Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377-0430
Vol. 5, No. 2, 2025

A Multi-Scale Edge Feature Network for Robust Object Detection

Eamon Lindström
Department of Computer Science, Lund University, Lund, Sweden
eamon57@cs.lu.se

Abstract:

Object detection, a foundational task in computer vision, entails accurately identifying and localizing objects in images, which remains challenging due to issues like object occlusion and multiscale detection imbalance. This paper proposes the Multi-Scale Edge Feature Enhancement Network (MEFENet), a novel one-stage object detection framework designed to address these challenges. MEFENet introduces two key innovations: (1) the Multi-Scale Edge Feature Extraction (MEFE) structure, which fuses extracted edge features with multi-scale feature maps, enriching semantic representations to improve occluded object detection; and (2) the Receptive Field Enhancement (RFE) module, which refines feature semantics and mitigates multiscale detection imbalances. MEFENet leverages a residual network (ResNet) backbone and combines outputs from the Feature Pyramid Network (FPN) and MEFE structures, which are subsequently processed through the RFE module for enhanced semantic feature extraction. Extensive experiments on the PASCAL VOC 2007+2012 and Microsoft COCO datasets demonstrate that MEFENet achieves state-of-the-art detection accuracy, outperforming nine representative methods in key evaluation metrics. These results validate the effectiveness of the proposed innovations in addressing occlusion and multiscale detection challenges.

Keywords:

Deep Learning; Occlusion Object Detection; Multi-scale Object Detection.

1. Introduction

A fundamental task in computer vision is object detection, which not only has many practical applications but also is a crucial first step toward computer picture understanding and analysis [1,2]. The difficulty of the target detection task lies in the need to identify objects in an image and locate their positions precisely. The object detection task is usually decomposed into two subtasks: object localization and object classification [3,4]. In the target localization task, the precise position of the target in the image needs to be determined, taking into account the variability of the object and the randomness of the image, which requires algorithms with high accuracy and robustness. In the target classification task, there is a need to identify classes of objects in an image, as well as to distinguish similarities and differences between different targets [5-7]. These pose challenges to the object detection task.

In the anchor-based object detection method, the two-stage approach first extracts the Region of Interest (RoI) using selective search and then performs prediction, while the one-stage approach forgoes the use of RoI for better speed performance. Usually, the two-stage approach has advantages in detection accuracy, while the one-stage approach has better real-time performance. However, since the one-stage method gives up selective search, a large number of negative samples are generated in the one-stage method. Meanwhile, the missing information and confusion of the occluded objects themselves make the number of valid positive anchors in the training network small.

However, numerous one-stage methods have made great progress, but they still have shortcomings

in occlusion object detection and multiscale object detection imbalance. To alleviate these problems, in this paper, we propose the multiscale edge feature enhancement network.MEFENet based on the one-stage object detection method, the main innovation points of this paper are as follows:

1. To improve the detection performance of occluded objects, we propose the edge feature extraction structure (MEFE). the MEFE can fuse the extracted edge features with the multi-scale feature map, which enriches the semantic features of the MEFENet and improves the detection performance of the occluded objects.

2. To alleviate the multiscale object detection imbalance, we propose the receptive field enhancement (RFE) module, which further acquires the edge features extracted from the MEFE structure and enriches the feature semantic information, thereby alleviating the multi-scale object detection imbalance and improving the multi-scale object detection performance of MEFENet.

3. Our proposed multi-scale edge feature enhancement network (MEFENet) has good experimental results on PASCAL VOC 2007+2012 and Microsoft COCO datasets, which effectively improves the performance of object detection.

2. Related Work

Object detection has achieved remarkable progress due to advancements in deep learning, especially in tackling challenges such as multiscale detection imbalance and occlusion. Classic convolutional neural networks, such as VGGNet and ResNet, have significantly influenced feature extraction methodologies. He et al. [8] evaluated the performance of VGG19 in handling complex visual data and highlighted its limitations in multiscale feature fusion. To address these limitations, reinforcement learning methods like those proposed by Huang et al. [9] introduce adaptive strategies for dynamic feature optimization, indirectly inspiring solutions such as MEFENet's multiscale feature enhancement.

Occlusion remains a major challenge in object detection. Zheng et al. [10] employed fully convolutional networks (FCNs) for high-precision medical image analysis, emphasizing spatial consistency in feature learning. Inspired by such advancements, MEFENet's Multi-Scale Edge Feature Extraction (MEFE) structure enhances edge features and integrates them with multiscale feature maps to improve occlusion handling. Furthermore, receptive field optimization plays a critical role in detecting objects of varying scales. Yan et al. [11] explored neural architecture search (NAS) to optimize receptive fields, a concept reflected in MEFENet's Receptive Field Enhancement (RFE) module, which refines semantic features to mitigate multiscale detection imbalance.

Recent advancements in contextual learning and few-shot adaptation have also influenced the development of MEFENet. Hu et al. [12] proposed adaptive weight masking in conditional GANs for few-shot learning, demonstrating the importance of feature alignment in scenarios with limited positive samples. Liang et al. [13] highlighted the role of contextual learning in combining local and global features for sensitive information detection, a technique mirrored in MEFENet's multiscale feature integration module.

