Angus Academic Press CSECEE 2025
ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168

AFA-Underwater: Dual-Path Adaptive Feature Augmentation for Few-Shot Deep-Sea Trash Detection

Guangdong Zhu [a], Li Zhang [b] *

College of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China.

[a] guangdong2513@163.com, [b] zhangli_dx@xpu.edu.cn

Abstract. As marine pollution intensifies, developing efficient deep-sea garbage detection technologies has become increasingly urgent. However, deep-sea environments present two major challenges: (i) the scarcity of target samples (few-shot problem), which makes it difficult to train deep learning models effectively; and (ii) harsh underwater imaging conditions—such as poor illumination and scattering—that lead to object blurring and background confusion. To address these challenges, this paper proposes AFA-Underwater, a framework specifically designed for few-shot deep-sea garbage detection. The core of this framework is a dual-path Adaptive Feature Augmentation (AFA) module, which generates more robust class prototypes to effectively mitigate the "prototype shift" problem caused by underwater noise. In addition, we introduce a novel Tri-Path Aggregation (TPA) mechanism that jointly models the raw, common, and differential representations between support and query features, thereby alleviating base-class bias. Extensive experiments on the re-partitioned TrashCAN 1.0 dataset demonstrate that AFA-Underwater significantly outperforms existing FSOD methods under multiple few-shot settings, achieving a remarkable 14.4% improvement in novel-class mAP under the 10-shot scenario. The algorithm also shows strong generalization capability on the PASCAL VOC dataset.

Keywords: Few-Shot Object Detection; Deep-Sea Trash Detection; Meta-Learning.

1. Introduction

Marine pollution has become an increasingly severe issue, with plastic waste posing a particularly destructive threat to marine ecosystems, making automatic detection and recognition of deep-sea debris a critical task[1]. However, this task faces two major challenges in practical applications:(i) Data scarcity: Deep-sea data collection and annotation are extremely costly (e.g., labeling a single sample can take several hours), resulting in a limited number of available samples and a severely imbalanced category distribution (long-tail problem).(ii) Poor underwater image quality: Due to light attenuation, scattering, and medium absorption, underwater images typically suffer from low contrast, color distortion, and object blurring. As a result, debris targets (e.g., plastic bags, fishing nets) are easily confused with the background (e.g., rocks, seaweed), while exhibiting large intra-class variations caused by corrosion or biofouling[2].

These challenges make existing **few-shot object detection (FSOD)** methods difficult to apply directly[3]. Most FSOD approaches are designed for natural scenes with high-resolution and clean features. When faced with the strong noise and large intra-class variance of deep-sea images, they tend to suffer from **prototype shift**—that is, the feature prototypes extracted from support samples fail to accurately represent their corresponding classes, leading to mismatched predictions and significant drops in detection accuracy.

2. RELATED WORK

2.1 General Object Detection

General object detection techniques rely heavily on large-scale annotated datasets (e.g., COCO, PASCAL VOC) and thus experience a significant performance drop in real-world scenarios with limited samples or high noise levels. For instance, YOLOv4 suffers a 37% decrease in mAP under a

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168
10-shot setting on PASCAL VOC, while DETR shows strong sensitivity to image quality—its false detection rate in noisy environments is approximately 2.1 times higher than in clean scenes[4][5].

To address underwater challenges such as optical distortion and low contrast[6], several studies have explored improvements from different perspectives, including attention mechanisms, pseudo-label learning, and noise suppression[7]. However, these approaches still depend on large amounts of annotated data and fail to overcome the generalization bottleneck inherent in few-shot scenarios.

2.2 Few-shot object detection

Few-shot object detection (FSOD) aims to enable rapid adaptation to novel categories using only a limited number of labeled samples. Transfer learning–based approaches: These methods typically pretrain on a large set of base classes and then fine-tune on novel classes[8]. For example, TFA freezes the backbone and fine-tunes only the classifier[9]; FSCE introduces contrastive proposal encoding[10]; and DeFRCN enhances module cooperation through gradient decoupling[11]. However, these approaches often suffer from slow convergence and limited generalization when applied to complex environments. Meta-learning–based approaches: These methods train by simulating few-shot scenarios, allowing the model to “learn how to learn.” Early representatives such as Meta R-CNN and FSRW combine support set features with query RoI features[12]. Subsequent works have explored more advanced feature aggregation and attention mechanisms[13], offering new insights for handling challenges such as object blurriness and texture similarity in few-shot detection.

