

Variance-Entropy Guided Fine-Grained Prototypes Distillation for Robust Few-Shot Object Detection

Mohammad Hussein Adib¹ and Mohammad javad Fadaeieslam¹*

Abstract—Few-shot object detection (FSOD) aims to enable models to recognize novel objects from minimal training examples; however, the task faces significant hurdles due to unstable feature learning and inherently biased class prototypes when training data is severely limited. Standard detectors often struggle to generalize in these scenarios, as they frequently overfit to the limited support set, leading to poor intra-class compactness. To mitigate these issues, we propose a novel Fine-Grained Prototype Distillation (FPD) framework that utilizes a unique variance-entropy guided refinement strategy. Our approach introduces two core components: first, a variance-aware module for Prototype Distillation (PD) that enhances prototype discriminability through adaptive intra-class dispersion weighting; and second, an entropy-based mechanism for Prototype Assignment (PA) that effectively suppresses low-confidence support samples during training. By explicitly accounting for the statistical reliability of support features, FPD fosters more stable and representative prototype construction, effectively bridging the critical gap between noisy, limited support instances and robust, reliable class representations. Extensive experimental results on the PASCAL VOC benchmark demonstrate that our method consistently achieves superior average precision (AP) compared to state-of-the-art FSOD models in both 1-shot and 5-shot scenarios. These results confirm the effectiveness of our framework, providing a more robust foundation for accurate category recognition in extreme, data-scarce regimes.

Index Terms: Meta-learning, Classification Confidence, Prototype Refinement, Similarity weighting

I. INTRODUCTION

Object detection is a cornerstone of computer vision, requiring both the classification of object categories and the precise spatial localization of objects through bounding boxes. Progress in this domain is essential for the evolution of intelligent systems, including autonomous driving, surveillance, medical diagnostics, and smart city infrastructure [1]. As these technologies move from controlled environments into complex, real-world deployments, the demand for robust and accurate object detection capabilities becomes increasingly critical. Historically, object detection relied on handcrafted features and traditional machine learning paradigms, which often struggled to generalize in complex and noisy visual environments. The emergence of deep learning has fundamentally transformed this field, with convolutional neural networks (CNNs) and, more recently, Vision Transformers (ViTs) demonstrating strong capabilities in learning hierarchical feature representations. These advances have significantly improved model accuracy, robustness, and adaptability across diverse scenarios [2]-[3].

Modern object detection frameworks are generally categorized into three main architectural paradigms: two-stage detectors, which generate region proposals followed by classification and bounding-box refinement (e.g., the R-CNN family); one-stage detectors, which prioritize inference speed by predicting object classes and locations in a single pass (e.g.,

Received; 2026-05-03 Revised; 2026-07-07 Accept; 2026-08-17

1. Faculty of Electrical & Computer Engineering, Semnan University, Semnan, Iran.

*Corresponding author Email: fadaei@semnan.ac.ir

Cite this article as:

Adib, M. H., and Fadaeieslam, M. J. (2026) Variance-Entropy Guided Fine-Grained Prototypes Distillation for Robust Few-Shot Object Detection. *Journal of Modeling & Simulation in Electrical & Electronics Engineering (MSEEE)*. Semnan University Press. 6 (3), 1-9.
DOI: <https://doi.org/10.22075/mseee.2026.41071.1270>

YOLO and SSD); and transformer-based detectors, which exploit self-attention mechanisms to achieve competitive end-to-end detection performance.

Despite these advancements, the performance of object detectors remains limited in data-scarce scenarios, where annotating large-scale datasets for every new category is often computationally expensive or practically infeasible. Few-shot learning (FSL) has emerged as a promising paradigm to address this challenge by enabling models to generalize to novel classes from only a small number of annotated examples [4]. In this context, Few-Shot Object Detection (FSOD) extends the principles of FSL to detection tasks, requiring models to simultaneously predict object categories and their spatial locations using limited support information.

Existing FSOD approaches mainly rely on meta-learning, metric learning, or attention-based mechanisms to bridge the knowledge gap between base and novel classes [5]–[7]. However, these approaches often suffer from instability due to the inherent uncertainty present in support features. In particular, many prototype-based methods construct coarse and overly generalized class prototypes that fail to capture fine-grained semantic distinctions. As a result, the learned class representations may become biased, leading to suboptimal feature matching between support and query samples. This sensitivity to noise and uncertainty in the support set ultimately reduces detection robustness and overall performance.

To address these limitations, we propose an enhanced version of the Fine-Grained Prototype Distillation (FPD) framework by incorporating uncertainty-aware mechanisms to improve the reliability of prototype-based learning. Although the original FPD framework provides a strong foundation for few-shot tasks, its reliance on conventional feature aggregation can lead to feature instability under extreme data scarcity.