Graph-based learning methods, as demonstrated by Mei et al. [14] and Gao et al. [15], have shown promise in modeling complex relationships, such as disease risk assessment and sequential visit prediction. These approaches provide insights into the potential of contextual relationships, which MEFENet leverages to enhance multiscale feature representation. Furthermore, Liu et al. [16] investigated calibration learning for few-shot tasks, emphasizing the importance of feature generalization in novel scenarios. Lastly, Song and Liu [17] explored norm-based feature selection for improving the robustness of extracted features, which aligns with MEFENet’s emphasis on efficient feature fusion.

2.1 Anchor-free Method

2.1.1 Key Point-based Method

Heatmap is used by the keypoint-based detector to forecast key points, which are then grouped to produce bounding boxes.CornerNet [18] locates an object's top left and bottom right corners and embeds them in the feature space of an abstract representation. With the addition of a centroid detection branch and centroid verification, CenterNet [19] significantly boosts CornerNet's speed. After corner point prediction, CentripetalNet [20] suggests using a centripetal displacement module to combine corner points, which lowers the false detection rate while maintaining the recall rate.

2.1.2 Center-based Method

In the Center-based method, YOLOv1 [21] divides the image into cells and directly predicts the objects whose object centroids fall within the cells. in order to provide end-to-end detection, DenseBox [22] incorporates fully convolutional networks (FCN) to the field of object detection. This directly regresses the confidence level and relative position of object occurrence. UnitBox [23] regresses the four boundaries collectively using intersection over union (IoU) loss. The recall of these detectors is low because there are not a lot of positive samples. FCOS [24] uses the object bounding box to treat all points as positive samples in order to address this issue. It finds all points that are positive and measures their distance from the boundary of the enclosing box.

2.2 Anchor-based Method

2.2.1 Two-stage Method

The two-stage method, which was evolved from the R-CNN [25] family of methods, extracts regions of interest (RoI) using a selective search strategy before classifying and regressing them. a region proposal network (RPN) is used by Faster R-CNN [26] to produce RoI. In order to improve detection, Mask R-CNN [27] discovered pixel bias in the RoI Pooling layer and employed bilinear interpolation to swap it out for the RoI Align layer. Additionally, the segmentation method used by its mask head is top-down.

2.2.2 One-stage Method

To complete the object identification task using the one-stage method, the pre-defined anchor is directly classified and regressed. SSD [28] achieves the object detection by classifying and regressing the anchor with various step sizes using feature maps from several convolutional layers. To further extract the depth features for regression and classification based on SSD, DSSD [29] uses the residual module. In order to address the issue of gradient disappearance and explosion while minimizing hyperparameters, YOLOv2 [30] adds a batch normalization (BN) layer after each convolutional layer. YOLOv2 also further takes into account fine-grained features. And with accuracy comparable to ResNet-101, the DarkNet-53 developed by YOLOv3 [31] significantly decreases the number of network layers. The preceding strategies, however, frequently lead to a situation where the training is dominated by negative samples because of the imbalance between the quantity of negative and positive samples. RetinaNet [32] proposes the usage of Focal Loss to address the sample balance issue in order to address the positive and negative sample imbalance.

3. Our Proposed Method

3.1 Network Structure Design

Our proposed multi-scale edge feature enhancement network (MEFENet) object detection method uses ResNet-50 as the backbone, and the features extracted from the backbone are first divided into two branches for processing. The first branch is connected to the feature pyramid networks (FPN) to further extract multi-scale features. The other branch feeds the backbone features into the multi-scale edge feature extraction (MEFE) structure. Then, the edge features are fused with the multi-scale features output from the feature pyramid and sent to the receptor field enhancement (RFE) module. After that, the semantic information is further enhanced using the Receptor Field Enhancement module and sent to the detection head. The structure of multi-scale edge feature enhancement network (MEFENet) is shown in Figure 1.

