascal VOC split-1/1-shot score increases from 33.1 to 34.6 with only augmentation steps.

S/M	TFA [65]	TFA+Hal [84]	TFA+Meta-ScaledDynamic+Aug
1	3.4	3.8	**4.7**
2	4.6	5.0	**5.8**
3	6.6	6.9	**7.1**

Table 5. Low-shot (1-shot, 2-shot and 3-shot) experiments on MS-COCO dataset with novel classes.

Very low-shot experiments. Finally, we evaluate the meta-tuning approach in low-shot many-class settings. [84] proposes TFA+Hal method that uses the TFA baseline and conducts 1-shot, 2-shot, and 3-shot FSOD on the MS-COCO dataset. As we already observe the positive effects of the loss terms and augmentation magnitudes obtained from the MPSR on the DeFRCN, we similarly apply the learned parameters to the TFA baseline. The results are presented in Table 5. We observe that results are consistently improved using the meta-tuned functions on the TFA baseline.

5. Conclusion

Fine-tuning based frameworks offer simple and reliable approaches to building detection models from few samples. However, a major limitation of the existing fine-tuning-based FSOD models is their focus on the hand-crafting the design of fine-tuning details for few-shot training, which is inherently difficult and likely to be sub-optimal. Towards addressing this limitation, we propose to meta-learn the fine-tuning based learning dynamics as a way of introducing learned inductive biases for few-shot learning. The proposed tuning scheme uses meta-learning principles with reinforcement learning, and obtains interpretable loss functions and augmentation magnitudes for few-shot training. Our comprehensive experimental results on Pascal VOC and MS COCO datasets show that the proposed meta-tuning approach consistently provides significant performance improvements over the strong fine-tuning based few-shot detection baselines in both FSOD and G-FSOD settings.

While we restrict our experiments to loss and augmentation functions, meta-tuning other learning components, *e.g.* initial model, and applications to other few-shot learning problems can be interesting future work directions.

Acknowledgements. This work was supported in part by the TUBITAK Grant 119E597 and a Google Faculty Research Award.

References

[1] Peyman Bateni, Jarred Barber, Jan-Willem van de Meent, and Frank Wood. Enhancing few-shot image classification with unlabelled examples. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2796–2805, January 2022. 2

[2] Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal. Improved Few-Shot Visual Classification. *arXiv e-prints*, page arXiv:1912.03432, Dec. 2019. 2

[3] Luca Bertinetto, Joao F. Henriques, Philip Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *Proc. Int. Conf. Learn. Represent.*, 2019. 2

[4] Malik Boudiaf, Hoel Kervadec, Ziko Imtiaz Masud, Pablo Piantanida, Ismail Ben Ayed, and Jose Dolz. Few-Shot Segmentation Without Meta-Learning: A Good Transductive Inference Is All You Need? In *arXiv:2012.06166 [cs]*, 2021. 1

[5] Kaidi Cao, Maria Brbić, and Jure Leskovec. Concept learners for few-shot learning. In *Proc. Int. Conf. Learn. Represent.*, 2021. 2

[6] Yuhang Cao, Jiaqi Wang, Ying Jin, Tong Wu, Kai Chen, Ziwei Liu, and Dahua Lin. Few-shot object detection via association and discrimination. *Proc. Adv. Neural Inf. Process. Syst.*, 34:16570–16581, 2021. 1, 5, 7

[7] Liangyu Chen, Tong Yang, Xiangyu Zhang, Wei Zhang, and Jian Sun. Points as queries: Weakly semi-supervised object detection by points. *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8819–8828, 2021. 1

[8] Tung-I Chen, Yueh-Cheng Liu, Hung-Ting Su, Yu-Cheng Chang, Yu-Hsiang Lin, Jia-Fong Yeh, Wen-Chin Chen, and Winston Hsu. Dual-awareness attention for few-shot object detection. *IEEE Transactions on Multimedia*, 2021. 1, 2

[9] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A Closer Look at Few-shot Classification. In *ICLR 2019*, 2019. 1, 2

[10] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501*, 2018. 3

[11] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshops*, pages 702–703, 2020. 3, 4

