Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **2**, June **2025**

EnAF: A network for Enhancing image quality based on Adaptive Frequency to support traffic violation image recognition

Pho Hai Dang

Faculty of Management Information Systems
Ho Chi Minh University of Banking
Ho Chi Minh City, Vietnam
dangph@hub.edu.vn

Received February 11, 2025, revised April 16, 2025, accepted April 17, 2025.

Abstract. *Traffic violation detection using traffic cameras is one of the key challenges in the field of traffic safety. The application of deep learning techniques, particularly modern deep learning approaches such as few-shot learning, is essential for supporting and enhancing traffic safety management capabilities. Therefore, in this study, we develop a few-shot deep learning model named EnAF for traffic violation detection, leveraging adaptive frequency features. This approach enables the model to understand different frequency components in an image, corresponding to various color scales, to improve recognition quality. We conduct experiments on the COCO dataset and the T-Traffic dataset (collected from traffic cameras) to evaluate the effectiveness of the proposed model.*
Keywords: Traffic violation detection, few-shot learning, image recognition

1. **Introduction.** In recent years, the rate of traffic accidents has increased dramatically [1]. As a result, many traffic safety management agencies have developed policies and tools to support and warn about traffic accidents. The deployment of camera systems on roads has been widely implemented in many countries, including Vietnam. However, the integration of AI [2, 3] technologies into these camera systems has not been extensively developed due to limitations in infrastructure and AI technology.

Over the past few decades, deep learning models [4] have significantly advanced in tasks such as image recognition and object detection. Building models to recognize traffic violation behaviors is an essential and critical task to help reduce traffic accidents. However, labeling and processing traffic violation data remain limited at present, as these tasks require human intervention and are time-consuming. Additionally, traffic cameras are usually positioned at high altitudes, making it challenging to detect traffic violation behaviors accurately.

In this study, we propose a few-shot deep learning model capable of learning pixel-level features through adaptive frequency analysis on images. Learning adaptive frequencies allows us to better understand each pixel in the image, determining which color scales affect the recognition of objects and which color scales influence image quality. The ability to respond to adaptive frequencies helps address issues related to motion blur or smoothness in images, providing more comprehensive features and a clearer understanding of the overall image structure. Furthermore, we explore the model in the direction of learning with small datasets, which aligns with the limited data availability and better meets real-world scenarios.

In summary, the key contributions of our study include:

- Proposing the EnAF model, a few-shot deep learning model capable of understanding features corresponding to adaptive scanning frequencies to support object recognition.
- Introducing the T-Traffic dataset, which we collected from traffic cameras installed on roads in Hanoi, Vietnam.
- Evaluating the proposed EnAF model on both the COCO and T-Traffic datasets using qualitative and quantitative metrics.

2. Related Works.

2.1. **Few-shot semantic segmentation.** Few-shot deep learning [5, 6, 7, 8] is a deep learning approach that achieves significant performance in tasks such as image segmentation [9], image recognition [10], and image detection [7], among others. However, the accuracy of these methods remains limited due to their reliance on less data compared to traditional deep learning methods. Recently, Lang et al. [11] developed a feature extraction method through segmentation domains to enhance the predictive quality of few-shot deep learning models. Sun et al. [12] constructed feature maps from convolutional layers in few-shot deep learning segmentation models to establish a mechanism for supporting image recognition. The team of Xiao [13] integrated few-shot deep learning with Segment Anything for image recognition. In summary, while research groups have successfully developed few-shot deep learning models, they have not yet thoroughly examined adaptive frequency details or adaptive resolution under varying weather conditions.

2.2. **Few-shot learning based frequency.** Few-shot deep learning trained on adaptive frequency domains [14, 15, 16, 17] aims to address complex tasks in image analysis, such as recognizing traffic violations on roadways. Ma et al. [18] proposed a few-shot deep learning method to solve image recognition tasks on combined adaptive frequency domains, enabling pixel-level understanding of images. Zhu et al. [19] demonstrated that features within the network can extract information distributed across frequency domains. Research groups have constructed the RGB color domain of images to locate and predict objects using information such as color, texture, and surface features. However, these representation methods are not fully adequate and often result in significant information loss.