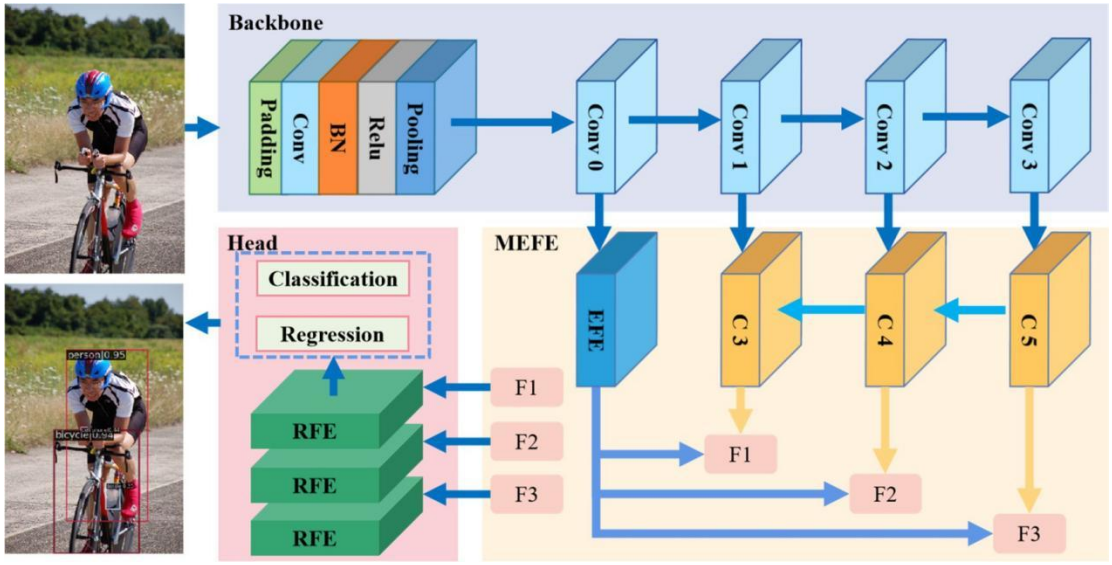


Figure 1. Architecture diagram of multi-scale edge feature enhancement network

3.2 Multi-scale Edge Feature Extraction

Detection of occluded objects in object detection has been a challenging problem, and the current methods for this problem have some shortcomings, mainly in the missing information caused by occlusion and the information confusion caused by occlusion. In addition, due to the relatively small dataset of occluded objects and the strong scene-specific and object-specific nature of the object occlusion problem, which varies greatly from scene to scene and from object to object, a large and diverse dataset is required for training and evaluation. For this reason, in the absence of publicly available datasets for obscured objects, it becomes a better choice based on practical considerations to extract as much feature information as possible in the detection network using existing datasets. Therefore, we propose the multi-scale edge feature extraction (MEFE) module.

Our proposed multi-scale edge feature extraction (MEFE) structure consists of multiple-scale edge feature extraction (EFE) modules. The feature maps of the first branch of the backbone network extraction are fed into the multi-scale edge feature extraction module, which consists of Laplace operator and down-sampling structure. The multi-scale edge feature extraction module is shown in Figure 2.

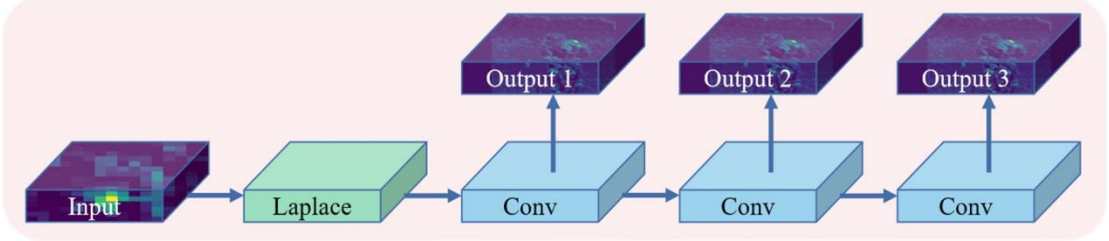


Figure 2. Architecture diagram of multi-scale edge feature extraction

The Laplace operator can be used to enhance edge information in deep learning object detection. In object detection tasks, edge information is often very important because the shape and contour of an object can be described by edges. By applying the Laplace operator, the edge information can be enhanced to help deep learning models detect objects better. Specifically, a Laplacian filter can be added to a convolution neural network (CNN) to enhance the edge information in the feature map of that layer. This filter performs a convolution operation on the feature map, resulting in a feature map with enhanced edge information. This feature map can be fed into subsequent convolutional or fully connected layers for further processing, thus improving the performance of the model. An important advantage of using the Laplace operator is that it can enhance the edge information without adding much computational burden. Because the Laplace operator only performs convolution operations on the feature map, it does not incur much computational overhead. Overall, the Laplace operator works by enhancing the edge information in the feature map, thus helping deep learning models to better detect objects. This method can be used in combination with other methods of enhancing edge information to further improve the performance of the model.

3.3 Receptor Field Enhancement

In object detection, objects at multiple scales need to be detected because different objects may have different scales in the image. However, the number of objects at different scales may vary greatly, leading to the imbalance problem of multi-scale object detection. The imbalance problem of multi-scale object detection is alleviated by using multi-scale feature fusion, but it is still insufficient. Therefore, we propose the receptor field enhancement (RFE) module in order to further enhance the semantic information of edge features while alleviating the multi-scale object detection imbalance problem.

Our proposed receptor field enhancement (RFE) module consists of a maximum pooling, a down-sampling structure, a null convolution and a residual structure. The receiver field enhancement module feeds the acquired feature maps of the second branch of the backbone network extraction into two separate null convolutions with different null rates. Also, the feature maps of the second branch extracted from the backbone network are fed into the maximum pooling module after using down-sampling for feature map resizing. Finally, the residual structure is combined to reduce overfitting while mitigating gradient disappearance or gradient explosion. The receptor field enhancement module is shown in Figure 3.