Our approach extends the framework by introducing two key components: a variance-aware module within the Prototype Distillation (PD) process and an entropy-based mechanism for Prototype Assignment (PA). The variance-aware module improves prototype discriminability by dynamically weighting support features according to their intra-class variance. In parallel, the entropy-based mechanism reduces the influence of ambiguous or low-confidence support samples through entropy regularization.

By integrating these uncertainty-aware strategies, the proposed framework enables the construction of more reliable and discriminative class prototypes, ultimately enhancing the robustness and stability of the classification process in few-shot scenarios.

The main contributions of this work can be summarized as follows:

- **Uncertainty-Aware Extension of FPD:** We extend the Fine-Grained Prototype Distillation (FPD) framework by incorporating uncertainty-aware mechanisms designed to alleviate the effects of feature noise and ambiguity in few-shot scenarios.
- **Variance-Guided Prototype Distillation:** We introduce a variance-guided weighted distillation strategy that leverages intra-class feature variance to construct more stable and

representative class prototypes. By assigning greater importance to low-variance features, the proposed method effectively suppresses the influence of noisy or unstable regions in the support set.

- **Entropy-Based Prototype Assignment:** We propose an entropy-based assignment mechanism that utilizes the information entropy of similarity scores to guide query–prototype matching. This strategy emphasizes high-confidence regions while reducing the contribution of ambiguous background areas, resulting in more accurate object classification.

- **Refined Fine-Grained Feature Aggregation:** We further refine the Fine-Grained Feature Aggregation (FFA) module within the FPD architecture to effectively incorporate the proposed variance- and entropy-driven guidance, enabling tighter integration between prototype distillation and feature matching.

II. RELATED WORKS

The field of Few-Shot Object Detection (FSOD) has undergone significant evolution, progressing from general meta-learning frameworks to more specialized prototype-based approaches. This section reviews the key methodologies that have shaped the current research landscape.

Early advancements in FSOD were largely driven by meta-learning paradigms, with Meta R-CNN [10] serving as a foundational approach. Meta R-CNN adopts a Siamese-style architecture in which a support branch and a query branch share a common backbone for feature extraction. A meta-learner aggregates support features to form class-level prototypes, which are subsequently fused with query features to guide the detection head. While this framework established an effective mechanism for transferring knowledge from base classes to novel classes, it exhibits limitations in capturing fine-grained spatial information. In particular, reliance on coarse class-level feature aggregation can lead to the loss of local semantic details that are essential for accurately localizing novel objects in complex scenes.

To enhance feature discriminability, several approaches have incorporated metric-learning strategies or attention mechanisms. Transformer-based architectures, such as DETR [1], have significantly influenced modern object detection by reformulating the task as a set prediction problem using global self-attention. Although DETR effectively models long-range dependencies, its direct application to few-shot scenarios remains challenging. Under extreme data scarcity, global attention mechanisms may struggle to accurately localize novel objects without sufficient local guidance derived from reliable support features.

More recent research has shifted toward prototype-based FSOD methods that aim to bridge the gap between base and novel classes through representative class prototypes. The Fine-Grained Prototype Distillation (FPD) framework [4] represents a notable advancement in this direction. Within FPD, the Fine-Grained Feature Aggregation (FFA) module distills support features into detailed prototypes, while additional components such as Balanced Class-Agnostic Sampling (B-CAS) and Non-Linear Fusion (NLF) enhance feature matching between support and query samples. Similarly, approaches such as VFA

attempt to reduce bias through class-agnostic sampling strategies. However, these prototype-based methods, including the original FPD framework, typically assume that support features are inherently reliable. In practice, they remain vulnerable to the uncertainty problem, where noisy or ambiguous regions in the support set lead to biased prototype representations. This issue can significantly degrade both classification accuracy and localization precision in highly data-constrained environments.

III. METHODOLOGY

A. The Overall Architecture

The proposed framework is built upon the Siamese architecture of Meta R-CNN [10], which is designed to facilitate knowledge transfer from a support set to a query set in few-shot object detection (FSOD) tasks. As illustrated in Fig. 1, the architecture consists of two parallel branches—a support branch and a query branch—that share a common backbone to ensure feature consistency.

Feature extraction is performed using a ResNet-50/101 backbone [1]. Mid-level features are obtained from the first three stages of the network and subsequently processed by the Fine-Grained Feature Aggregation (FFA) module. This module aggregates support features to form fine-grained prototypes, enabling more effective interaction between the support and query branches.

Following this stage, the fourth stage of the backbone is used to extract high-level semantic representations, producing Region of Interest (RoI) features from the query branch and class-level prototypes from the support branch. These representations are then integrated through a Non-Linear Fusion (NLF) module before being forwarded to the detection head for final classification and localization.

Importantly, the Region Proposal Network (RPN) operates on query features that have already interacted with support-derived information, which enhances the model's ability to generate proposals and accurately localize novel object instances.