3. Proposed Method.

3.1. **Problem Definition.** The objective of this study is to develop a deep learning model capable of recognizing rare or previously unseen traffic violations using a few-shot learning approach. These violations may include illegal parking, red-light running, or wrong-way driving, detected from RGB camera footage. Given the challenge of acquiring large labeled datasets for every possible violation type, the model must generalize to new violation categories with minimal labeled examples (*K-shot learning*). To achieve this, it leverages adaptive RGB frequency analysis, extracting discriminative spatio-spectral features that enhance robustness and accuracy in prediction.

The problem is formulated as a 1-way K-shot weakly supervised few-shot learning (WFSL) task, where the model learns to classify novel traffic violations with limited supervision. The training process relies on two datasets: a base dataset (D_{base}), containing frequently observed traffic violations such as speeding or lane deviation, and a novel dataset (D_{novel}), consisting of rare or unseen violation types like pedestrian lane encroachment or emergency lane misuse. Notably, the classes in these two datasets are

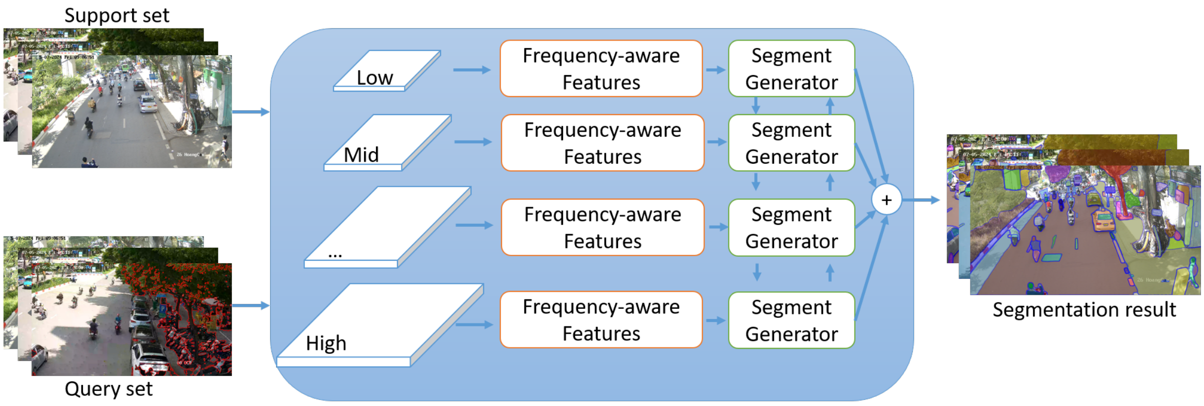

FIGURE 1. The framework of the proposed EnAF

disjoint, meaning there is no overlap between the violations seen during training (C_{train}) and those encountered during testing (C_{test}), ensuring:

$$C_{\text{train}} \cap C_{\text{test}} = \varnothing \tag{1}$$

The model is trained on D_{base} and then evaluated on D_{novel}, where it must recognize new violations using only image-level labels during meta-training, without requiring pixel-wise annotations. This weakly supervised learning approach enables the system to adapt to new traffic violations efficiently, making it suitable for real-world applications where labeled data is scarce.

3.2. **Model Architecture.** The EnAF framework (show in Figure 1) is designed for traffic violation recognition using surveillance cameras by leveraging frequency-aware multi-scale feature extraction. The framework consists of multiple processing stages, where input RGB images are analyzed at different levels of feature granularity to improve segmentation accuracy.

At its core, the Cross-granularity Frequency-aware Module (CFM) [18] extracts feature representations from low, mid, and high levels of the backbone network. Each feature level is processed through a Frequency-aware Module, which decomposes RGB domain information into high-frequency and low-frequency components across different granularities. This decomposition helps realign spatial structural information in the frequency domain, optimizing feature representation for segmentation.

Segment Generator module is responsible for refining frequency-aware feature maps into high-quality segmentation outputs. After the Cross-granularity Frequency-aware Module (CFM) decomposes features into high-frequency and low-frequency components, these features are passed to the Segment Generator. It applies convolutional refinement layers, followed by upsampling and normalization operations, to produce pixel-wise predictions of traffic violation regions. Each Segment Generator operates on a different granularity level (low, mid, high), and their outputs are later fused to form the final segmentation mask.

