# Mitigating Negative Gradient Bias in Test-Time Learning for Few-Shot Object Detection

Taijin Zhao, Heqian Qiu*, Lanxiao Wang*, Yu Dai, Qingbo Wu, Fanman Meng, Hongliang Li*
*School of Information and Communication Engineering*
*University of Electronic Science and Technology of China*
Chengdu, China
zhtjww@std.uestc.edu.cn, hqqiu@uestc.edu.cn, lanxiaowang@uestc.edu.cn, ydai@std.uestc.edu.cn,
qbwu@uestc.edu.cn, fmmeng@uestc.edu.cn, hlli@uestc.edu.cn

***Abstract*—Test-Time Learning for Few-Shot Object Detection (TTL-FSOD) aims to adapt few-shot detectors to unlabeled test data, typically via teacher-student pseudo-labeling. However, this paradigm is undermined by the intrinsic sensitivity of detectors to pseudo-labeling noise. In this work, we identify three critical challenges: (1) semantic corruption, where false negative pseudo-labels cause the model to distort learned correct class features; (2) long-tailed class imbalance, where dominant base classes suppress novel class learning; and (3) foreground-background proposal imbalance, where excessive background proposals overwhelm sparse foreground proposals. Accordingly, we introduce feature orthogonality to decouple objectness from classification, shielding class semantics from corrupting gradients. Additionally, we design a Prototype-based Exponential-Moving-Average (P-EMA) classifier update to provide stable supervision for novel classes, and employ background resampling to rebalance foreground-background proposals. Experiments demonstrate our method effectively mitigates these core biases, improving TTL-FSOD performance.**

***Index Terms*—few-shot learning, object detection, test-time learning**

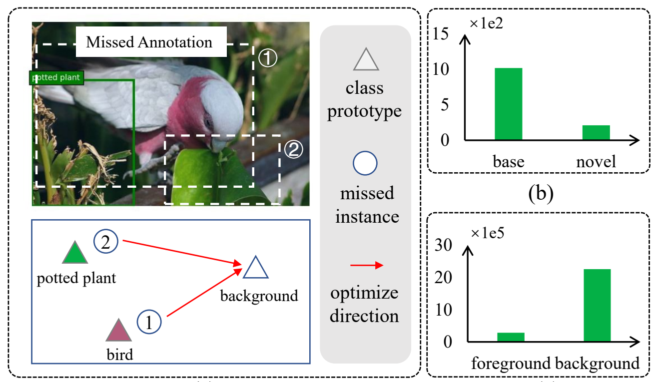



Fig. 1. (a) Illustration of semantic corruption from missed annotations. For PASCAL VOC Split-1 1-shot using our method, we report (b) average pseudo-sample number per class for base and novel classes, and (c) number of foreground and background proposals.

## I. Introduction

Few-Shot Object Detection (FSOD) [1]–[4] focuses on localizing and classifying objects from novel classes by leveraging abundant training samples from base classes alongside only a few annotated samples per novel class. However, conventional FSOD models operate statically, remaining fixed after deployment and incapable of learning from the ongoing stream of unlabeled data encountered during inference. This offline paradigm fundamentally limits their ability to adapt to real-world deployment scenarios and exploit potentially valuable test-time information.

To bridge this gap, Test-Time Learning for FSOD (TTL-FSOD) [5] investigates methods to dynamically adapt the model during inference. A recent work [5] establishes a baseline TTL-FSOD framework leveraging the teacher-student architecture prevalent in semi-supervised object detection [6]. In this paradigm, the pre-trained initial FSOD model is duplicated into a teacher and a student branch. The teacher generates pseudo-labels for unlabeled test data, which subsequently supervise the student's updates. Concurrently, the teacher's parameters are refined through an Exponential-Moving-Average (EMA) of the student's weights.

While promising, we find that the direct application of this paradigm to TTL-FSOD is undermined by the limited detection capability of the initial few-shot detector. This limitation triggers a set of biases arising from pseudo-label inaccuracies and the intrinsic sensitivity of standard detectors. Specifically, we identify three critical challenges: **(1) Semantic Corruption from Missed Annotations:** As illustrated in Fig. 1 (a), when a missed foreground object is mislabeled as background, the standard coupled optimization repels the feature from its true class prototype. Since this update direction is generally not aligned with the feature vector, it introduces tangential gradients that rotate the feature, corrupting the learned semantic direction. **(2) Long-Tailed Foreground Class Imbalance:** The generated pseudo-labels exhibit a severe long-tailed distribution. Well-trained base classes (head classes) produce numerous high-confidence pseudo-labels, while few-shot novel classes (tail classes) yield very few. As quantified in Fig. 1 (b), this imbalance causes head classes to generate dominating negative gradients for tail class classifiers, drowning out their

* *Corresponding Author*
*This work was supported in part by the National Natural Science Foundation of China (No.62301121, No.U23A20286), the Central Guiding Local Science and Technology Development Funding Project (No.2025ZYDF051), and the China Postdoctoral Science Foundation (No.2025M783502, No.GZB20240120).*

sparse positive signals and suppressing their learning [7]. **(3) Foreground-Background Proposal Imbalance:** Due to the teacher's limited recall, the number of foreground proposals is scarce. Consequently, standard sampling strategies incorporate an excessive number of background proposals to gather sufficient training data, creating a severe imbalance that destabilizes objectness learning, as shown in Fig. 1 (c).