B. The FPD Framework

To achieve robust few-shot detection, our framework builds upon the Fine-Grained Prototype Distillation (FPD) framework [2], which improves the alignment between support and query feature distributions through a set of specialized modules.

1) Fine-Grained Feature Aggregation (FFA)

Instead of directly matching raw support and query features, the FFA module first distills support features into fine-grained prototypes. These prototypes are then aligned with the query feature maps and injected according to their similarity scores. This process enables the model to better distinguish foreground objects from background regions while simultaneously capturing similarities and distinctions among different classes. In the following subsections, we describe the prototype distillation and assignment processes in detail. We also discuss the extension of the framework to novel classes and the fusion of multi-shot prototypes during inference. An overview of the FFA module is illustrated in Fig. 2.

2) Balanced Class-Agnostic Sampling (B-CAS)

To mitigate the inherent bias present in standard meta-learning approaches—such as Meta R-CNN, which relies primarily on class-specific matching—the B-CAS module introduces a class-agnostic sampling strategy. While previous methods, such as VFA [3], attempt to reduce bias through random sampling, this stochastic process can introduce additional noise into the training procedure. In contrast, B-CAS provides a more controlled sampling mechanism that enables the model to learn robust and generalized features without being adversely influenced by irrelevant or non-informative class-agnostic samples.

3) Non-Linear Fusion (NLF)

To effectively bridge the semantic gap between support prototypes and query features, we introduce the Non-Linear Fusion (NLF) module. Unlike conventional meta-learning frameworks that rely on simple element-wise multiplication or linear summation—operations that often fail to capture complex, non-linear cross-class relationships—the proposed NLF module adopts a multi-path transformation strategy. This design enables the model to capture diverse relational cues through multiplicative, subtractive, and concatenative interactions.

Formally, let $f_{roi} \in \mathbb{R}^{C \times H \times W}$ denote the feature map extracted for a query Region of Interest (RoI), and let $p_{cls} \in \mathbb{R}^{C \times 1 \times 1}$ represent the distilled class prototype for the target category. To capture higher-order correlations between these representations, we first construct an augmented feature representation f' by aggregating multiple interaction paths:

$$f' = [F_1(f_{roi} \odot p_{cls}), F_2(f_{roi} - p_{cls}), F_3[f_{roi}, p_{cls}], f_{roi}] \quad (1)$$

$$f = F_{agg}(f') \quad (2)$$

where $\odot$ denotes the Hadamard (element-wise) product, capturing similarity-based correlations between the query feature and the prototype. The subtraction operator $-$ models the residual difference between the two representations. The functions F_1, F_2, F_3 represent independent fully-connected layers followed by the ReLU activation function. Finally, F_{agg} integrates the concatenated features to produce the fused representation f , which is subsequently used for classification and localization in the detection head.

C. Proposed Uncertainty-Aware Enhancements

While the FPD framework provides a strong foundation for few-shot object detection, it implicitly assumes that support features are clean and fully representative of the target class. In practice, however, support sets in few-shot scenarios often contain noisy or ambiguous regions—such as background clutter or unclear object boundaries—which may result in polluted or unreliable prototypes.

To address this limitation, we introduce two uncertainty-aware mechanisms: Weighted Variance-Based Prototype Distillation and Entropy-Based Prototype Assignment, which aim to improve the reliability and discriminative quality of the learned prototypes.

1) Weighted Variance-Based Prototype Distillation

To mitigate the influence of noisy or non-representative features during prototype construction, we introduce a variance-based prioritization mechanism. This module aims to differentiate stable semantic features from transient spatial noise, thereby ensuring more reliable prototype representations even in extreme few-shot settings, such as the 1-shot case.

In the 1-shot setting, we disable the variance-based weighting component and rely exclusively on the entropy-guided prototype assignment mechanism. This design ensures mathematical consistency across all shot configurations while preserving the effectiveness of the proposed uncertainty-aware framework under conditions of extreme data scarcity.

Formally, let $F \in \mathbb{R}^{C \times H \times W}$ denote the feature map extracted from a single support instance. For each channel $c \in \{1, \dots, C\}$ the spatial variance σ_c^2 is computed as:

$$\sigma_c^2 = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (f_{c,i,j} - \mu_c)^2 \quad (3)$$

where $f_{c,i,j}$ denotes the activation at the i -th row and j -th column of the c -th channel, and μ_c represents the mean activation of that channel:

$$\mu_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_{c,i,j} \quad (4)$$

To emphasize highly discriminative features and suppress those characterized by high spatial instability (noise), we derive an