Each frequency-aware feature set is then passed through a Segment Generator, which refines the extracted features for enhanced segmentation. Finally, the outputs from multiple feature levels are aggregated to generate the final segmentation results, which highlight different objects and regions in the traffic scene.

The EnAF framework improves segmentation accuracy in traffic monitoring scenarios, even under challenging conditions, by leveraging multi-scale frequency domain learning.

This structured approach ensures better object differentiation, making the model effective for real-world traffic violation detection applications.

3.3. **Loss function.** The loss function for the EnAF model is designed to optimize both segmentation accuracy and feature consistency across different frequency levels. The total loss function integrates three key components: segmentation loss, frequency-aware consistency loss, and multi-scale feature alignment loss.

First, the segmentation loss ($\mathcal{L}_{seg}$) ensures that the predicted segmentation masks align with the ground truth annotations. This can be formulated using the Cross-Entropy (CE) loss:

$$\mathcal{L}_{seg} = -\sum_i M_i \log \hat{M}_i \tag{2}$$

where M_i is the ground truth mask and $\hat{M}_i$ is the predicted probability for each pixel. Alternatively, Dice loss can be used to address class imbalance:

$$\mathcal{L}_{dice} = 1 - \frac{2\sum_i M_i \hat{M}_i}{\sum_i M_i + \sum_i \hat{M}_i} \tag{3}$$

Second, the frequency-aware consistency loss ($\mathcal{L}_{freq}$) is introduced to maintain consistency between high-frequency and low-frequency feature representations extracted by the Cross-granularity Frequency-aware Module (CFM). This loss minimizes the discrepancy between high-frequency (F_l^H) and low-frequency (F_l^L) components across multiple layers:

$$\mathcal{L}_{freq} = \sum_l ||F_l^H - F_l^L||^2 \tag{4}$$

Lastly, the multi-scale feature alignment loss ($\mathcal{L}_{align}$) ensures proper alignment of features extracted from different levels before being processed by the Segment Generators. This loss reduces feature discrepancies between different granularity levels (F_i and F_j) using a transformation function $\mathcal{T}$ to match their representations:

$$\mathcal{L}_{align} = \sum_{i,j} ||F_i - \mathcal{T}(F_j)||^2 \tag{5}$$

The total loss function is then formulated as a weighted combination of these three components:

$$L_{total} = \lambda_1 L_{seg} + \lambda_2 L_{freq} + \lambda_3 L_{align} \tag{6}$$

where: $\lambda_1 = 1.0$; $\lambda_2 = 0.5$; and $\lambda_3 = 0.5$

By jointly optimizing these loss components, the EnAF model enhances the robustness of traffic violation segmentation, ensuring both accurate object delineation and frequency-domain consistency in real-world surveillance applications.

4. **Experiments.**

4.1. **Dataset.** To evaluate the performance of few-shot deep learning models, we utilize two datasets: COCO-20i and T-Traffic, each with distinct characteristics.

COCO [20] is derived from the MS COCO dataset, which contains over 80,000 images categorized into 80 classes. This dataset is particularly challenging due to its diverse object categories and variations in object appearances. To facilitate few-shot learning evaluation, we divide the dataset into four folds, each containing 20 distinct categories. The mean intersection over union (mIoU) is used as the primary evaluation metric. The

IoU for each category is computed from the confusion matrix, and the final mIoU is obtained by averaging the IoU values across all categories.

The T-Traffic dataset is specifically designed for traffic safety violation detection and is conceptually similar to COCO. It consists of 22,000 images collected from traffic surveillance cameras in Hanoi, Vietnam. These images are annotated with 9 distinct traffic violation categories. These categories include motorcyclists without helmets, two or more people on a motorcycle without helmets, overloaded motorcycles, motorcycles moving in the wrong direction, cars or buses in restricted lanes, illegal parking, and cyclist violations. The dataset structure follows the COCO framework to ensure compatibility with few-shot learning models. This enables direct performance comparison with COCO while addressing real-world traffic law enforcement challenges.