In this paper, we propose a novel TTL-FSOD framework designed to specifically address these core issues. First, to resolve semantic corruption, we introduce the principle of feature orthogonality [8]–[10] into the learning objective. This strategy decouples objectness prediction from classification, using the feature vector's magnitude to determine the objectness score and its angle to predict the foreground class. Consequently, a missed detection labeled as background will only affect the feature's magnitude, shielding class semantics determined by the feature's angle from corrupting gradients. We demonstrate that this approach inherently mitigates the negative semantic bias induced by standard optimization objectives. Second, to address the long-tailed foreground class distribution, we propose a Prototype-based Exponential-Moving-Average (P-EMA) update for the classifier, augmented with a class-agnostic learnable prompt. This mechanism provides stable positive supervision for novel classes. Finally, we employ a simple background proposal resampling strategy to discard redundant background proposals, mitigating the foreground-background imbalance. Empirical evaluations validate the effectiveness of our proposed designs in resolving these specific biases within the test-time learning process.

## II. Related Works

### A. Few-Shot Object Detection

To learn from limited samples per novel class, detectors should absorb useful knowledge from massive base class samples. According to the strategy of inheriting knowledge, FSOD solutions can be classified into meta-learning methods and fine-tuning methods. Most state-of-the-art FSOD frameworks build upon Faster R-CNN [11], incorporating various adaptations for few-shot scenarios. For instance, DeFRCN [2] updates the detector backbone with ultralow learning rates to prevent overfitting. Recently, AD-UOFS [8] introduced feature orthogonality to decouple objectness prediction from classification in few-shot scenarios. In this work, we identify the specific value of this geometric property for test-time learning, demonstrating its capability to shield the model from the semantic corruption caused by missed pseudo-labels.

Researchers have observed that base class images often contain unlabeled novel class instances. Semi-supervised FSOD methods [12], [13] exploit these instances by generating pseudo-labels for novel class objects during training, primarily focusing on enhancing pseudo-label quality. LVC [12] rectifies pseudo-samples using external feature encoders and cascade box refinement, whereas APLDet [13] employs percentile threshold filtering to mitigate class imbalance. However, in TTL-FSOD with sequential data streams, percentile filtering becomes challenging to apply to balance pseudo-samples due to the one-pass data sequence. Moreover, efficient sample utilization becomes critical to accelerate convergence. Our proposed P-EMA update addresses these challenges from an optimization perspective with the given fixed pseudo-samples.

Some researchers studied how to reduce the interference of mislabeled foreground instances. Methods including gradient blocking [14], label smoothing [15], and loss reweighting [16], [17] have proven effective in mitigating the impact of these unannotated samples via filtering absent classes or gradient manipulation. However, they do not fully resolve the interference of such false negatives on semantic learning. In this paper, we conduct a systematic analysis of how gradient updates impact the semantic learning of foreground classes, and demonstrate that feature orthogonality can completely shield foreground semantic learning against interference from false negative samples.

### B. Test-Time Learning in FSOD

PS-TTL [5] is the first work to introduce TTL into FSOD. This approach establishes a semi-supervised baseline for online test streams, and further proposes a prototype-based knowledge distillation to leverage medium-confidence detections as soft-labeled samples. Although this mitigates class imbalance, the strategy remains constrained by detection bias: well-learned base classes still recall more soft-labeled samples, maintaining the imbalance. In contrast, our P-EMA classifier update offers a more stable and efficient mechanism that avoids class-specific negative gradients, demonstrating superior efficacy.

## III. Method

We follow PS-TTL to define the task of TTL-FSOD. The datasets consist of a base set $D_B$ with abundant samples for base classes, a few-shot novel set $D_N$ with $K$ annotated samples per novel class under the $K$-shot setting, and a test set $D_T$. One-epoch evaluation setting is adopted: an initial detector is first trained on $D_B$ and $D_N$, then undergoes a single-pass tuning stage on the test set $D_T$ without access to its labels. Finally, the adapted detector is evaluated on $D_T$ using its ground-truth labels. Our approach is designed to mitigate several critical biases that arise during this single-pass tuning phase. The overall framework shown in Fig. 2 is built upon a teacher-student structure, where an Exponential-Moving-Average updated teacher generates pseudo-labels for student supervision.