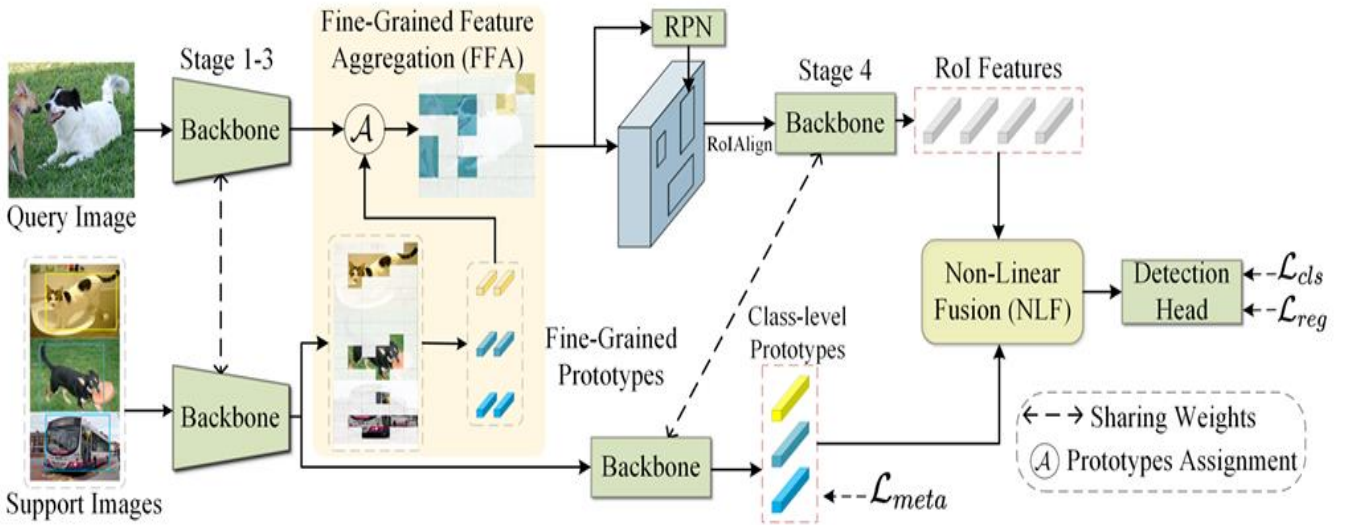


Fig 1 The FPD model architecture [2]

importance weight ω_c for each channel using an inverse-variance weighting scheme:

$$\omega_c = \frac{1}{\sigma_c^2 + \varepsilon} \quad (5)$$

where $\varepsilon > 0$ is a small constant for numerical stability. The weighted prototype $v_{weighted}$ is then obtained by aggregating the channel-wise features using these weights:

$$v_{weighted} = \frac{\sum_{c=1}^C \omega_c \cdot p_c}{\sum_{c=1}^C \omega_c} \quad (6)$$

Finally, to preserve the integrity of the original feature distribution while incorporating the stability-aware information, the final refined prototype v_{final} is formulated as a convex combination:

$$v_{final} = \alpha v_{original} + (1 - \alpha) v_{weighted} \quad (7)$$

where $v_{original}$ is the original prototype, and $\alpha \in [0,1]$ is a hyperparameter that balances the contribution between the original feature representation and the stability-weighted representation. This mechanism ensures that the resulting

prototype is robust against localized spatial noise and irregular feature activations. For our work, we set α to 0.8.

2) Entropy-based Prototype Assignment:

A critical challenge in the assignment stage of few-shot object detection (FSOD) lies in accurately associating query regions with their corresponding class prototypes. Approaches that rely solely on local similarity measures often produce false positives in background regions or areas with low semantic confidence.

To address this issue, we introduce an uncertainty-aware assignment mechanism that leverages Shannon entropy to adaptively regulate the contribution of query features according to their classification confidence.

Initially, we compute the similarity score between the query feature map $q \in \mathbb{R}^{H \times W \times d}$ and the set of fine-grained prototypes $K = \{k_1, k_2, \dots, k_{n_c}\} \subset \mathbb{R}^d$ using cosine similarity. For each spatial location (i, j) , the similarity with the c -th prototype is defined as:

$$sim_{i,j,c} = \frac{q_{i,j} \cdot k_c}{\|q_{i,j}\| \|k_c\|} \quad (8)$$

To quantify the uncertainty of the assignment at each spatial location, we transform the similarity scores into a probability distribution via a Softmax function. Let $p_{i,j,c}$ denote the probability that the feature at location (i, j) belongs to class c :

$$p_{i,j,c} = \frac{\exp(\text{sim}_{i,j,c})}{\sum_{m=1}^{n_c} \exp(\text{sim}_{i,j,m})} \quad (9)$$