[12] Guneet S Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A Baseline for Few-Shot Image Classification. In *ICLR*, page 20, 2020. 1, 2

[13] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 2, 5

[14] Qi Fan, Wei Zhuo, Chi-Keung Tang, and Yu-Wing Tai. Few-shot object detection with attention-rpn and multi-relation detector. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4013–4022, 2020. 3

[15] Zhibo Fan, Yuchen Ma, Zeming Li, and Jian Sun. Generalized few-shot object detection without forgetting. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4527–4536, 2021. 1, 2, 3, 5, 7

[16] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017. 1, 2

[17] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. Int. Conf. Mach. Learn.*, volume 70, pages 1126–1135, 2017. 2

[18] Golnaz Ghiasi, Yin Cui, A. Srinivas, Rui Qian, Tsung-Yi Lin, Ekin Dogus Cubuk, Quoc V. Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2917–2927, 2021. 1

[19] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord. Boosting few-shot visual learning with self-supervision. In *Proc. IEEE Int. Conf. on Computer Vision*, 2019. 2

[20] Santiago Gonzalez and Risto Miikkulainen. Improved training speed, accuracy, and data utilization through loss function optimization. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. IEEE, 2020. 2, 3

[21] Guangxing Han, Yicheng He, Shiyuan Huang, Jiawei Ma, and Shih-Fu Chang. Query adaptive few-shot object detection with heterogeneous graph convolutional networks. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 3263–3272, 2021. 3, 5

[22] Guangxing Han, Jiawei Ma, Shiyuan Huang, Long Chen, and Shih-Fu Chang. Few-shot object detection with fully cross-transformer. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5321–5330, 2022. 1, 2, 5

[23] Bharath Hariharan and Ross B. Girshick. Low-shot visual recognition by shrinking and hallucinating features. *Proc. IEEE Int. Conf. on Computer Vision*, pages 3037–3046, 2017. 2

[24] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 4

[25] Daniel Ho, Eric Liang, Xi Chen, Ion Stoica, and Pieter Abbeel. Population based augmentation: Efficient learning of augmentation policy schedules. In *Proc. Int. Conf. Mach. Learn.*, pages 2731–2741. PMLR, 2019. 3

[26] Ting-I Hsieh, Yi-Chen Lo, Hwann-Tzong Chen, and Tyng-Luh Liu. One-shot object detection with co-attention and co-excitation. *arXiv preprint arXiv:1911.12529*, 2019. 3

[27] Hanzhe Hu, Shuai Bai, Aoxue Li, Jinshi Cui, and Liwei Wang. Dense relation distillation with context-aware aggregation for few-shot object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 10185–10194, 2021. 3, 5

[28] Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim, and Timothy M. Hospedales. Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference. *arXiv e-prints*, page arXiv:2204.07305, Apr. 2022. 2

[29] Zeyi Huang, Yang Zou, BVK Kumar, and Dong Huang. Comprehensive attention self-distillation for weakly-supervised object detection. *Proc. Adv. Neural Inf. Process. Syst.*, 33, 2020. 1

[30] Taewon Jeong and Heeyoung Kim. Ood-maml: Meta-learning for few-shot out-of-distribution detection and classification. In *Advances in Neural Information Processing Systems*, volume 33, pages 3907–3916, 2020. 1

[31] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 8420–8429, 2019. 1, 2, 3, 5

[32] Prannay Kaul, Weidi Xie, and Andrew Zisserman. Label, verify, correct: A simple few shot object detection method. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14237–14247, 2022. 1, 3, 5, 6, 7

[33] Michalis Lazarou, Yannis Avrithis, and Tania Stathaki. Tensor feature hallucination for few-shot learning. *ArXiv*, abs/2106.05321, 2021. 2

[34] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 10649–10657, 2019. 2

[35] Aoxue Li and Zhenguo Li. Transformation invariant few-shot object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3094–3102, 2021. 3, 5

[36] Bohao Li, Boyu Yang, Chang Liu, Feng Liu, Rongrong Ji, and Qixiang Ye. Beyond max-margin: Class margin equilibrium for few-shot object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7363–7372, 2021. 1, 2, 3, 5