### A. Semantic Corruption from Missed Annotations

To specifically address semantic corruption, we adopt the Orthogonal Feature Space (OFS) [8]. In this section, we formally define the notation and analyze the gradient. Let $\boldsymbol{f} \in \mathbb{R}^d$ denote the feature vector of a proposal extracted by the detector, where $d$ is the feature dimension. Let $N_c$ be the number of foreground classes.

**Probability Construction.** In the OFS framework, objectness prediction is decoupled from foreground classification. The

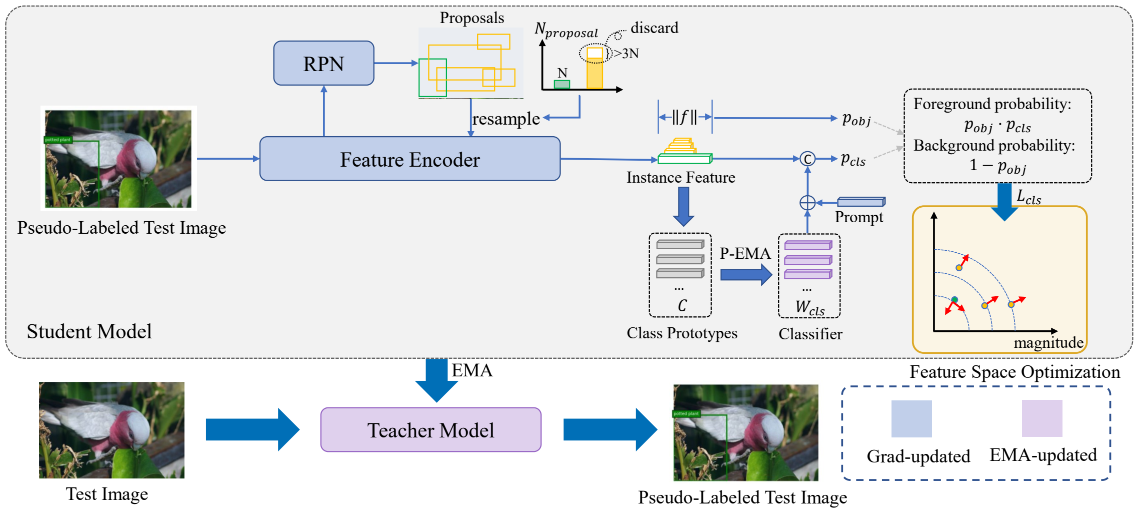



Fig. 2. The framework of our proposed method for TTL-FSOD. We inherit the network structure of AD-UOFS. For convenience, we omit structure details.

objectness score $p_{\text{obj}}$ is determined solely by the $L_2$-norm of the feature vector $\|\boldsymbol{f}\|_2$. The probability is modeled as:

$$p_{\text{obj}} = \sigma\left(-|w_{\text{obj}}| \cdot \|\boldsymbol{f}\|_2 + b_{\text{obj}}\right) \tag{1}$$

where $w_{\text{obj}}, b_{\text{obj}} \in \mathbb{R}$ are learnable scalars, and $\sigma(\cdot)$ denotes the Sigmoid function.

The foreground class probabilities $\boldsymbol{p}_{\text{cls}} \in \mathbb{R}^{N_c}$ are predicted based on the cosine similarity between the feature $\boldsymbol{f}$ and the learnable class prototypes (i.e., classifier weight vectors) $\boldsymbol{W}_{\text{cls}} \in \mathbb{R}^{N_c \times d}$. Let $\boldsymbol{w}^i_{\text{cls}} \in \mathbb{R}^d$ denote the $i$-th row of $\boldsymbol{W}_{\text{cls}}$ (i.e., the prototype for class $i$). The probability for the $i$-th foreground class is computed as:

$$p_{\text{cls},i} = \frac{\exp(\tau \cdot \cos(\boldsymbol{f}, \boldsymbol{w}^i_{\text{cls}}))}{\sum_{j=1}^{N_c} \exp(\tau \cdot \cos(\boldsymbol{f}, \boldsymbol{w}^j_{\text{cls}}))} \tag{2}$$

where $\cos(\boldsymbol{a}, \boldsymbol{b})$ is the cosine similarity, and $\tau$ is a temperature hyperparameter scaling the distribution. The final detection probability is $p_{\text{obj}} \cdot p_{\text{cls},i}$ for foreground class $i$, and $1 - p_{\text{obj}}$ for the background.