The predictive uncertainty is then quantified using the Shannon entropy $H_{i,j}$ at each location:

$$H_{i,j} = - \sum_{c=1}^{n_c} p_{i,j,c} \log(p_{i,j,c}) \quad (10)$$

A low entropy indicates a highly confident, unimodal assignment, whereas a high entropy suggests semantic ambiguity or noise. To prioritize high-confidence regions and suppress uncertain activations, we compute a spatial scaling factor $\gamma_{i,j}$ based on the exponential decay of the entropy:

$$\gamma_{i,j} = \exp(-H_{i,j}) \quad (11)$$

Finally, the refined similarity map S' , which serves as the guiding signal for the subsequent detection head, is obtained by applying the scaling factor to the original similarity scores:

$$S'_{i,j,c} = \gamma_{i,j} \cdot \text{sim}_{i,j,c} \quad (12)$$

By integrating this uncertainty-aware modulation, our framework effectively suppresses spurious responses in low-confidence regions, thereby enhancing the robustness and reliability of the few-shot object detection process.

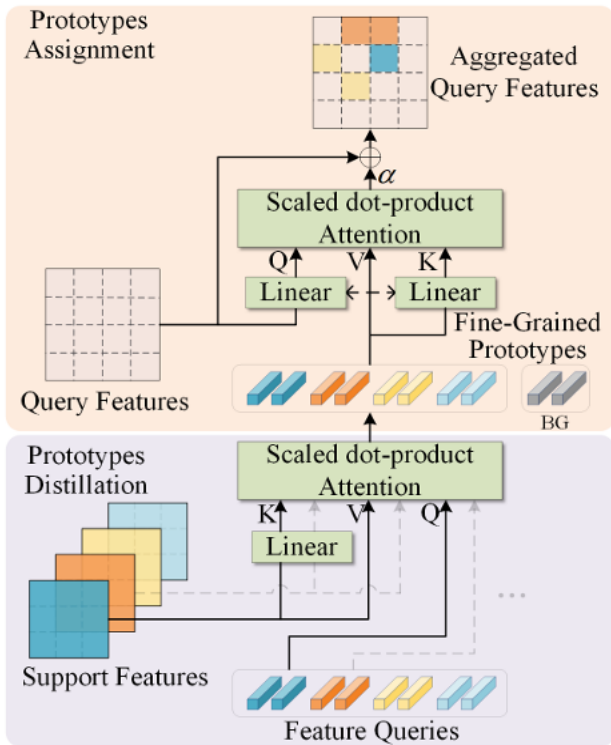


Fig. 2 FFA module [2]

D. The Model Architecture

As illustrated in Fig. 1, our framework extends Meta R-CNN, which adopts a Siamese architecture consisting of query and support branches that share a common backbone. Mid-level features from both the query and support images are extracted using the first three stages of ResNet-50/101 [1]. These features are then processed by the Fine-Grained Feature Aggregation (FFA) module, which distills support features into fine-grained prototypes and injects them into the query branch.

Subsequently, the final stage (Stage 4) of the backbone extracts high-level semantic features, producing Region of Interest (RoI) features from the query branch and class prototypes from the support branch. These representations are further refined through the Non-Linear Fusion (NLF) module before being forwarded to the detection head for final predictions.

Importantly, the Region Proposal Network (RPN) operates on query features that have already been enhanced through interaction with support information, thereby improving the model's ability to localize novel objects.

IV. EXPERIMENTS

We evaluate our method on a widely used few-shot object detection (FSOD) benchmark: PASCAL VOC.

E. Dataset

We evaluate the performance of the proposed method on the widely recognized PASCAL VOC benchmark for Few shot object classification. The dataset comprises 20 object categories, partitioned into 15 base classes and 5 novel classes. To ensure a rigorous evaluation, three distinct class partitions (splits) are utilized. For the training phase,

the training and validation set from VOC2007 and VOC2012 are combined, while the VOC2007 test set is strictly reserved for evaluation. Performance is measured using the mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 (AP50) across various few-shot settings, specifically $K \in \{1, 2, 3, 5, 10\}$ shots per novel class.

F. Implementation Details

The proposed framework is implemented using the MMDetection [4] toolbox. We employ a ResNet-101 backbone [1], pre-trained on ImageNet [5], for feature extraction. Following the architectural configuration of the baseline FPD framework, a single-scale feature map is utilized for detection without the integration of a Feature Pyramid Network (FPN) [6]. For the input processing, query images are resized to a maximum resolution of 1333×800 pixels, while support instances are resized to 224×224 pixels. The training is conducted on NVIDIA P100 GPUs with a total batch size of 8, utilizing the Stochastic Gradient Descent (SGD) optimizer. During the base training stage on the PASCAL VOC dataset, the model undergoes 20k iterations with an initial learning rate of 0.004, which is decayed to 0.0004 at the 17k iteration mark. In the subsequent fine-tuning stage, the learning rate is set to 0.001. The loss functions employed are identical to those used in the Meta R-CNN [7] framework.