By leveraging these two datasets, we systematically evaluate the generalization capability of few-shot deep learning models in both generic object recognition and traffic safety violation detection. In our few-shot evaluation setup, we divided the 80 object categories of the COCO dataset into 4 folds, each containing 20 distinct classes. During training and testing, each fold is treated as a class group — C0 to C3 — corresponding to different splits of the dataset used in cross-validation. These class groups are not semantic categories (e.g., animals, vehicles) but rather represent partitions for evaluation. For example, C0 may contain classes like "person," "bicycle," etc., while C1 contains another disjoint group, and so on. In contrast, the T-Traffic dataset includes 9 real-world traffic violation categories (e.g., no helmet, wrong direction, overloaded motorcycles, etc.). For reporting clarity and to maintain alignment with our few-shot segmentation framework, we grouped these 9 violation types into 4 evaluation clusters (C0–C3), each containing 2–3 semantically related classes. For instance, C0 includes helmet-related violations, while C1 focuses on wrong-lane or direction violations. While both datasets are divided into C0–C3 for evaluation purposes, the COCO splits are for general object segmentation using arbitrary folds of standard object classes, whereas the T-Traffic splits are grouped based on traffic violation semantics. Thus, the Cn groups in COCO and T-Traffic are not directly comparable and serve different roles in the respective contexts.

4.2. **Experiment setup.** In our evaluation, we adopt a rigorous protocol to validate the EnAF model under the few-shot traffic violation recognition task. All input images are resized to 448×448 to balance computational efficiency and spatial detail retention. For the COCO dataset, we utilize ResNet50 as the backbone to leverage its deep hierarchical features, while T-Traffic, a domain-specific traffic violation dataset, employs ResNet101 for its robust mid-level feature extraction. Both models undergo training for 48 epochs, with COCO using a batch size of 24 and a learning rate of 5×10^{-4}, whereas T-Traffic adopts a larger batch size of 18 with the same learning rate to accommodate its distinct data distribution. During meta-testing, we evaluate generalization by randomly sampling 1,200 episodes (support-query pairs) from each test set, assessing performance in a 5-shot setting to mirror real-world scarcity of annotated violations. The framework integrates the Cross-granularity Frequency-aware Module (CFM), which decomposes RGB features from backbone layers (3, 9, 12) into multi-frequency components, and the CLIP-guided Spatial-adapter Module (CSM), aligning text embeddings with visual features to refine pseudo-mask generation. Experiments are conducted on a system equipped with dual RTX 3090 GPUs (64GB VRAM), 128GB RAM, and an Intel i5 10th-gen CPU, ensuring efficient large-scale tensor operations and end-to-end training. This setup ensures robust validation of EnAF's ability to fuse adaptive frequency analysis with semantic prior knowledge for few-shot segmentation.

Our experiments were conducted to address two key research questions (RQs):

TABLE 1. Results(%) of various methods on COCO dataset and T-Traffic dataset.

Method	Backbone	COCO dataset					T-Traffic dataset				
		C0	C1	C2	C3	mIoU	C0	C1	C2	C3	mIoU
DRNet [21]	Resnet50	0.477	0.517	0.470	0.493	0.490	0.665	0.671	0.657	0.651	0.661
CANet [22]		-	-	-	-	0.516	0.682	0.693	0.681	0.697	0.688
AFANet [18]		0.410	0.495	0.430	0.469	0.451	0.615	0.608	0.593	0.602	0.6045
HPA [23]		0.455	**0.554**	0.489	0.502	0.500	0.643	0.658	0.627	0.659	0.647
EnAF (ours)		**0.492**	0.531	**0.518**	**0.527**	**0.517**	**0.702**	**0.725**	**0.713**	**0.736**	**0.719**
DRNet [21]	Resnet101	0.520	0.545	0.479	0.498	0.511	0.702	0.746	0.684	0.691	0.706
MGNet [24]		0.452	0.469	0.441	0.438	0.450	0.626	0.617	0.609	0.613	0.616
AFANet [18]		0.442	0.526	0.457	0.503	0.482	0.602	0.695	0.618	0.683	0.6495
HPA [23]		0.492	0.578	0.520	0.506	0.524	0.647	0.709	0.651	0.667	0.6685
EnAF (ours)		**0.556**	**0.561**	**0.537**	**0.528**	**0.5455**	**0.726**	**0.753**	**0.732**	**0.758**	**0.744**

- **RQ1:** How does the EnAF model perform compared to other models with the same concept?
- **RQ2:** How does the EnAF model predict segmentation and detect traffic violations in real-world scenarios?