In contrast, the standard Euclidean Feature Space (EFS) computes probabilities via a unified linear projection followed by a global Softmax over $N_c + 1$ classes (including background). Let $\boldsymbol{W} \in \mathbb{R}^{(N_c+1)\times d}$ denote the unified prototype matrix, where the last row $\boldsymbol{w}_{N_c+1}$ corresponds to the background class. The probability for class $i$ is:

$$p_i = \frac{\exp(\boldsymbol{w}_i \cdot \boldsymbol{f})}{\sum_{j=1}^{N_c+1} \exp(\boldsymbol{w}_j \cdot \boldsymbol{f})} \tag{3}$$

We denote the background probability as $p_{\text{bg}} = p_{N_c+1}$, and similarly denote $\boldsymbol{w}_{\text{bg}} = \boldsymbol{w}_{N_c+1}$.

**Gradient Analysis.** Based on these formulations, we analyze the gradients for a missed annotation sample (a true foreground object incorrectly labeled as background). Intuitively, EFS corrupts semantics by pulling the feature towards the background, while OFS strictly preserves the semantic angle by only elongating the feature radially.

For EFS, minimizing the Cross-Entropy loss $\mathcal{L}^{\text{EFS}} = -\log(p_{\text{bg}})$ generates a parameter update direction (negative gradient) derived from Eq. (3):

$$-\nabla_{\boldsymbol{f}} \mathcal{L}^{\text{EFS}} = \underbrace{(1 - p_{\text{bg}})\boldsymbol{w}_{\text{bg}}}_{\text{Pull to Bg}} - \underbrace{\sum_{j \neq \text{bg}} p_j \boldsymbol{w}_j}_{\text{Push from Fg}} \tag{4}$$

This formula reveals two destructive forces: an attractive force from the background prototype $\boldsymbol{w}_{\text{bg}}$ pulling $\boldsymbol{f}$ towards the background cluster, and a repulsive force pushing it away from foreground semantics. Decomposing this update vector into radial and tangential components with respect to the feature direction $\boldsymbol{n} = \boldsymbol{f}/\|\boldsymbol{f}\|_2$:

$$-\nabla_{\boldsymbol{f}} \mathcal{L}^{\text{EFS}} = \underbrace{\left(-\nabla_{\boldsymbol{f}} \mathcal{L}^{\text{EFS}} \cdot \boldsymbol{n}\right) \boldsymbol{n}}_{\boldsymbol{g}_{\text{rad}}} + \underbrace{\left(-\nabla_{\boldsymbol{f}} \mathcal{L}^{\text{EFS}} - \boldsymbol{g}_{\text{rad}}\right)}_{\boldsymbol{g}_{\text{tan}}} \tag{5}$$

The update inevitably introduces a non-zero tangential component $\|\boldsymbol{g}_{\text{tan}}\| > 0$. This forces the feature vector to rotate, distorting the angular semantic information learned from foreground proposals.

In contrast, for OFS, the background supervision acts solely on the objectness branch (Eq. (1)). For a background sample, the loss is $\mathcal{L}^{\text{OFS}} = -\log(1 - p_{\text{obj}})$. Using the chain rule, the update direction is:

$$\begin{aligned} -\nabla_{\boldsymbol{f}} \mathcal{L}^{\text{OFS}} &= -\frac{\partial \mathcal{L}^{\text{OFS}}}{\partial p_{\text{obj}}} \cdot \frac{\partial p_{\text{obj}}}{\partial \|\boldsymbol{f}\|_2} \cdot \nabla_{\boldsymbol{f}} \|\boldsymbol{f}\|_2 \\ &= \underbrace{-\frac{1}{1 - p_{\text{obj}}} \cdot p_{\text{obj}}(1 - p_{\text{obj}}) \cdot (-|w_{\text{obj}}|)}_{\text{Scalar } \lambda = p_{\text{obj}}|w_{\text{obj}}|} \cdot \frac{\boldsymbol{f}}{\|\boldsymbol{f}\|_2} \\ &= \lambda \boldsymbol{n} \end{aligned} \tag{6}$$

This strictly collinear update direction implies that the tangential component is identically zero ($\boldsymbol{g}_{\text{tan}} \equiv \boldsymbol{0}$). Consequently, the feature vector undergoes only radial scaling (confidence suppression) without rotation, ensuring that the class-specific semantic direction remains intact despite the false negative supervision. Furthermore, since the background loss $\mathcal{L}^{\text{OFS}}$ depends only on $p_{\text{obj}}$ and is independent of the classifier weights $\boldsymbol{W}_{\text{cls}}$, the gradients of $\boldsymbol{W}_{\text{cls}}$ with respect to this loss are identically zero, providing protection at both the feature and classifier levels.