G. Comparison with the State-of-the-Art Methods

We compare our method against several state-of-the-art (SOTA) FSOD approaches, including FSDetView [8], DCNet [9], Meta-DETR, DeFR CN, FCT, VFA, and FPD, across three standard PASCAL VOC splits for $K \in \{1, 2, 3, 5, 10\}$ shot settings. Results are reported as the mean $\pm$ standard deviation over 5 independent runs. As presented in Table I, our method achieves competitive or superior performance in most configurations. Notably, in several key scenarios, the proposed

model demonstrates significant improvements. For instance, on Novel Set 1, our approach achieves an AP50 of 43.4% in the 1-shot setting and 57.1% in the 2-shot setting, outperforming the most competitive prior method by up to 3.2%. While there are specific instances where other models perform similarly (e.g., in certain 1-shot scenarios in Novel Set 2), our approach demonstrates strong overall robustness and scalability. The performance comparison between our model and the leading model across different shot settings is visualized in Fig. 3.

TABLE 1
FSOD (AP50) Results on three Parts of the Pascal VOC Dataset as the Mean $\pm$ Standard over 5 Independent Runs

Method/ shot	Novel Set 1					Novel Set 2					Novel Set 3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
FSDetView	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
DCNet	33.9	37.4	43.7	51.1	59.7	23.2	24.8	30.6	36.7	46.6	32.3	34.9	39.7	42.6	50.7
Meta-DETR	35.1	49.0	53.2	57.4	62.0	27.9	32.3	38.4	43.2	51.8	34.9	41.8	47.1	54.1	58.2
DeFR CN	40.2	53.6	58.2	63.6	66.5	<u>29.5</u>	<u>39.7</u>	43.4	48.1	<u>52.8</u>	35.0	38.3	52.9	57.7	60.8
FCT	38.5	49.6	53.5	59.8	64.3	25.9	34.2	<u>40.1</u>	<u>44.9</u>	47.4	34.7	43.9	49.3	53.1	56.3
VFA	<u>47.4</u>	<u>54.4</u>	<u>58.5</u>	64.5	66.5	33.7	38.2	<u>43.5</u>	<u>48.3</u>	52.4	43.8	<u>48.9</u>	53.3	58.1	60.0
FPD	41.5	52.8	58.4	<u>64.9</u>	<u>67.1</u>	28.2	<u>38.7</u>	<u>43.8</u>	<u>50.3</u>	53.6	34.9	48.6	<u>54.0</u>	<u>58.4</u>	<u>61.5</u>
Ours	43.4 ± 0.85	57.1 ± 0.55	60.2 ± 0.4	68.5 ± 0.25	68.0 ± 0.15	28.6 ± 1.10	43.5 ± 0.75	45.2 ± 0.5	52.8 ± 0.35	52.5 ± 0.2	35.6 ± 0.95	49.4 ± 0.6	55.3 ± 0.45	61.3 ± 0.3	64.2 ± 0.15



Fig. 3

Qualitative results of the proposed method on novel object categories. The bounding boxes demonstrate the model's ability to localize novel objects using the refined prototypes

To provide a more comprehensive evaluation of localization quality, we additionally report the results under the stricter AP_{75} metric. While AP_{50} measures detection performance with a relatively lenient IoU threshold, AP_{75} better reflects the model's ability to produce precise bounding boxes. The AP_{75} results of the proposed method across all novel sets and shot settings are summarized in Table II. As shown, our method maintains competitive performance under this stricter evaluation criterion,

indicating that the uncertainty-aware prototype refinement not only improves classification accuracy but also contributes to better localization of novel objects. Note that most competing methods have not reported AP_{75} results on PASCAL VOC; therefore, no direct comparison is provided under this metric.

TABLE II
Comparison of AP_{75} on PASCAL VOC Under Different Shot Settings in 5 Independent Runs

Metric (AP_{75})	Novel Set 1	Novel Set 2	Novel Set 3
1-shot	24.8 ± 0.9	16.1 ± 1.1	20.3 ± 1.0
2-shot	35.2 ± 0.6	26.8 ± 0.8	30.9 ± 0.7
3-shot	38.9 ± 0.5	28.4 ± 0.6	35.7 ± 0.5
5-shot	47.3 ± 0.4	34.7 ± 0.5	41.8 ± 0.4
10-shot	48.6 ± 0.3	35.2 ± 0.4	45.1 ± 0.3

H. Visualize Detection Results

To qualitatively evaluate the performance of our proposed method, we present the detection results in Fig. 4. The visualizations illustrate the model's ability to accurately localize and classify novel object categories. These results are obtained by training the model on the Novel Set 1 of the PASCAL VOC dataset under a 10-shot setting and subsequently evaluating it on the VOC2007 test set. The captured bounding boxes demonstrate the effectiveness of our uncertainty-aware distillation in producing more discriminative features for novel object recognition.