2.3 Feature Enhancement and Underwater Image Enhancement Methods

In terms of feature enhancement, modules such as SE[14], CBAM[15], and DCN[16] have demonstrated strong performance in general object detection tasks. However, they lack dynamic adaptability to address both few-shot scarcity and underwater noise interference, and their computational cost remains high. For underwater image enhancement, methods such as UWCNN, Ucolor, and WaterNet have been proposed[17][18][19]. Nevertheless, these approaches either rely on large amounts of paired training data or exhibit high model complexity, making them difficult to generalize effectively or deploy efficiently under few-shot conditions.

In summary, existing methods fail to adequately address the dual challenges inherent in deep-sea few-shot scenarios. Therefore, this paper proposes a dual-path AFA module and a TPA aggregation mechanism, designed to maintain model efficiency while significantly improving the robustness and generalization of feature representations.

3. Method

3.1 Model Overview

The proposed AFA-Underwater framework (shown in Fig. 1) is built upon the classic meta-learning detector Meta R-CNN. It is designed to enhance few-shot object detection performance in complex underwater environments. The core workflow is as follows: features from support images are fed into the proposed AFA module for adaptive enhancement to generate robust class prototypes; features from query images are processed by the RPN to generate RoIs; the enhanced support features and query RoI features are then input into the TPA feature aggregation module for interaction; finally, the fused features are passed to the detection head for object classification and bounding box regression, enabling efficient detection and precise localization of novel-class targets.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168

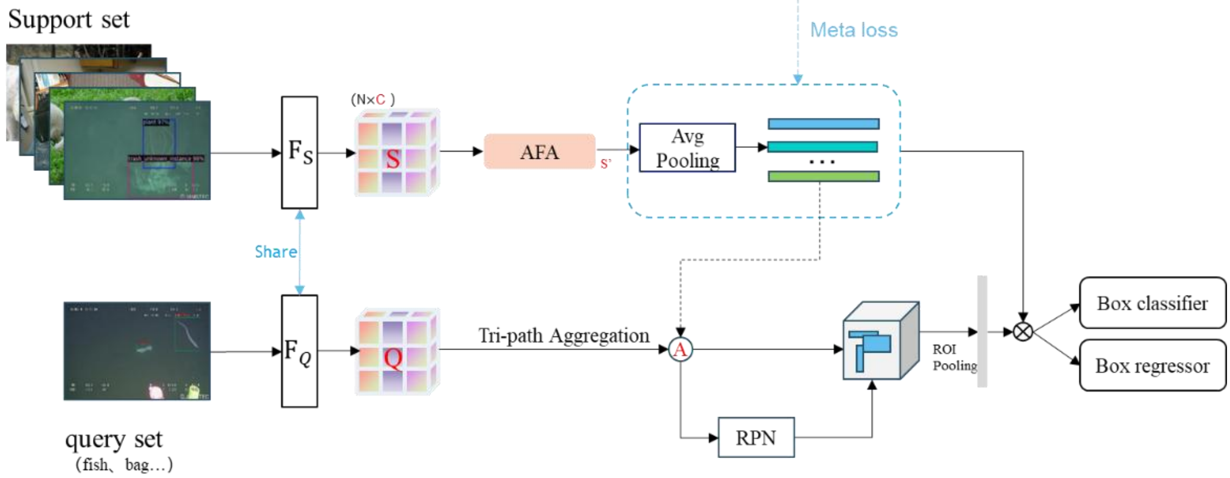

Fig. 1 Overview of the AFA-Underwater Framework

3.2 Adaptive Feature Augmentation (AFA) Module

In few-shot tasks, the support features F_s carry the core semantic information for constructing class prototypes. However, in underwater environments, noise caused by light attenuation, scattering, and other factors makes F_s highly unstable, leading to "prototype shift," where the generated prototypes fail to accurately represent the target's intrinsic properties, thus weakening matching accuracy. To address this, we design the AFA module (Fig. 2), which performs fine-grained enhancement of F_s along both channel and spatial dimensions.