### *B. Long-tailed Foreground Bias*

Pseudo-labels generated during TTL exhibit a long-tailed distribution, where numerous base class samples suppress learning for sparse novel classes [7] by generating overwhelming negative gradients. Inspired by Momentum SGD optimizer, we propose a Prototype-based Exponential-Moving-Average (P-EMA) update to provide stable positive update signals. We maintain a ROI feature bank following [5], which is initialized with $D_N$ and updated during TTL with pseudo-labeled samples. The prototype $\boldsymbol{C}_i$ for each class is computed by averaging its stored ROI features. The classifier weights $\boldsymbol{w}^i_{\text{cls}}$ are then updated to align with these robust class representations:

$$\boldsymbol{w}^i_{\text{cls}} \leftarrow (1-\alpha)\boldsymbol{w}^i_{\text{cls}} + \alpha \text{Norm}(\boldsymbol{C}_i) \tag{7}$$

where $\alpha$ is an update ratio and $\text{Norm}(\cdot)$ denotes the magnitude scaling operator [21] that normalizes prototypes.

While P-EMA provides crucial positive supervision to enforce intra-class compactness, it lacks a mechanism for inter-class separation. To reintroduce discriminative learning with less class bias, we add a class-agnostic learnable prompt $\boldsymbol{\pi} \in \mathbb{R}^d$ to the classifier:

$$\boldsymbol{w}^{i\,\prime}_{\text{cls}} = \frac{\boldsymbol{w}^i_{\text{cls}}}{\|\boldsymbol{w}^i_{\text{cls}}\|} + \boldsymbol{\pi} \tag{8}$$

By adjusting the orientation of each classifier weight $\boldsymbol{w}^{i\,\prime}_{\text{cls}}$, the prompt $\boldsymbol{\pi}$ refines the angular decision boundaries between classes. This design offers two benefits: the shared prompt $\boldsymbol{\pi}$ is updated with the classification loss, capturing negative gradients in a class-agnostic way that prevents head classes from dominating tail classes. Simultaneously, as a dynamic residual component, it helps compensate for minor feature space shifts during fine-tuning.

### *C. Foreground-Background Proposal Bias*

Due to the teacher's limited recall, the number of foreground proposals is small. To gather sufficient training data, standard sampling strategies incorporate a large number of background proposals, creating a severe imbalance that destabilizes objectness learning. We mitigate this with a simple resampling strategy: for any image, if the number of background proposals exceeds three times the number of foreground proposals $N$, we randomly sample $3N$ background proposals and discard the rest.

## IV. Experiments

### *A. Implementation Details*

We adopt AD-UOFS [8] with a ResNet-101 backbone [22] as our baseline. Following [5], we remove the PCB [2] of the baseline. The P-EMA update ratio $\alpha$ is 1e-4, the pseudo-sample confidence threshold is 0.7, and the feature bank length is 512. The GCD ratio [2] is 0.01. For PASCAL VOC, the adaptation process utilizes only the unlabeled test stream $D_T$ for TTL, and $D_N$ is only utilized to initialize the feature memory bank. Evaluation uses the student model for PASCAL VOC and the teacher model for MS COCO due to the latter's greater complexity. Other settings are the same as PS-TTL.

For evaluation, we report the widely adopted metrics: novel class AP50 (nAP50) on PASCAL VOC [23], and novel class AP/AP75 (nAP/nAP75) on MS COCO [24].

### *B. Comparison Results*

**PASCAL VOC**. Table I presents our PASCAL VOC results. Our method outperforms non-TTL approaches by leveraging semantic information from unlabeled test images in general. Furthermore, it achieves superior performance compared to another TTL method PS-TTL, demonstrating the efficacy of our proposed bias mitigation strategy. Under Split-2, our method's performance is constrained by the detection capability of AD-UOFS baseline. While less competitive with advanced methods in this setting, our method still achieves competitive accuracy.

**MS COCO**. On the more complex MS COCO dataset, the initial detector's weaker performance creates more misclassified pseudo instances. Compared to the PASCAL VOC implementation, we implement the following modifications: (1) Maintain $D_N$ during TTL following PS-TTL. (2) Increase the teacher model's EMA momentum keep ratio from 0.9998 to 0.9999 and use the teacher model for performance evaluation. (3) Replace simple counting-based background resampling with the removal of background samples having objectness probability $p_{\text{obj}} > 0.5$ predicted by the student model to mitigate interference from undiscovered foregrounds. As shown in Table II, although baseline improvements are limited, our method still outperforms PS-TTL. This result demonstrates that our work's primary focus is on mitigating the optimization bias of the CE loss, rather than directly enhancing pseudo-label quality, which remains a key challenge on complex datasets.

### *C. Ablation Studies*

We conduct ablation studies on PASCAL VOC Split-1 to evaluate the impact of different components.