[37] Chuming Li, Xin Yuan, Chen Lin, Minghao Guo, Wei Wu, Junjie Yan, and Wanli Ouyang. Am-lfs: Automl for loss function search. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 8410–8419, 2019. 3

[38] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-SGD: Learning to Learn Quickly for Few-Shot Learning. *arXiv e-prints*, page arXiv:1707.09835, July 2017. 2

[39] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast autoaugment. *Proc. Adv. Neural Inf. Process. Syst.*, 32, 2019. 3

[40] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Proc. European Conf. on Computer Vision*, pages 740–755. Springer, 2014. 2, 5

[41] Bin Liu, Yue Cao, Yutong Lin, Qi Li, Zheng Zhang, Mingsheng Long, and Han Hu. Negative margin matters: Understanding margin in few-shot classification. *arXiv preprint arXiv:2003.12060*, 2020. 2

[42] Peidong Liu, Gengwei Zhang, Bochao Wang, Hang Xu, Xiaodan Liang, Yong Jiang, and Zhenguo Li. Loss function discovery for object detection via convergence-simulation driven search. *arXiv preprint arXiv:2102.04700*, 2021. 2, 3

[43] Shichen Liu, Mingsheng Long, Jianmin Wang, and Michael I Jordan. Generalized Zero-Shot Learning with Deep Calibration Network. In *NeurIPS*, pages 2005–2015. 2018. 4

[44] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sungju Hwang, and Yi Yang. Learning to propagate labels: Transductive propagation network for few-shot learning. In *Proc. Int. Conf. Learn. Represent.*, 2019. 2

[45] Yan Liu, Zhijie Zhang, Li Niu, Junjie Chen, and Liqing Zhang. Mixed supervised object detection by transferring mask prior and semantic similarity. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Proc. Adv. Neural Inf. Process. Syst.*, 2021. 1

[46] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *Proc. IEEE Int. Conf. on Computer Vision*, 2021. 1

[47] Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, and Adam Trischler. Rapid adaptation with conditionally shifted neurons. In *ICML*, 2018. 2

[48] Alex Nichol and John Schulman. Reptile: a scalable metalearning algorithm. *arXiv: Learning*, 2018. 2

[49] Boris N. Oreshkin, Pau Rodriguez Lopez, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In *NeurIPS*, 2018. 2, 4

[50] Matteo Papini, Andrea Battistello, and Marcello Restelli. Balancing learning speed and stability in policy gradient via adaptive exploration. In *Proc. Int. Conf. on Artif. Intellig. and Stat.*, pages 1188–1199, 2020. 5

[51] Eunbyung Park and Junier B. Oliva. Meta-curvature. In *NeurIPS*, 2019. 2

[52] Juan-Manuel Perez-Rua, Xiatian Zhu, Timothy M Hospedales, and Tao Xiang. Incremental few-shot object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13846–13855, 2020. 1, 2, 5

[53] Limeng Qiao, Yuxuan Zhao, Zhiyuan Li, Xi Qiu, Jianan Wu, and Chi Zhang. Defrcn: Decoupled faster r-cnn for few-shot object detection. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 8681–8690, 2021. 1, 2, 3, 5, 7

[54] Aravind Rajeswaran, Chelsea Finn, Sham M. Kakade, and Sergey Levine. Meta-learning with implicit gradients. In *NeurIPS*, 2019. 2

[55] Esteban Real, Chen Liang, David So, and Quoc Le. Automl-zero: Evolving machine learning algorithms from scratch. In *Proc. Int. Conf. Mach. Learn.*, pages 8007–8019. PMLR, 2020. 3

[56] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Proc. Adv. Neural Inf. Process. Syst.*, 28:91–99, 2015. 3

[57] Zhongzheng Ren, Zhiding Yu, Xiaodong Yang, Ming-Yu Liu, Yong Jae Lee, Alexander G. Schwing, and Jan Kautz. Instance-aware, context-focused, and memory-efficient weakly supervised object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 1

[58] Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In *Proc. Int. Conf. Learn. Represent.*, 2019. 2

[59] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *Proc. Int. Conf. Mach. Learn.*, volume 48, pages 1842–1850, 2016. 2