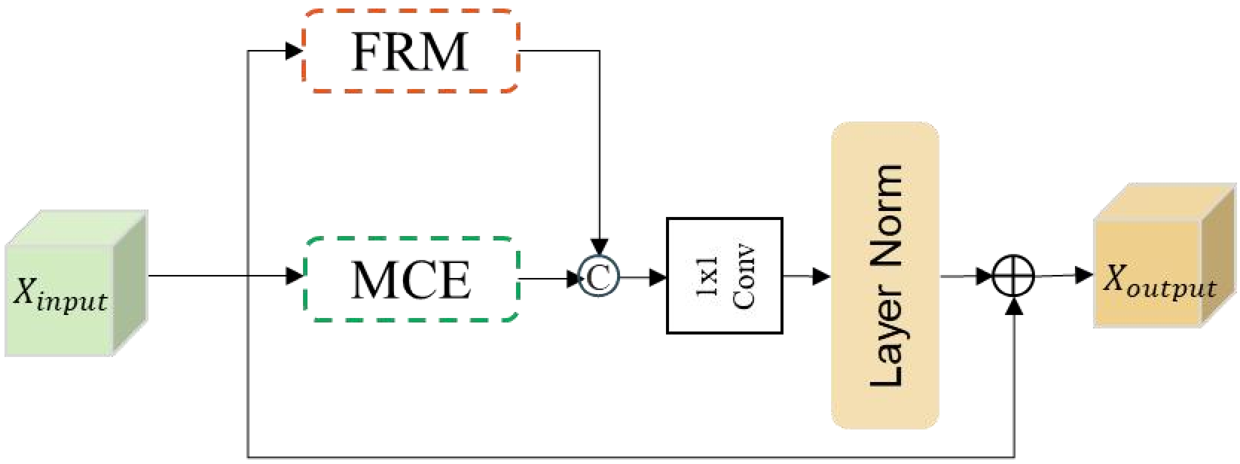

Fig. 2 The AFA module consists of two parallel paths: FRM (channel path) and MCE (spatial path).

Feature Refinement Module (FRM) – (Channel Path): As shown in Fig. 3(a), this path focuses on semantic enhancement along the channel dimension and consists of two parallel computational branches:Left Branch: A lightweight MLP (1×1 Conv → Norm → GELU, repeated twice) for feature refinement and optimization.Right Branch: An SE-like structure (GAP → FC → Sigmoid) generating channel attention weights that represent the relative importance of each channel.The outputs of the two branches are multiplied element-wise, amplifying key information channels while suppressing noisy or redundant channels. A residual connection is applied to ensure that the original crucial information is preserved while enhancing features.

Multi-Scale Context Extractor (MCE) – (Spatial Path): As shown in Fig. 3(b), this path aims to enhance the perception of targets at different spatial scales, addressing the large scale variation and diverse morphology of underwater objects. It employs three parallel depthwise separable convolutions with different receptive fields (DWConv 3×3, 5×5, and dilated 3×3) to capture fine

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168
details of small targets, mid-scale patterns, and larger contextual semantics, respectively. The resulting features are concatenated and fused to adapt to the large-scale variations of underwater targets.