The recepor field enhancement (RFE) module is designed to further process the edge features acquired by the multi-scale edge feature extraction (MEFE) module to obtain richer semantic information. First, the perceptual field enhancement module can enhance the model's ability to perceive the object and enable the model to detect the object better. By increasing the perceptual range of the model on the input image, the contextual information of the object object can be captured more accurately, and the accuracy and recall of object detection can be improved. Second, the perceptual field enhancement module can reduce unnecessary computations in object detection, thus speeding up object detection. In traditional object detection methods, object detection is usually performed by sliding window, which requires one detection for each position of the image and is computationally intensive and slow. By enhancing the perceptual field of the model, unnecessary calculations can be reduced, thus speeding up the object detection. After that, the perceptual field enhancement module can improve the robustness of the model and make the model better adaptable to noise and image changes. In practical applications, images may be disturbed by noise, image blur, lighting changes, etc., which may lead to a decrease in the accuracy of object detection. By enhancing the perceptual field of the model, the adaptability of the model to these factors can be improved, and thus the accuracy of object detection can be improved. Finally, the perceptual field enhancement module can make the model more robust to changes in the input image, thus reducing the risk of overfitting. The main reason for overfitting is that the model relies too much on specific image features in the training set and does not adapt well to new images. The perceptual field enhancement

structure allows the model to learn image features more comprehensively to mitigate overfitting.

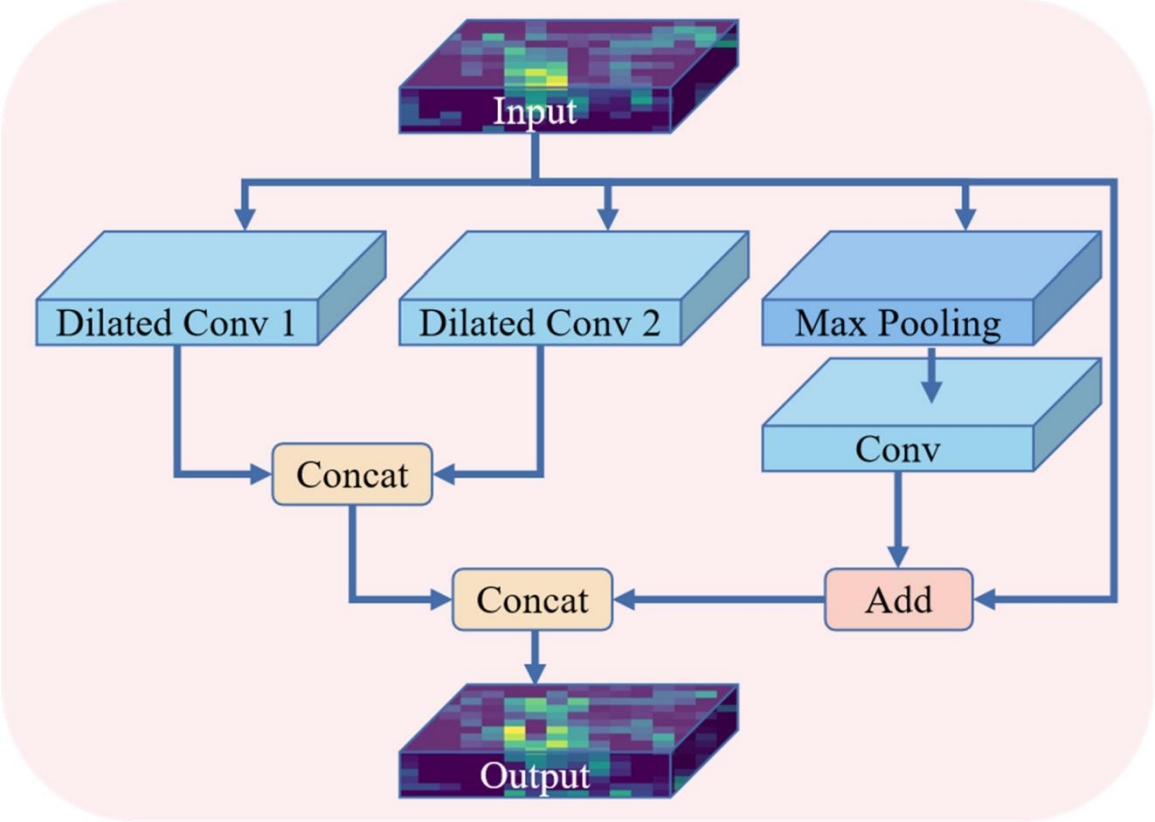


Figure 3. Architecture diagram of receptor field enhancement

4. Experiment

4.1 Datasets

The PASCAL VOC [33] and Microsoft COCO [34] datasets are standard datasets in the field of object detection. The experiments of our proposed multi-scale edge feature enhancement network (MEFENet) object detection method are also based on this.

Because the images in the PASCAL VOC 2007 datasets and the PASCAL VOC 2012 datasets are mutually exclusive, many object detection methods combine the PASCAL VOC 2007 datasets and the PASCAL VOC 2012 datasets for training and evaluation on the PASCAL VOC 2007 datasets. There are 16551 training images with 40,058 object objects at the same time after the merging. The evaluation images have 4952 images, including 12032 object objects. For the evaluation metric of the model on the PASCAL VOC 2007+2012 datasets, we used the mean average precision (mAP).