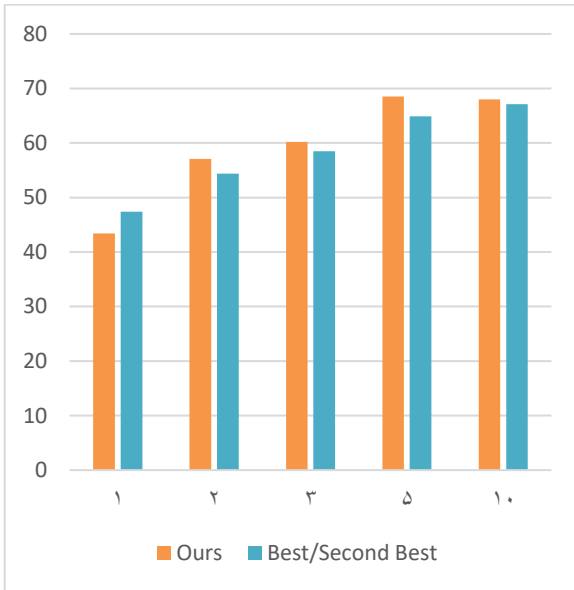


Fig. 4 Quantitative comparisons of mAP50 scores between the proposed method (orange) and the competitive baseline (blue) across different shot settings for Novel Set 1. The proposed model consistently outperforms the best baseline, especially in 1-shot and 2-shot scenarios, demonstrating enhanced robustness in few-shot object classification.

V. ABLATION STUDY

To evaluate the individual and combined contributions of the proposed modules, we conduct an ablation study on Novel Set 1 of the PASCAL VOC dataset. Specifically, we analyze the impact of Weighted Variance-Based Prototype Distillation and Entropy-Based Prototype Assignment across different few-shot settings ($K \in \{1, 5, 10\}$). The results are summarized in Table III.

As shown in the experimental results, the integration of both modules (“Both”) consistently achieves the best performance across all shot configurations. In the challenging 1-shot setting, the combined approach attains an AP_{50} of 43.4%, outperforming the individual variance-based (41.3%) and entropy-based (41.7%) modules. In the 5-shot and 10-shot settings, the full framework continues to demonstrate superior performance, reaching 68.5% and 68.0%, respectively.

The improvement observed with the variance-based module confirms its effectiveness in mitigating noise during prototype construction. Moreover, the additional gains introduced by the entropy-based module validate its role in reducing uncertainty during the prototype-query assignment stage. The complementary benefits of these two components highlight the importance of simultaneously improving prototype stability and assignment reliability for robust few-shot object classification.

TABLE III
Ablation Study (AP_{50}) on PASCAL VOC (Novel Set 1)

Shot(K)	w/Variance	w/Entropy	Ours(both)
1-shot	41.3	41.7	43.4
5-shot	67.4	67.8	68.5
10-shot	67.1	67.5	68.0

To investigate the sensitivity of the proposed method to the weighting parameter α in (7), we conduct an ablation study by varying α from 0.0 to 1.0 in increments of 0.1. As reported in Table IV, the model achieves the highest performance when α is set to 0.8 across all shot settings on Novel Set 1. This indicates that assigning a higher weight to the variance term while retaining a moderate contribution from the entropy term yields the most effective prototype distillation. Furthermore, the performance remains relatively stable within the range of $\alpha \in [0.7, 0.9]$, demonstrating that the proposed framework is robust to moderate variations in the weighting coefficient. Based on these results, we adopt $\alpha=0.8$ as the default value in all experiments.

TABLE IV
Sensitivity Analysis of the Weighting Coefficient α in (7) on Novel Set 1 of PASCAL VOC

α	1-shot	2-shot	5-shot	10-shot
1.0	41.3	55.8	67.4	67.1
0.9	42.7	56.4	67.9	67.6
0.8	43.4	57.1	68.5	68.0
0.7	42.9	56.7	68.1	67.8
0.6	42.1	55.9	67.6	67.3
0.5	41.5	55.2	67.2	66.9
0	41.7	54.8	67.8	67.5

VI. CONCLUSION

In this work, we addressed the fundamental challenge of recognizing novel object categories in few-shot scenarios where labeled data are extremely scarce. Conventional object detection frameworks often incur high computational costs and require extensive retraining when adapting to new classes. To overcome these limitations, we proposed an enhanced version

of the Fine-Grained Prototype Distillation (FPD) framework by introducing uncertainty-aware refinement mechanisms.

Specifically, the proposed approach improves the quality of class representations by exploiting the statistical stability of support features. Through a weighted variance-based prototype distillation strategy, the model prioritizes stable and representative features while suppressing the influence of noisy or less informative regions. In addition, an entropy-based prototype assignment mechanism is introduced to enable more precise alignment between query regions and class prototypes.