4.3. **Performance Compare (RQ1).** The performance comparison table 1 evaluates different methods on the COCO and T-Traffic datasets using two backbone architectures, ResNet50 and ResNet101. The assessment focuses on Intersection over Union (IoU) for four class groups (C0–C3), which are derived from a total of nine labels, as well as the mean IoU (mIoU). The proposed EnAF model (ours) is benchmarked against state-of-the-art methods, including DRNet, CANet, AFANet, MGNet, and HPA. The results demonstrate EnAF's superior performance, particularly in real-world scenarios such as traffic monitoring.

On the COCO dataset, EnAF achieves significant improvements over other models. With ResNet50, it reaches a mIoU of 51.7%, outperforming HPA (50.0%) and CANet (51.6%). It particularly excels in C0 (49.2), C2 (51.8), and C3 (52.7), showing its ability to capture diverse features effectively. When using ResNet101, EnAF further improves to 54.55% mIoU, surpassing HPA (52.4%) and DRNet (51.1%), demonstrating its ability to leverage deeper architectures for better generalization.

On the T-Traffic dataset, which consists of real-world traffic surveillance images, EnAF significantly outperforms competing models, achieving mIoU scores of 71.9% (ResNet50) and 74.4% (ResNet101), representing a 3–8% improvement over other methods. The class-wise IoU results further highlight its advantages. With ResNet50, EnAF achieves C0 (70.2), C1 (72.5), C2 (71.3), and C3 (73.6). When switching to ResNet101, these values increase to C0 (72.6), C1 (75.3), C2 (73.2), and C3 (75.8). These results confirm EnAF's capability to process real-world traffic data efficiently, even with limited training data.

Comparing different methods, DRNet and AFANet exhibit lower performance across both datasets, especially on COCO, where their mIoU scores remain around 0.45–0.49 with ResNet50. This indicates that these models may struggle to learn adaptive frequency features, which are crucial for accurate object recognition. HPA performs reasonably well on COCO (mIoU = 50.0% with ResNet50), but its performance is inconsistent on T-Traffic (mIoU between 64.7% and 66.85%), indicating weak generalization to real-world conditions. CANet achieves an overall mIoU of 51.6% on COCO, but its lack of class-wise results makes comprehensive evaluation difficult.

The choice of backbone network significantly influences performance. Using ResNet101 consistently enhances results compared to ResNet50. For EnAF, the mIoU improves from 51.7% to 54.55% on COCO and from 71.9% to 74.4% on T-Traffic. DRNet also

Backbone	mIoU (T-Traffic)	FLOPs (G)	FPS (batch=1)
ResNet50	71.9%	84.2	21.3 FPS
ResNet101	74.4%	118.6	14.8 FPS

Table 2. Performance comparison of different backbone architectures

Model Variant	CFM	CSM	mIoU (%)
Baseline (no modules)	×	×	66.2
+ CFM only	✓	×	70.1
+ CSM only	×	✓	68.5
Full EnAF (CFM + CSM)	✓	✓	74.4

Table 3. Ablation study of different model variants with CFM and CSM modules

sees an improvement, with its mIoU increasing from 49.0% to 51.1% (COCO) and from 66.1% to 70.6% (T-Traffic). These findings emphasize that a more powerful backbone can substantially enhance model performance.

EnAF provides several key advantages. Its few-shot learning capability enables it to maintain high accuracy even with limited training data, making it particularly well-suited for datasets like T-Traffic. Moreover, its ability to optimize frequency-based feature extraction helps it distinguish important pixels that impact recognition quality, especially in scenarios involving motion blur and occlusions. This is evident in T-Traffic, where EnAF achieves class-wise IoU values above 0.7 for C0 and C3. Furthermore, EnAF generalizes well across both general-purpose datasets (COCO) and specialized datasets (T-Traffic), proving its adaptability to various real-world applications.