Since, the Microsoft COCO 2017 dataset has more images and object objects than Microsoft COCO 2014, which makes the Microsoft COCO 2017 datasets more challenging, we select the Microsoft COCO 2017 datasets. Microsoft COCO 2017 datasets the training set has more than 118,000 images, the number of object annotations reaches 910670, and the number of evaluations set images is 5000. We also use the evaluation criteria of Microsoft COCO, such as: average accuracy (AP), average accuracy of small-sized objects (AP_S), average accuracy of medium-sized objects (AP_M), and average accuracy of large-sized objects (AP_L).

4.2 Experimental Setup

Our proposed multi-scale edge feature enhancement network (MEFENet) approach for object detection is implemented through MMDetection [35], a toolbox based on Pytorch implementation of object detection. In ablation experiments, quantitative experiments, and qualitative experiments, we

use 1 GeForce RTX 3090 on training and prediction.

The experimental parameters of our proposed MEFENet method on the Microsoft COCO 2017 datasets are set as follows: the backbone is ResNet-50; meanwhile, the maximum size of the input image is rescaled to 1333*800 without changing the aspect ratio; the optimizer is SGD, the learning rate is 2*10-2 and weight decay is 10-4; the method is trained with other comparative representative methods for 12 epochs, and the batch size is also set to 8.

The experimental parameters of our proposed MEFENet method on the PASCAL VOC2007+2012 datasets are set as follows: the backbone is ResNet-50; the image input size is 1000*600; the optimizer is SGD, the learning rate is 2*10-2, and the weight decay is 10-4; the method is trained for 12 Epoch with other comparative representative methods, and the batch size is also set to 16.

4.3 Quantitative Analysis of Ablation Experiments

Our ablation experiments are based on the Microsoft COCO 2017 datasets with a ResNet-50 backbone, using a 12-epoch training scheme. Also, the maximum size of the input image is rescaled to 1333*800 without changing the aspect ratio. in addition, the ablation experiments in this subsection are based on the multi-scale edge feature enhancement network (MEFENet) object detection method to remove the multi-scale edge in addition, the ablation experiment in this subsection is based on the baseline method obtained by removing the multi-scale edge feature extraction (MEFE) module and the receptor field enhancement (RFE) module from the multi-scale edge feature enhancement network (MEFENet) object detection method.

Our proposed multi-scale edge feature enhancement network (MEFENet) object detection method is designed to use the multi-scale edge feature extraction (MEFE) module to The MEFE module is used to obtain more semantic information about the edge features, and then the RFE module is used to enhance the obtained edge feature information. The richer edge semantic information is obtained through enhancement to alleviate the problem of missing information of occluded objects and achieve higher detection performance of occluded objects.

First, to verify that our proposed multi-scale edge feature extraction (MEFE) module can obtain edge feature information to alleviate the problem of missing information of occluded objects. The experimental results using Baseline+MEFE module show that the MEFE module can help Baseline to obtain more edge feature information and thus improve the detection performance with an average precision (AP) of 39.3%. The AP reaches 39.3%, which is 0.7% higher than that of Baseline, and the other five detection metrics of Microsoft COCO 2017 datasets are also improved, as shown in the third row of Table 1.

Second, to demonstrate that the receptor field enhancement (RFE) module can enhance the acquired feature semantic information to improve the detection performance. The experiments using Baseline+RFE module show that RFE module can enhance the detection performance by enhancing the semantic information of the features, and the AP of Baseline+RFE module reaches 39.1%. The AP of Baseline+RFE module reaches 39.1%, which is 0.5% higher than that of Baseline, and the detection results of other indicators are also improved, as shown in the fourth row of Table 1.

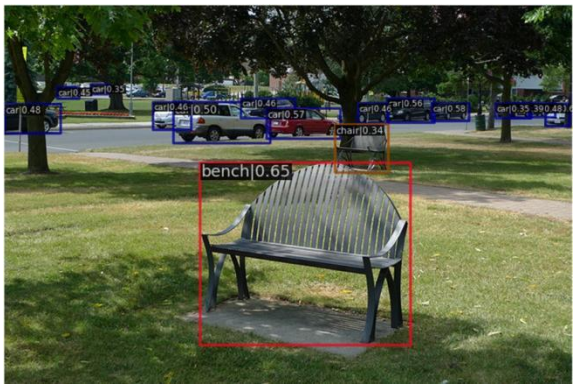
Finally, to confirm that the receptor field enhancement (RFE) module can enhance the semantic information of edge features obtained by the multi-scale edge feature extraction (MEFE) module and thus improve the detection performance of occluded objects. The experimental results using Baseline+MEFE+RFE module show that the perceptual field enhancement module further enhances the edge feature information to improve the detection accuracy, and the AP of Baseline+MEFE+RFE reaches 39.7%. Compared with Baseline, the average accuracy of Baseline+MEFE+RFE improved by 1.1%, and 1.3%, 1.6%, and 2.1% in the average accuracy (AP_S) of small objects, average accuracy (AP_M) of medium-sized objects, and average accuracy (AP_L) of large-sized objects, respectively. The analysis of quantitative experimental results demonstrates the effectiveness of our proposed multi-scale edge feature enhancement network (MEFENet) object detection method, as shown in the fifth

row of Table 1 (the bolded font is the highest detection accuracy in this category).