Experimental results on the PASCAL VOC benchmark demonstrate that the proposed enhancements significantly improve the robustness and discriminative capability of the learned prototypes. As a result, the proposed framework achieves competitive performance compared with existing state-of-the-art approaches and demonstrates strong generalization ability to novel object categories under highly data-constrained conditions.

Despite these improvements, the proposed method has certain limitations. First, the current implementation does not utilize a Feature Pyramid Network (FPN), which may limit its ability to handle objects of varying scales. Second, all experiments were conducted exclusively on the PASCAL VOC dataset, and the effectiveness of the method on larger and more challenging benchmarks, such as MS COCO, remains to be validated. Additionally, the variance estimation in the 1-shot setting relies on spatial variance within a single support sample, which may not fully capture inter-sample uncertainty.

In future work, we intend to incorporate multi-scale feature representations through FPN or more advanced backbone architectures. Furthermore, exploring the application of the uncertainty-aware distillation strategy to other few-shot tasks, such as instance segmentation and video object detection, presents a promising direction for research.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this article.

AUTHORS CONTRIBUTION STATEMENT

All authors contributed equally to this work.

REFERENCES

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Computer Vision and Pattern Recognition*, Boston, 2015.
- [2] Z. Wang, B. Yang, H. Yue, and Z. Ma, "Fine-Grained Prototypes Distillation for Few-shot Object Detection," in *AAAI*, Vancouver, 2024.
- [3] Q. Fan and W. Zhuo, "Few-Shot Object Detection via Variational Feature Aggregation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, 2022.
- [4] K. Chen, "MMDetection: Open MMLab Detection Toolbox and Benchmark," in *arXiv*, Washington, 2020.
- [5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, and A. Karpathy, "ImageNet Large Scale Visual Recognition Challenge," *IJCV*, vol. 115, no. 3, pp. 211-252, 2015.
- [6] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 2017.
- [7] X. Fan, H. Yang, and J. Yan, "Meta-RCNN: Meta Learning for Few-Shot Object Detection," in *Proceedings of the 28th ACM International Conference on Multimedia*, Washington, 2020.
- [8] Y. Xiao, V. Lepetit, and R. Marlet, "Few-Shot Object Detection and Viewpoint Estimation for Objects in the Wild," in *Computer Vision and Pattern Recognition*, Seattle, 2020.
- [9] X. Hu, "Dense Relation Distillation with Context-aware Aggregation for Few-Shot Object Detection," in *CVPR*, Virtual, 2021.
- [10] A. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," *IEEE*, pp. 35479-35516, 2023.
- [11] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, pp. 257-276, 2023.
- [12] R. Q. Qu and H. Liang, "A Survey of Deep Learning-Based Object Detection," *IEEE*, p. 126637-126668, 2019.
- [13] B. Kang and Y. Xue, "Few-Shot Object Detection: A Survey," *ACM*, pp. 1-41, 2022.
- [14] Y. Song and T. Y. Wang, "A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities," *ACM*, pp. 1-57, 2023.
- [15] X. Zhang, X. Tan, and X. Weng, "Few-Shot Object Detection: Research Advances and Challenges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 4608-4627, 2022.
- [16] C. Sun and Y. Li, "Dual-Attention Model for Camera-Based Few-Shot Object Detection," *IEEE Sensors*, pp. 2045-2059, 2025.
- [17] T. Wang, "A Generalized Few-Shot Object Detection Method via Extraction of Base-Novel Commonality With Memory Distillation of Category Prototypes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 7, pp. 456-467, 2025.
- [18] R. Xiao and L. Lepetit, "Few-Shot Object Detection and Viewpoint Estimation for Objects in Wild," *IEEE*, vol. 23, no. 4, pp. 2045-2052, 2022.
- [19] J. He, "Recent Few-shot Object Detection Algorithms: A Survey with Performance Comparison," *Springer*, vol. 54, no. 13, pp. 9043-9064, 2024.
- [20] S. Zhang and W. Wang, "Extreme R-CNN: Few-Shot Object Detection via Sample Synthesis and Knowledge Distillation," *Sensors*, vol. 24, no. 23, pp. 9023-9048, 2024.
- [21] N. Carion, F. Massa, and G. Synnaeve, "End-to-End Object Detection with Transformers," in *Computer Vision and Pattern Recognition*, 2020.

BIOGRAPHIES



Mohammad Hossein Adib is an M.Sc. graduate in Artificial Intelligence with a strong interest in research and development in computer vision. His research interests include object detection and classification, with a particular focus on few-shot learning. His work aims to improve the performance, robustness, and generalization of object detection models in scenarios where only limited annotated training data are available.



Mohammad Javad Fadaeieslam received his MSc and PhD degrees in artificial intelligence from Iran University of Science and Technology (IUST) in 2005 and 2013. He is currently an assistant professor in the Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran. His research interests include computer vision, deep learning, multimedia, and data mining.