**Impact of semantic corruption.** Building on the feature orthogonality inherent in AD-UOFS, our approach effectively mitigates semantic corruption from mislabeled background samples during TTL. As Table III shows, applying TTL to the baseline consistently improves performance. To demonstrate the importance of angle-interference-free learning, we reintroduce the angular optimization for background proposals by adding a background prototype to construct an $(N_c+1)$-class probability distribution $\boldsymbol{p}_{\text{cls-bg}}$:

$$\boldsymbol{p}_{\text{cls-bg}} = \text{Softmax}\left(\tau \cdot \cos(\boldsymbol{f}, [\boldsymbol{W}_{\text{cls}}, \boldsymbol{W}_{\text{bg}}])\right) \tag{9}$$

TABLE I
COMPARISON RESULTS AMONG OUR METHOD AND OTHER FSOD METHODS ON PASCAL VOC. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD. *FOLLOWING PS-TTL, WE REMOVE THE PCB MODULE IN AD-UOFS AND OUR METHOD.

| Method | Novel Split1 | | | | | Novel Split2 | | | | | Novel Split3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 5 | 10 | 1 | 2 | 3 | 5 | 10 | 1 | 2 | 3 | 5 | 10 |
| TFA w/cos [1] | 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 |
| DeFRCN [2] | 57.0 | 58.6 | 64.3 | 67.8 | 67.0 | 35.8 | 42.7 | 51.0 | **54.5** | 52.9 | 52.5 | 56.6 | 55.8 | 60.7 | 62.5 |
| APLDet [13] | 61.3 | 58.1 | 61.9 | 65.3 | 67.5 | 40.6 | 39.2 | 43.1 | 50.0 | 55.0 | 43.8 | 53.3 | 48.6 | 53.3 | 58.1 |
| FPD [18] | 46.5 | 62.3 | 65.4 | 68.2 | **69.3** | 32.2 | 43.6 | 50.3 | 52.5 | **56.1** | 43.2 | 53.3 | 56.7 | 62.1 | 64.1 |
| UNP [16] | 43.7 | 58.3 | 59.8 | 63.7 | 64.2 | 28.1 | 42.8 | 47.7 | 49.5 | 50.3 | 38.4 | 49.3 | 53.8 | 57.7 | 58.7 |
| MPFSOD [19] | 61.7 | 63.4 | 64.2 | 68.5 | **69.3** | 41.9 | 42.2 | 42.6 | 50.6 | 53.1 | 55.0 | 56.4 | 57.6 | 61.8 | 62.7 |
| DeFRCN+FR [20] | 61.2 | 65.8 | 67.1 | 67.6 | 67.9 | 36.4 | 46.5 | 49.5 | 53.8 | 54.2 | 53.5 | 56.4 | 57.2 | 60.3 | 62.1 |
| PS-TTL* [5] | 58.4 | 65.7 | 67.9 | **69.3** | 68.1 | 38.4 | 47.8 | **52.8** | 53.6 | 49.1 | 53.0 | 58.8 | 59.2 | 63.8 | 64.1 |
| AD-UOFS* [8] | 62.1 | 64.6 | 64.4 | 68.3 | 66.4 | 43.8 | 46.1 | 51.1 | 49.9 | 48.1 | 59.1 | 60.5 | 60.0 | 62.7 | 62.4 |
| AD-UOFS*+Ours | **67.3** | **68.2** | **68.5** | 68.5 | 68.4 | **48.2** | **49.2** | 51.2 | 48.4 | 50.5 | **63.1** | **63.9** | **63.1** | **64.8** | **64.6** |

TABLE II
COMPARISON RESULTS AMONG OUR METHOD AND OTHER FSOD METHODS ON MS COCO. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD. *FOLLOWING PS-TTL, WE REMOVE THE PCB MODULE IN AD-UOFS AND OUR METHOD.

| Method | 10-shot | | 30-shot | |
|---|---|---|---|---|
| | nAP | nAP75 | nAP | nAP75 |
| TFA w/cos [1] | 10.0 | 9.3 | 13.7 | 13.4 |
| DeFRCN [2] | 18.6 | 17.6 | 22.5 | 22.3 |
| LVC [12] | 17.8 | 17.8 | **24.5** | **25.0** |
| FPD [18] | 16.5 | – | 20.1 | – |
| UNP [16] | 12.3 | 11.5 | 15.3 | 14.8 |
| PS-TTL* [5] | 17.3 | 16.7 | 20.9 | 21.3 |
| AD-UOFS* [8] | 19.1 | 19.1 | 21.5 | 21.7 |
| AD-UOFS*+Ours | **19.5** | **19.8** | 21.6 | 21.7 |

TABLE III
IMPACT OF BACKGROUND ANGULAR LOSS. THE WEIGHT OF $\mathcal{L}_{BA}$ IS 1.