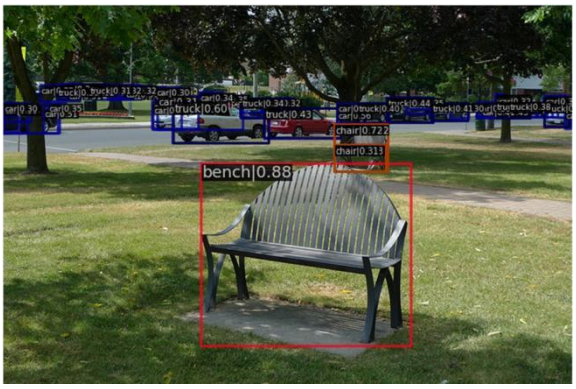
Table 1. Quantitative results of ablation experiment

method	AP	AP50	AP75	AP_S	AP_M	AP_L
Baseline	38.6	56.4	41.7	21.6	42.4	49.0
+MEFE	39.3	56.8	42.9	22.7	43.4	50.9
+RFE	39.1	56.5	42.2	22.3	43.6	50.5
+MEFE+RFE	**39.7**	**56.9**	**43.1**	**22.9**	**44.0**	**51.1**

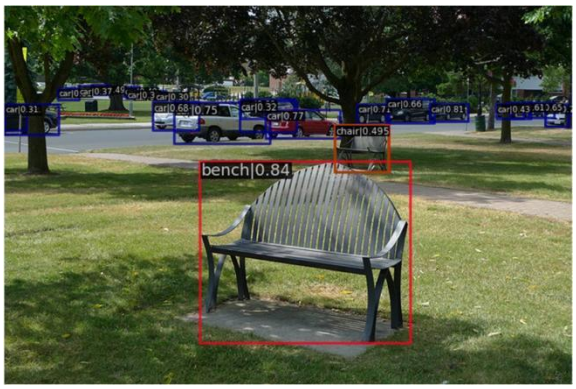
4.4 Qualitative Analysis of Ablation Experiments



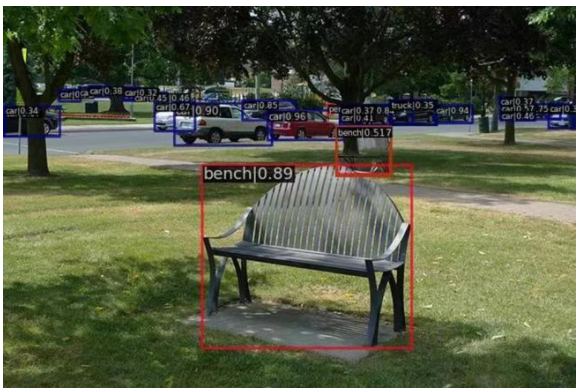
(a) Baseline



(b) Baseline+MEFE



(c) Baseline+RFE



(d) Baseline+MEFE+RFE

Figure 4. Visualization of ablation results

To demonstrate the ability of the multi-scale edge feature enhancement network (MEFENet) object detection method to enhance the detection of occluded objects and alleviate the imbalance problem of multi-scale object detection, the qualitative analysis of the ablation experiment is shown in Figure 4. To better demonstrate the feature extraction capability of the multi-scale edge feature extraction (MEFE) module and the semantic enhancement capability of the receptor field enhancement (RFE) module, we have selected a number of objects in the PASCAL VOC 2007+ 2012 datasets with the presence of occlusion.

As shown in Figure 4, Figure (a) shows the visualization results of Baseline, where there are missed

objects due to object occlusion in the upper middle of the image, while the confidence level of detected objects of Baseline is generally low, and the confidence level of occluded objects and normal objects is not reasonable. Figure (b) shows the visualization results of Baseline+MEFE structure. Due to the use of edge feature information, the problem of missed detection of occluded objects is improved, and the confidence level of each object is significantly improved. However, the bench in front of the large tree in the middle of the image is mistakenly detected as a chair. The analysis suggests that the multi-scale edge feature extraction (MEFE) module highlights the edge feature information of the bench, while the semantic information for performing the classification task is not rich enough. Figure (c) uses the Baseline+RFE structure, and similar to the results in Figure (b), the Bench in front of the tree is mistakenly detected as a Chair, which also has shortcomings. Finally, the Baseline+MEFE+RFE structure achieves more satisfactory detection results, as shown in Figure (d), the confidence level of each object is significantly improved, and the bench is no longer mis detected.

4.5 Quantitative Analysis of Comparative Experiments

To demonstrate the effectiveness of our proposed multi-scale edge feature enhancement network (MEFENet) object detection method, in this subsection, we compare MEFENet with nine other representative methods on the PASCAL VOC2007+2012 datasets and Microsoft COCO 2017 datasets, respectively, to complete quantitative experiments.