[60] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 30, 2017. 2

[61] Bo Sun, Banghuai Li, Shengcai Cai, Ye Yuan, and Chi Zhang. Fsce: Few-shot object detection via contrastive proposal encoding. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7352–7362, 2021. 1, 2, 3, 4, 5, 7

[62] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H.S. Torr, and Timothy M. Hospedales. Learning to compare: Relation network for few-shot learning. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1199–1208, 2018. 2

[63] Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B. Tenenbaum, and Phillip Isola. Rethinking Few-Shot Image Classification: a Good Embedding Is All You Need? In *Proc. European Conf. on Computer Vision*, 2020. 1, 2

[64] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, koray kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 29, 2016. 2

[65] Xin Wang, Thomas E Huang, Trevor Darrell, Joseph E Gonzalez, and Fisher Yu. Frustratingly simple few-shot object detection. *arXiv preprint arXiv:2003.06957*, 2020. 1, 2, 3, 5, 7, 8

[66] Xiaobo Wang, Shuo Wang, Cheng Chi, Shifeng Zhang, and Tao Mei. Loss function search for face recognition. In *Proc. Int. Conf. Mach. Learn.*, pages 10029–10038. PMLR, 2020. 3, 4

[67] Yan Wang, Wei-Lun Chao, Kilian Q. Weinberger, and Laurens van der Maaten. Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. *arXiv preprint arXiv:1911.04623*, 2019. 2

[68] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-Shot Learning from Imaginary Data. *arXiv:1801.05401 [cs]*, Jan. 2018. 2

[69] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Meta-learning to detect rare objects. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 9925–9934, 2019. 5

[70] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992. 4

[71] Aming Wu, Suqi Zhao, Cheng Deng, and Wei Liu. Generalized and discriminative few-shot object detection via svd-dictionary enhancement. *Proc. Adv. Neural Inf. Process. Syst.*, 34:6353–6364, 2021. 3

[72] Jiaxi Wu, Songtao Liu, Di Huang, and Yunhong Wang. Multi-scale positive sample refinement for few-shot object detection. In *Proc. European Conf. on Computer Vision*, pages 456–472. Springer, 2020. 1, 2, 3, 4, 5, 6, 7

[73] Xiongwei Wu, Doyen Sahoo, and Steven Hoi. Meta-rcnn: Meta learning for few-shot object detection. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1679–1687, 2020. 1

[74] Yongqin Xian, Bernt Schiele, and Zeynep Akata. Zero-shot learning-the good, the bad and the ugly. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4582–4591, 2017. 5

[75] Yang Xiao and Renaud Marlet. Few-shot object detection and viewpoint estimation for objects in the wild. In *Proc. European Conf. on Computer Vision*, pages 192–210. Springer, 2020. 1, 2, 3, 5

[76] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. Meta r-cnn: Towards general solver for instance-level low-shot learning. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 9577–9586, 2019. 1, 2, 5, 7

[77] Huaxiu Yao, Linjun Zhang, and Chelsea Finn. Meta-learning with fewer tasks through task interpolation. In *Proceeding of the 10th International Conference on Learning Representations*, 2022. 2

[78] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8805–8814, 2020. 2, 4

[79] Li Yin, Juan M Perez-Rua, and Kevin J Liang. Sylph: A hypernetwork framework for incremental few-shot object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9035–9045, 2022. 1, 2, 3

[80] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. Deepemd: Few-shot image classification with differentiable earth mover's distance and structured classifiers. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, June 2020. 2

[81] Gongjie Zhang, Zhipeng Luo, Kaiwen Cui, Shijian Lu, and Eric P Xing. Meta-detr: Image-level few-shot detection with inter-class correlation exploitation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1, 2, 3

[82] Lu Zhang, Shuigeng Zhou, Jihong Guan, and Ji Zhang. Accurate few-shot object detection with support-query mutual guidance and hybrid loss. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14424–14432, 2021. 3

[83] Shan Zhang, Lei Wang, Naila Murray, and Piotr Koniusz. Kernelized few-shot object detection with efficient integral aggregation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 19207–19216, 2022. 1, 2, 5

[84] Weilin Zhang and Yu-Xiong Wang. Hallucination improves few-shot object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13008–13017, 2021. 1, 2, 5, 7, 8