| Baseline | TTL | $\mathcal{L}_{ba}$ | 1-shot | 2-shot | 3-shot |
|---|---|---|---|---|---|
| ✓ | | | 62.1 | 64.6 | 64.4 |
| | ✓ | | 62.8 | 65.8 | 67.5 |
| | ✓ | ✓ | 63.0 | 65.3 | 66.5 |

where $[\boldsymbol{W}_{\text{cls}}, \boldsymbol{W}_{\text{bg}}]$ denotes the concatenation of foreground prototypes and the background prototype along the class dimension. We apply the Cross-Entropy loss on this distribution using the background label, and term this auxiliary supervision as Background Angular Loss ($\mathcal{L}_{\text{ba}}$). Results indicate that introducing $\mathcal{L}_{\text{ba}}$ causes performance degradation in general. This degradation occurs because while feature orthogonality mitigates missing-label issues by strictly limiting updates to the radial direction, optimizing with $\mathcal{L}_{\text{ba}}$ actively disrupts semantic feature learning. It introduces harmful tangential gradients that force the features of missed foreground objects to rotate away from their true class prototypes.

**Impact of classifier update method**. Table IV compares different classifier update strategies. While Vanilla Fine-Tuning (Vanilla FT) with standard loss gradients improves baseline performance, our P-EMA achieves superior results. These benefits derive from two mechanisms: first, it balances updates across foreground classes to alleviate long-tailed distribution effects in pseudo-labeled samples; second, it accelerates classifier convergence while minimizing disruptive negative gradients. Although effective, this method's limited discriminative learning constrains performance. Joint optimization with both Vanilla FT and P-EMA further enhances accuracy. Here, Vanilla FT introduces discriminative learning via negative gradient updates, while P-EMA provides stable positive updates that counteract class imbalance inherent in the fine-tuning process. Additionally, the prompt-based update mechanism refines the P-EMA updated classifier in a class-agnostic way, enabling further performance gains.

TABLE IV
IMPACT OF CLASSIFIER UPDATING METHOD.

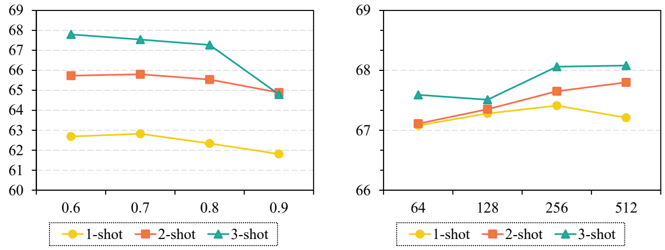
| Vanilla FT | P-EMA | Prompt | 1-shot | 2-shot | 3-shot |
|---|---|---|---|---|---|
| | | | 62.1 | 64.6 | 64.4 |
| ✓ | | | 62.8 | 65.8 | 67.5 |
| | ✓ | | 66.4 | 67.3 | 67.5 |
| ✓ | ✓ | | 66.8 | 67.5 | 68.0 |
| | ✓ | ✓ | 67.2 | 67.8 | 68.1 |

TABLE V
IMPACT OF FOREGROUND-BACKGROUND RESAMPLE.

| Resample | 1-shot | 2-shot | 3-shot |
|---|---|---|---|
| | 67.2 | 67.8 | 68.1 |
| ✓ | 67.3 | 68.2 | 68.5 |

**Impact of foreground-background resample**. Foreground-background imbalance creates an additional imbalance distinct from foreground class distribution imbalance. Given the redundancy of background samples, we implement random downsampling. Table V demonstrates that this straightforward numerical resampling strategy consistently enhances performance.

**Hyperparameter sensitivity analysis**. We ablate critical hyperparameters in Fig. 3. For the confidence filter threshold, higher values retain fewer pseudo-labeled samples with higher reliability, while lower thresholds increase sample quantity at the cost of introducing noise. Evaluated on our CE-updated TTL baseline, Fig. 3 (a) shows that thresholds of 0.6/0.7 yield optimal performance. Regarding feature memory bank length (Fig. 3 (b)) in P-EMA, storing 256/512 features per foreground class achieves competitive accuracy, whereas reducing capacity significantly harms performance.




(a) (b)

Fig. 3. (a) Impact of filter threshold. (b) Impact of the feature memory bank length.

## V. Discussion

Our design restricts angular optimization to foreground proposals, entailing two trade-offs. First, in standard EFS, background proposals help compact foreground distributions and spread background features uniformly across classes. Removing this may reduce boundary compactness and cause the model to misclassify undefined class instances as defined foreground classes. However, under noisy pseudo-labels, angular optimization corrupts missed foreground semantics, making semantic protection the more robust choice. Second, when a false negative occurs, OFS only suppresses the feature magnitude, preserving classification ability but inevitably reducing the objectness score of the missed instance. Furthermore, our approach inherently relies on the initial detector's capability. As observed on MS COCO, performance gains are bounded by the baseline's initial performance.