Table 2. Quantitative experimental results of Microsoft COCO 2017 datasets

method	Backbone network	FPN	AP	AP50	AP75	AP_S	AP_M	AP_L
RetinaNet	ResNet-50	√	36.2	55.1	38.7	20.4	39.8	46.8
FSAF	ResNet-50	√	37.0	56.2	39.4	20.3	40.1	47.8
Repponits	ResNet-50	√	37.4	56.8	40.3	21.9	41.4	48.3
FCOS	ResNet-50	√	36.9	45.8	39.3	20.7	40.1	47.2
ATSS	ResNet-50	√	38.6	56.4	41.7	21.6	42.4	49.0
Foveabox	ResNet-50	√	35.5	54.9	37.8	19.8	39.1	46.1
GFL	ResNet-50	√	39.6	**57.3**	42.7	21.8	43.5	**51.8**
VFNet	ResNet-50	√	37.5	53.9	40.5	21.0	41.0	49.0
Free Anchor	ResNet-50	√	38.2	56.7	40.7	20.8	41.6	49.8
MEFENet	ResNet-50	√	**39.7**	56.9	**43.1**	**22.9**	**44.0**	51.1

The quantitative experimental results based on the Microsoft COCO 2017 datasets are shown in Table 2 (the bolded font in the table is the highest detection accuracy for this category). Our proposed multi-scale edge feature enhancement network (MEFENet) object detection method compares with nine other representative methods, and our proposed MEFENet object detection method on the Microsoft COCO 2017 datasets which achieved the highest experimental results for four of the six evaluation metrics. According to the quantitative experimental results of the Microsoft COCO 2017 datasets, our proposed MEFENet for object detection method achieves an AP of 39.7%, which is the highest detection accuracy among the 10 object detection methods. It also achieves the highest detection accuracy in the evaluation metrics of AP_S for small-sized objects and average accuracy (AP_M) for medium-sized objects, which are 1% and 0.5% higher than the second place in each evaluation metric, respectively. Our suggested object detection strategy for MEFENet is proven to be effective by quantitative experimental findings using the Microsoft COCO 2017 datasets.

The results of quantitative experiments based on the PASCAL VOC 2007+2012 datasets are shown in Table 3 (the bolded font in the table is the highest detection accuracy for this category). Our proposed multi-scale edge feature enhancement network (MEFENet) for object detection method achieves 79.08% in category mAP compared with the other nine representative methods. Meanwhile, the detection results on 9 out of 20 categories (bird, boat, cup, car, cat, chair, horse, human, and potted plant) in the PASCAL VOC 2007+2012 datasets are the best detection results among the 10 object detection methods, which verifies the effectiveness of MEFENet.

Table 3. Quantitative experimental results of PASCAL VOC 07+12 datasets (a)

Method	mAP	plane	bike	bird	boat	cup	bus	car	cat	chair
RetinaNet	**79.13**	**87.2**	**86.2**	78.1	66.4	71.0	84.7	87.7	88.4	62.9
FSAF	76.31	79.3	79.3	76.0	65.1	67.6	83.1	86.7	87.1	59.8
Repponits	79.47	83.5	82.4	77.1	72.4	**71.6**	85.1	87.8	88.3	63.4
FCOS	71.59	78.5	78.7	68.3	61.8	57.6	78.0	82.2	83.0	54.8
ATSS	77.77	84.7	81.9	76.8	67.9	69.5	**85.4**	86.4	88.1	61.7
Foveabox	76.67	79.8	80.2	77.0	66.9	66.7	82.5	86.9	87.3	62.1
GFL	77.04	85.4	83.6	76.1	63.9	67.6	82.2	86.5	86.9	59.6
VFNet	77.83	83.1	84.3	76.7	68.4	69.5	84.5	86.7	87.3	61.3
Free Anchor	78.16	85.0	83.6	76.0	65.5	69.7	85.4	86.9	87.6	62.5
MEFENet	79.08	83.5	84.4	**79.3**	**73.0**	**71.6**	84.1	**87.9**	**88.4**	**63.9**

Continued Table 3 (b)

Method	cow	table	dog	horse	mbike	human	plant	sheep	sofa	train	TV
RetinaNet	85.2	**72.6**	85.8	85.7	82.7	84.1	53.4	82.9	**77.0**	82.9	**77.7**
FSAF	83.2	69.5	85.3	85.1	81.9	84.4	48.2	76.3	71.6	82.8	74.0
Repponits	86.3	75.7	**87.5**	85.8	**84.1**	83.8	50.7	84.0	76.2	**86.3**	77.4
FCOS	80.2	65.8	80.4	78.4	77.4	76.5	41.4	74.5	66.9	81.1	66.2
ATSS	**86.4**	72.3	85.1	85.4	80.2	83.1	48.3	81.1	72.3	81.7	77.1
Foveabox	85.6	69.4	85.1	85.9	78.9	84.4	48.8	79.1	71.2	79.4	76.5
GFL	83.4	72.8	83.9	84.9	83.2	83.3	48.7	78.2	70.6	83.9	76.2
VFNet	83.7	70.5	84.8	85.4	83.4	84.2	49.8	79.0	73.1	84.3	76.6
Free Anchor	82.3	72.3	85.1	86.0	84.4	85.0	47.5	**82.3**	74.9	85.1	76.1
MFSMNet	84.5	72.1	84.9	**86.3**	81.4	**85.2**	**53.5**	81.7	76.6	82.1	77.1

4.6 Qualitative Analysis of Comparative Experiments

Figure 5. Comparison of the visual detection results of our proposed MFSMNet with the third method on the PASCAL VOC datasets: (a) Input image; (b) Ours; (c) GFL ; (d) ATSS