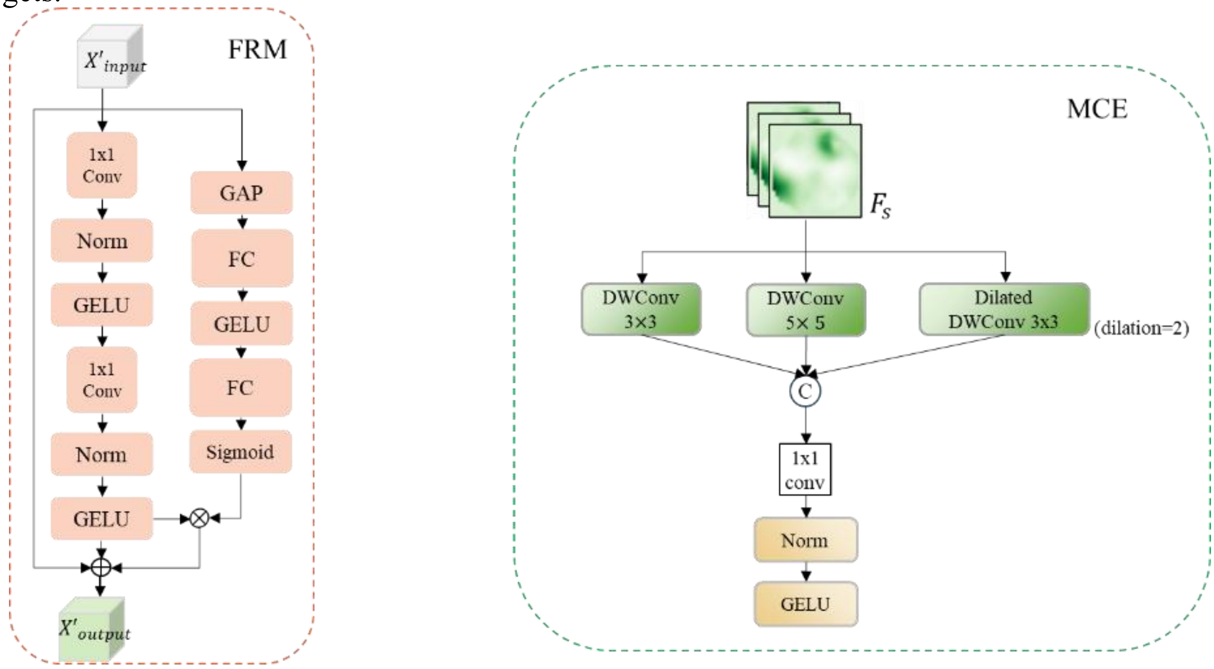

Fig.3 (a) Feature Refinement Module (FRM) (b) Multi-Scale Context Extractor (MCE)

3.3 Tri-path Aggregation (TPA)

In FSOD, feature aggregation is a critical component. Many methods adopt a Class-Specific Aggregation (CSA) strategy, as illustrated in Fig. 6(a), where only support features S_i of the same class are aggregated with the query RoI features Q_i. Although intuitive, this approach ignores inter-class relationships. When there is an imbalance between base and novel classes, such "isolated" aggregation can exacerbate the model's bias toward base classes.

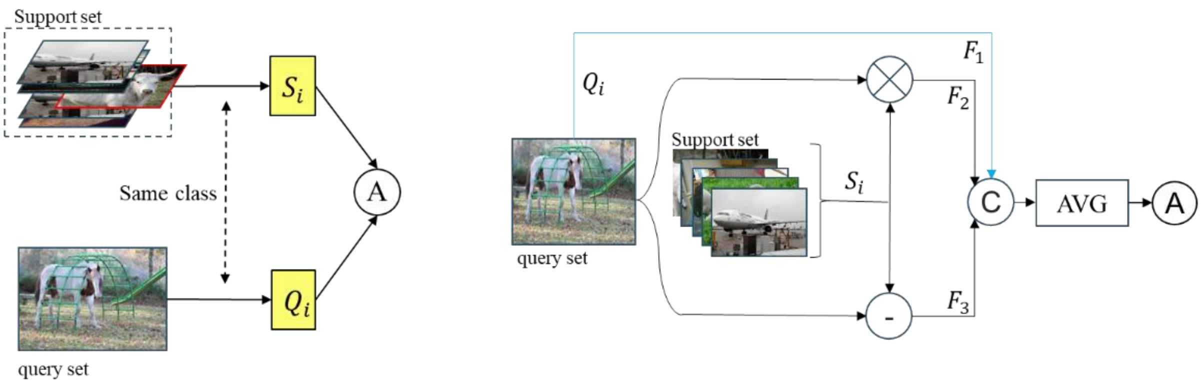

Fig. 6 Intuitive comparison between CSA and TPA.