Despite its strong performance, EnAF has some limitations. Certain methods, such as CANet, do not provide detailed class-wise results, reducing the transparency of comparisons. Additionally, the use of ResNet101 increases computational complexity, which could impact real-time deployment speed. Finally, while T-Traffic provides valuable real-world data, its scale and diversity need expansion for a more comprehensive evaluation. The results confirm that EnAF is a highly effective model for deep learning-based object recognition, particularly in real-world applications like traffic monitoring. By leveraging ResNet101, EnAF not only enhances accuracy but also maintains stable performance across different datasets. Its strong adaptability, even with limited training data, makes it a promising candidate for intelligent traffic surveillance systems.

We evaluated the FLOPs and inference speed (FPS) of the EnAF model using both ResNet50 and ResNet101 backbones on a system with dual RTX 3090 GPUs show in Table 2. While ResNet101 offers higher segmentation accuracy (+2.5% mIoU), it incurs approximately 41% more FLOPs and 30% drop in FPS compared to ResNet50. This trade-off highlights that EnAF with ResNet50 remains a viable option for real-time or resource-constrained deployments, while ResNet101 provides superior performance for high-precision scenarios.

4.4. **Qualitative Study (RQ2).** The results in Table 3 show that both CFM and CSM contribute significantly to the performance of the model. The CFM module provides the largest gain (+3.9%), highlighting the importance of frequency-aware feature learning in handling image quality variations. The CSM module also boosts performance (+2.3%) by integrating semantic context. The full model, combining both modules, achieves the highest accuracy, validating their complementarity.

The analysis of the T-Traffic dataset demonstrates the effectiveness of the EnAF model in detecting and segmenting various traffic violations using surveillance camera footage,

FIGURE 2. The detection and segmentation results of samples on T-Traffic dataset.

show in Figure 2. The framework successfully identifies multiple types of violations, including wrong direction driving, bulky goods transportation, illegal unloading, helmet violations, improper parking, missing mirrors, phone usage while riding, lane violations, and covered license plates. Each detected infraction is accurately segmented, with a highlighted region outlining the object of interest. Additionally, zoomed-in visualizations of the detected objects provide a clearer perspective of the violations.

The segmentation masks, marked in blue, effectively delineate motorcyclists, vehicles, and objects involved in infractions. The model's ability to detect small but critical features, such as helmets, mirrors, and mobile phones, highlights its robustness in real-world traffic scenarios. The multi-scale frequency-aware segmentation mechanism, employed by EnAF, enhances its ability to distinguish fine details, even in challenging conditions such as occlusions, varying lighting, and complex urban traffic environments. By decomposing RGB domain features into high-frequency and low-frequency components, the system optimally realigns spatial structures, improving segmentation accuracy.

Furthermore, the multi-stage processing architecture of EnAF enables efficient feature extraction, making it well-suited for traffic enforcement applications. The framework's ability to generalize across different traffic scenarios underscores its potential for deployment in intelligent traffic monitoring systems. The high detection accuracy of EnAF ensures that safety violations are effectively identified, providing a foundation for automated law enforcement and urban traffic management. Through its structured frequency-aware learning approach, the framework not only enhances object differentiation but also contributes to improving road safety and regulatory compliance in real-world traffic environments.

5. **Conclusions.** In this paper, the results from the T-Traffic dataset demonstrate the effectiveness of the EnAF model in detecting and segmenting various traffic violations with high accuracy. By leveraging multi-scale frequency-aware feature extraction, EnAF successfully identifies critical infractions such as wrong direction driving, bulky goods transportation, helmet violations, improper parking, missing mirrors, and mobile phone usage while riding. The integration of high-frequency and low-frequency decomposition allows the model to enhance object differentiation, even under challenging real-world conditions such as occlusions, varying lighting, and complex traffic environments. The

structured multi-stage processing approach ensures that segmentation is both precise and adaptable across different traffic scenarios. These findings highlight the potential of EnAF as a reliable and scalable solution for intelligent traffic monitoring and automated law enforcement, contributing to improved road safety, regulatory compliance, and urban traffic management.

REFERENCES

[1] Mohammed, Ali Ahmed, et al. "A review of traffic accidents and related practices worldwide." The Open Transportation Journal 13.1 (2019).