## VI. Conclusion

In this paper, we identify and address three critical biases in Test-Time Learning for Few-Shot Object Detection (TTL-FSOD) that arise from pseudo-labeling noise: semantic corruption, long-tailed class imbalance, and foreground-background proposal imbalance. To overcome these challenges, we introduce a novel framework that uses feature orthogonality to shield class semantics, a Prototype-based Exponential-Moving-Average (P-EMA) classifier update to provide stable supervision, and background resampling to rebalance proposals. Experiments demonstrate that our method effectively mitigates these core biases.

## Acknowledgements

The authors used Gemini to refine the language and grammar of this manuscript.

## References

[1] X. Wang, T. Huang, J. Gonzalez, T. Darrell, and F. Yu, "Frustratingly simple few-shot object detection," in *Proceedings of the International Conference on Machine Learning*, 2020, pp. 9919–9928.

[2] L. Qiao, Y. Zhao, Z. Li, X. Qiu, J. Wu, and C. Zhang, "Defrcn: Decoupled faster r-cnn for few-shot object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 8681–8690.

[3] B. Sun, B. Li, S. Cai, Y. Yuan, and C. Zhang, "Fsce: Few-shot object detection via contrastive proposal encoding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7352–7362.

[4] H. Mei, T. Zhao, S. Tang, H. Qiu, L. Wang, M. Zhang, F. Meng, and H. Li, "Grsdet: Learning to generate local reverse samples for few-shot object detection," *Neurocomputing*, p. 130903, 2025.

[5] Y. Gao, Y. Zhang, Z. Huang, N. Liu, and D. Huang, "Ps-ttl: Prototype-based soft-labels and test-time learning for few-shot object detection," in *Proceedings of the ACM International Conference on Multimedia*, 2024, pp. 8691–8700.

[6] Y.-C. Liu, C.-Y. Ma, Z. He, C.-W. Kuo, K. Chen, P. Zhang, B. Wu, Z. Kira, and P. Vajda, "Unbiased teacher for semi-supervised object detection," *Proceedings of the International Conference on Learning Representations*, 2021.

[7] J. Wang, W. Zhang, Y. Zang, Y. Cao, J. Pang, T. Gong, K. Chen, Z. Liu, C. C. Loy, and D. Lin, "Seesaw loss for long-tailed instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9695–9704.

[8] T. Zhao, H. Qiu, Y. Dai, L. Wang, F. Meng, Q. Wu, and H. Li, "Attention-disentangled uniform orthogonal feature space optimization for few-shot object detection," *arXiv preprint arXiv:2506.22161*, 2025.

[9] Z. Sun, J. Li, and Y. Mu, "Exploring orthogonality in open world object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2024, pp. 17 302–17 312.

[10] D. Chen, S. Zhang, J. Yang, and B. Schiele, "Norm-aware embedding for efficient person search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 615–12 624.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[12] P. Kaul, W. Xie, and A. Zisserman, "Label, verify, correct: A simple few shot object detection method," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14 237–14 247.

[13] Y. Tang, Z. Cao, Y. Yang, J. Liu, and J. Yu, "Semi-supervised few-shot object detection via adaptive pseudo labeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2151–2165, 2024.

[14] B.-B. Gao, X. Chen, Z. Huang, C. Nie, J. Liu, J. Lai, G. Jiang, X. Wang, and C. Wang, "Decoupling classifier for boosting few-shot object detection and instance segmentation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 18 640–18 652, 2022.

[15] B. Qiao, H. Zhou, L. Yang, and X. Xie, "Few shot object detection with incompletely annotated samples," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2023, pp. 1–7.

[16] B. Yan, C. Lang, G. Cheng, and J. Han, "Understanding negative proposals in generic few-shot object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[17] M. J. Jung, S. D. Han, and J. Kim, "Re-scoring using image-language similarity for few-shot object detection," *Computer Vision and Image Understanding*, vol. 241, p. 103956, 2024.

[18] Z. Wang, B. Yang, H. Yue, and Z. Ma, "Fine-grained prototypes distillation for few-shot object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 6, 2024, pp. 5859–5866.

[19] J. Ma and S. Bai, "Multi-view part-based few-shot object detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 8, pp. 14 749–14 763, 2025.

[20] Y. Zhang, J. Li, Q. Ji, K. Li, L. Liu, C. Zheng, and W. Qiang, "Intervening on few-shot object detection based on the front-door criterion," *Neural Networks*, vol. 185, p. 107251, 2025.

[21] Z. Yang, C. Zhang, R. Li, Y. Xu, and G. Lin, "Efficient few-shot object detection via knowledge inheritance," *IEEE Transactions on Image Processing*, vol. 32, pp. 321–334, 2022.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[23] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, pp. 303–338, 2010.

[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Proceedings of the European Conference on Computer Vision*, 2014, pp. 740–755.