To address this issue, we propose the Tri-Path Aggregation (TPA) mechanism (Fig. 6(b)). The core idea of TPA is to model the multi-dimensional relationships between the support feature and the query feature, rather than performing simple within-class matching. TPA processes and through three parallel paths:

Raw Path: $F_1 = Q_i$ directly preserves the original basic representation of the query feature (Q_i).

Common Path: $F_2 = S_i \otimes Q_i$ employs the Hadamard product (element-wise multiplication) to amplify jointly activated feature regions, enhancing their shared discriminative characteristics and highlighting similarity.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168

Difference Path: $F_3 = |S_i - Q_i|$ computes the absolute element-wise difference to capture subtle discrepancies between the two, which is critical for distinguishing highly similar objects (e.g., different types of trash or trash versus marine organisms).

Finally, the features from the three paths, F_1, F_2, F_3 are concatenated along the channel dimension, F_{cat} =Concat(F_1, F_2, F_3), and an averaging operation is applied to balance the contributions of each support sample.

4. Experiments and Analysis

4.1 Dataset Introduction

TrashCAN 1.0: A benchmark dataset specifically designed for deep-sea garbage detection. We divide it into three non-overlapping splits (Split1–3), each containing 15 base classes for pre-training and 5 novel classes for few-shot fine-tuning. Splits 1–3 focus on shape-confusable objects, man-made containers, and transparent objects, respectively. k-shot training sets are constructed with (k ∈{1, 2, 3, 5, 10}).

PASCAL VOC: Used to evaluate the generalization ability of AFA-Underwater beyond underwater scenarios, following the standard 15-base / 5-novel class split.

4.2 Ablation Study

We conducted a series of ablation experiments on the TrashCan 1.0 dataset to verify the effectiveness of each core component in the AFA-Underwater framework (for brevity, only 3-shot and 10-shot settings are reported).

Effectiveness of the AFA Module. As shown in Table 1, incorporating the full AFA module into the Meta R-CNN baseline results in substantial improvements in novel-class mAP across all Splits. For example, in Split 1 under the 3-shot setting, mAP increases from 16.5% to 27.3% (+10.8%). This strongly demonstrates that the AFA module, by enhancing support features and mitigating "prototype shift," significantly improves the model's few-shot recognition capability.

Table.1 Ablation results of the AFA module on TrashCAN 1.0 under the 3-shot setting (novel class mAP@50). The baseline model is Meta-RCNN.

	Split 1 3-shot (%)	Split 2 3-shot (%)	Split 3 3-shot (%)
Baseline (Meta-RCNN)	16.5	14.2	21.5
+ AFA	27.3 (+10.8)	23.8 (+9.6)	31.5 (+10.0)

Contribution of the Dual-Path AFA Structure. In Table 2, we evaluate the two paths of the AFA module separately: FRM (channel) and MCE (spatial). Under the 10-shot setting, using only FRM (channel path) improves mAP by 6.7%, and using only MCE (spatial path) improves it by 7.4%. Enabling both paths simultaneously (full AFA) achieves a substantial 11.7% increase. This indicates that the channel and spatial paths capture complementary information, and their combined effect is optimal, validating the rationale behind our dual-path design.

Table 2. Ablation study of AFA's dual-path structure (mAP@0.5 on TrashCan 1.0 SPLIT1, 10-shot)

Method	mAP@0.5
Baseline (Meta R-CNN)	22.8%
Baseline + AFA (FRM only - Path1)	29.5%
Baseline + AFA (MCE only - Path2)	30.2%
Baseline + AFA (FRM + MCE - Dual Path)	34.5%

Effectiveness of the TPA Module. As shown in Table 3, we replace the baseline CSA aggregation with the proposed TPA module. Under the 10-shot setting, TPA consistently improves

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168

performance across all splits, with an average gain of 2.3%. This demonstrates that TPA, by explicitly modeling raw, common, and difference information, provides a more comprehensive understanding of support-query relationships and effectively mitigates base-class bias.