[2] Phat Nguyen Huu* and Thuong Nguyen Thi Mai "Proposing a fruit classification system using Tensorflow model" Journal of Information Hiding and Multimedia Signal Processing, Vol. 14, No. 1, pp. 10-19, March 2023.

[3] Budi Setiyono, Dwi Ratna Sulistyaningrum, Ario Fajar Pratama, Ridho Nur Rohman Wijaya "A Modification of the Temporal Group Attention Method on Super-Resolution Video for Vehicle Number Plate Detection" Journal of Information Hiding and Multimedia Signal Processing, Vol. 15, No. 3, pp. 156-165, September 2024.

[4] Kuznetsova, Alina, et al. "The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale." International journal of computer vision 128.7 (2020): 1956-1981.

[5] Gharoun, Hassan, et al. "Meta-learning approaches for few-shot learning: A survey of recent advances." ACM Computing Surveys 56.12 (2024): 1-41.

[6] Parnami, Archit, and Minwoo Lee. "Learning from few examples: A summary of approaches to few-shot learning." arXiv preprint arXiv:2203.04291 (2022).

[7] Aboah, Armstrong, et al. "Real-time multi-class helmet violation detection using few-shot data sampling technique and yolov8." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

[8] Ye, Tianpeng, et al. "FLAG: Few-shot latent Dirichlet generative learning for semantic-aware traffic detection." IEEE Transactions on Network and Service Management 19.1 (2021): 73-88.

[9] Katsamenis, Iason, et al. "A few-shot attention recurrent residual U-Net for crack segmentation." International Symposium on Visual Computing. Cham: Springer Nature Switzerland, 2023.

[10] Ma, Wengang, et al. "Few-shot abnormal network traffic detection based on multi-scale deep-CapsNet and adversarial reconstruction." International Journal of Computational Intelligence Systems 14.1 (2021): 195.

[11] Lang, Chunbo, et al. "Learning what not to segment: A new perspective on few-shot segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.

[12] Sun, Xian, et al. "Research progress on few-shot learning for remote sensing image interpretation." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021): 2387-2402.

[13] Xiao, Aoran, et al. "Cat-sam: Conditional tuning for few-shot adaptation of segment anything model." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024.

[14] Feng, Han Cong, Bin Tang, and Tao Wan. "Radar pulse repetition interval modulation recognition with combined net and domain-adaptive few-shot learning." Digital Signal Processing 127 (2022): 103562.

[15] Li, Xinyu, et al. "Supervised domain adaptation for few-shot radar-based human activity recognition." IEEE Sensors Journal 21.22 (2021): 25880-25890.

[16] Zhang, Junjie, et al. "Frequency-Aware Multi-Modal Fine-Tuning for Few-Shot Open-Set Remote Sensing Scene Classification." IEEE Transactions on Multimedia (2024).

[17] Su, Binyi, et al. "FSRDD: An efficient few-shot detector for rare city road damage detection." IEEE Transactions on Intelligent Transportation Systems 23.12 (2022): 24379-24388.

[18] Ma, Jiaqi, et al. "AFANet: Adaptive Frequency-Aware Network for Weakly-Supervised Few-Shot Semantic Segmentation." arXiv preprint arXiv:2412.17601 (2024).

[19] Zhu, Hegui, et al. "Few-shot fine-grained image classification via multi-frequency neighborhood and double-cross modulation." IEEE Transactions on Multimedia (2024).

[20] Sun, Bo, et al. "Fsce: Few-shot object detection via contrastive proposal encoding." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

[21] Gao, Guangyu, et al. "Drnet: Double recalibration network for few-shot semantic segmentation." IEEE Transactions on Image Processing 31 (2022): 6733-6746.

[22] Zhang, Chi, et al. ”Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
[23] Cheng, Gong, Chunbo Lang, and Junwei Han. ”Holistic prototype activation for few-shot segmentation.” IEEE Transactions on Pattern Analysis and Machine Intelligence 45.4 (2022): 4650-4666.
[24] Chang, Zhaobin, et al. ”MGNet: Mutual-guidance network for few-shot semantic segmentation.” Engineering Applications of Artificial Intelligence 116 (2022): 105431.