Table 3. Ablation study on the effectiveness of the TPA module (mAP@0.5 on TrashCan 1.0 SPLIT1-3, 10-shot).

Method	Split 1 10-shot (%)	Split 1 10-shot (%)	Split 1 10-shot (%)
Baseline + AFA (with CSA)	34.5	23.7	37.7
Baseline + AFA + TPA (Ours)	37.3	31.5	41.5

4.3 Experimental Analysis

We present a comprehensive comparison of AFA-Underwater with 11 state-of-the-art methods on TrashCAN 1.0 in Table 4.

Table.4 Performance evaluation (nAP@50) of several state-of-the-art models on the TrashCAN 1.0 dataset. The results are grouped by novel set and sorted by average score (Avg.). All models use ResNet-50 as the backbone.

	SPLIT1					SPLIT2					SPLIT3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
RepMet	8.0	12.2	15.0	18.4	22.0	6.1	9.3	11.0	14.6	18.1	11.0	14.5	18.2	22.2	26.5
Meta R-CNN	7.8	11.9	14.4	12.8	21.3	14.9	17.3	17.8	23.8	24.5	12.4	17.8	20.7	27.6	27.1
Attention RPN	9.5	9.8	6.8	9.4	10.3	14.9	10.9	13.4	22.5	18.7	13.3	15.8	18.1	21.5	24.7
TFA	13.0	11.2	13.1	16.6	26.3	5.6	10.9	14.2	17.3	14.2	18.0	17.0	19.9	27.2	31.6
FsDet	14.3	21.5	13.9	16.5	22.4	14.6	13.6	14.2	18.3	30.5	22.3	22.9	24.8	30.5	38.0
Meta FR-CNN	9.9	18.6	18.8	22.5	24.5	9.6	13.8	14.9	22.5	25.3	17.3	20.3	27.3	32.1	40.6
FSCE	11.2	12.8	5.5	10.4	13.5	5.9	7.5	9.7	11.3	19.5	11.8	14.8	17.6	21.0	27.1
DeFRCN	14.9	15.1	19.1	20.7	21.8	16.5	15.4	17.6	19.9	20.5	25.2	24.7	28.3	32.8	41.0
MPSR	14.0	18.8	15.0	14.7	19.2	11.8	14.6	16.8	18.3	22.6	28.6	26.8	27.8	33.6	39.7
FCT	10.5	16.2	13.9	17.0	17.4	10.6	13.6	13.9	19.9	25.5	18.0	18.1	23.6	29.9	40.3
AFA-U (Ours)	20.5	23.6	25,2	32.2	37.3	18.8	21.0	21.7	27.3	31.5	24.6	25.4	28.8	31.5	41.5

The experimental results clearly demonstrate the following: AFA-Underwater consistently outperforms all other methods across all Splits and k-shot settings. In the most challenging 1-shot scenario, our method achieves an average nAP of 21.3%, significantly higher than DeFRCN's 13.0%, highlighting its strong feature learning capability under extreme data scarcity. In the 10-shot setting, AFA-Underwater attains 41.5% nAP on Split 3, representing a 14.4-point increase over the Meta-RCNN baseline (27.1%), fully validating the effectiveness and superiority of the AFA and TPA modules for deep-sea few-shot detection.

4.4 Generalization Ability on PASCAL VOC

To evaluate generalization, we conducted experiments on the PASCAL VOC dataset (Table 5). The results indicate that AFA-Underwater also achieves competitive performance on a general-purpose dataset. In the 10-shot setting, its overall performance is comparable to or slightly better than DeFRCN and FCT, and clearly superior to methods such as TFA and FSCE. This demonstrates that the proposed AFA and TPA modules are not limited to underwater scenarios and exhibit strong generalization capability.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168

Table.5 Few-shot detection performance (mAP@0.5) on PASCAL VOC dataset across three splits.

	SPLIT1					SPLIT2					SPLIT3				
	1	2	3	5	10	1	2	3	5	10	1	2	3	5	10
Meta-RCNN	19.9	25.5	35.5	45.7	51.5	10.4	19.4	29.6	34.8	45.4	14.3	18.2	27.5	41.2	48.1
TFA (2021)	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
FSCE	44.2	43.8	51.4	61.9	63.4	27.3	29.5	43.5	44.2	50.2	37.2	41.9	47.5	54.6	58.5
FCT	**49.9**	57.1	57.9	63.2	67.1	27.6	34.5	43.7	44.2	51.2	**49.5**	**54.7**	52.3	57.0	**58.7**
Meta FR-CNN	43.0	54.5	60.6	**66.1**	**65.4**	27.7	35.5	46.1	47.8	51.4	40.6	46.4	**53.4**	**59.9**	58.6
SQMG	48.6	51.1	52.0	53.7	54.3	**41.6**	**45.4**	45.8	46.3	48.0	46.1	51.7	52.6	54.1	55.0
DeFRCN	53.6	57.5	61.5	64.1	60.8	30.1	38.1	47.0	**53.3**	47.9	48.4	50.9	52.3	54.9	57.4
AFA-U (Ours)	48.7	**59.4**	**62.2**	63.9	65.2	27.5	43.2	**47.1**	51.2	**51.9**	38.6	48.7	50.1	59.4	57.4

5. Summary

This paper presents AFA-Underwater, a dual-path feature enhancement framework for few-shot deep-sea garbage detection. The method addresses two major challenges in deep-sea environments: data scarcity and low-quality imagery, through two key innovations: AFA module: Employing dual-path channel and spatial enhancement, it effectively mitigates the “prototype shift” caused by underwater noise, generating robust class prototypes. TPA mechanism: By introducing a three-branch aggregation of raw, common, and difference features, it replaces the conventional CSA, significantly improving support-query matching accuracy and alleviating base-class bias. Extensive experiments on the TrashCAN 1.0 dataset demonstrate that AFA-Underwater consistently outperforms current state-of-the-art methods across all few-shot settings, and it also exhibits strong generalization on PASCAL VOC.

References

[1] Jambeck J R, Geyer R, Wilcox C, et al. Plastic waste inputs from land into the ocean[J]. science, 2015, 347(6223): 768-771.

[2] Mathangi M. Detection of underwater trash objectsusing deep learning algorithms[J]. 2023.

[3] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[4] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[5] Jiang Z, Zhao L, Li S, et al. Real-time object detection method based on improved YOLOv4-tiny[J]. arXiv preprint arXiv:2011.04244, 2020.

[6] Zhou J, He Z, Zhang D, et al. Dgnet: Dynamic gradient-guided network with noise suppression for underwater image enhancement[J]. arXiv preprint arXiv:2312.06999, 2023.

[7] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.

[8] Kang B, Liu Z, Wang X, et al. Few-shot object detection via feature reweighting[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 8420-8429.

[9] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[10] Sun B, Li B, Cai S, et al. Fsce: Few-shot object detection via contrastive proposal encoding[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 7352-7362.

ISSN: 3007-9640 DOI: https://doi.org/10.70114/acmsr.2025.5.1.P168
[11] Qiao L, Zhao Y, Li Z, et al. Defrcn: Decoupled faster r-cnn for few-shot object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 8681-8690.

[12] Yan X, Chen Z, Xu A, et al. Meta r-cnn: Towards general solver for instance-level low-shot learning[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9577-9586.

[13] Fan Q, Zhuo W, Tang C K, et al. Few-shot object detection with attention-RPN and multi-relation detector[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 4013-4022.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[15] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[16] Chen F, Wu F, Xu J, et al. Adaptive deformable convolutional network[J]. Neurocomputing, 2021, 453: 853-864.

[17] Li C, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement[J]. Pattern recognition, 2020, 98: 107038.

[18] Li C, Anwar S, Hou J, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding[J]. IEEE Transactions on Image Processing, 2021, 30: 4985-5000.

[19] Li C, Guo C, Ren W, et al. An underwater image enhancement benchmark dataset and beyond[J]. IEEE transactions on image processing, 2019, 29: 4376-4389.