To better demonstrate the detection performance of our proposed multi-scale edge feature enhancement network (MEFENet) object detection method, the top three detection methods with quantitative experimental results in the Microsoft COCO 2017 datasets are therefore selected in this subsection: The first place is our proposed MEFENet (the second column of Figure 5); the second place is GFL (the second column of Figure 5); and the third place is ATSS (the second column of Figure 5) for visualizing the results. To ensure that the visualization results can actually reflect the real performance of the model, the images used for the inference of the visualization results in this subsection are derived from the PASCAL VOC 2007+2012 datasets, while the training set of the model uses the Microsoft COCO 2017 datasets. Meanwhile, to ensure the fairness of the visualization results, the parameters of the model inference process of the three methods are set as above.

In Figure 5, to demonstrate the advantages of our proposed MEFENet on occlusion object detection therefore, the object image of the occlusion object with multiple scales is selected. The first column shows the input image of the detection network; the second column shows the detection results of our proposed MEFENet; the third column shows (the second-best detection accuracy in Table 2 of the Microsoft COCO 2017 datasets); and the fourth column shows ATSS (the third best detection accuracy in Table 2 of the Microsoft COCO 2017 datasets). In the first row of images with only some slight occlusions, the object frame of our proposed MEFENet detection results have better localization with confidence. In the image of the motorcycle rider in the second row, the samples of

each category obscure each other (backpack, person and motorcycle) causing more difficulty in the detection task, and there is a more obvious gap in the localization and confidence of the object box in the detection images of GFL and ATSS with our proposed MEFENet object detection method. In the third row of the person and bicycle images, the localization of the object frame in the detected images of ATSS is not satisfactory. In the last row of images, the person, fence and horse are occluded from each other, where the person is also holding a bottle with a small size object in his hand, which is a greater challenge for all detectors. ATSS in the fourth column incorrectly detects the horse as a cow, while GFL in the third column detects the horse and the dog in the lower part of the image, and both methods perform unsatisfactorily. Of course, the performance of our proposed MEFENet also has shortcomings, and dogs are also incorrectly detected in the middle of the image. In summary, our proposed MEFENet object detection method demonstrates its actual detection effect through the visualization results in this subsection, and MEFENet has a more obvious advantage in the detection performance of occluded objects.

5. Conclusion

Our proposed multi-scale edge feature enhancement network (MEFENet) object detection method uses a residual network (ResNet) as the backbone, and the features extracted from the backbone are first divided into two branches for processing. The first branch is connected to the feature pyramid networks (FPN) to further extract multi-scale features. The other branch feeds the backbone features into the multi-scale edge feature extraction (MEFE) structure. Then, the edge features are fused with the multi-scale features output from the feature pyramid and sent to the receptor field enhancement (RFE) module. After that, the semantic information is further enhanced using the RFE module and sent to the detection head. Finally, the detection head performs classification and regression tasks to achieve object detection. In the PASCAL VOC 2007+2012 datasets and Microsoft COCO 2017 datasets, our proposed MEFENet object detection method has advantages in terms of detection accuracy compared with other 9 representative object detection methods, and most of the metrics are the highest detection accuracy among 10 methods.

References

[1] Wieczorek M, J. Siłka, M. Woźniak, et al., "Lightweight convolutional neural network model for human face detection in risk situations," **IEEE Transactions on Industrial Informatics**, vol. 18, no. 7, pp. 4820-4829, 2021.

[2] Yang Y., "Adversarial attack against images classification based on generative adversarial networks," **arXiv preprint**, arXiv:2412.16662, 2024.

[3] Kajo I., N. Kamel, and Y. Ruichek, "Incremental tensor-based completion method for detection of stationary foreground objects," **IEEE Transactions on Circuits and Systems for Video Technology**, vol. 29, no. 5, pp. 1325-1338, 2018.

[4] Dong Y., L. Shen, Y. Pei, et al., "Field-matching attention network for object detection," **Neurocomputing**, vol. 535, pp. 123-133, 2023.

[5] Huo F., X. Zhu, L. Zhang, et al., "Efficient context-guided stacked refinement network for RGB-T salient object detection," **IEEE Transactions on Circuits and Systems for Video Technology**, vol. 32, no. 5, pp. 3111-3124, 2021.

[6] Dong Y., W. Tan, D. Tao, et al., "CartoonLossGAN: Learning surface and coloring of images for cartoonization," **IEEE Transactions on Image Processing**, vol. 31, pp. 485-498, 2021.

[7] Dong Y., K. Zhao, L. Zheng, et al., "Refinement Co-supervision network for real-time semantic segmentation," **IET Computer Vision**, 2023.

[8] W. He, T. Zhou, Y. Xiang, Y. Lin, J. Hu, and R. Bao, "Deep learning in image classification: Evaluating VGG19's performance on complex visual data," **arXiv preprint**, arXiv:2412.20345, 2024.

[9] X. Huang, Z. Zhang, X. Li, and Y. Li, "Reinforcement learning-based Q-learning approach for optimizing data mining in dynamic environments," unpublished.

[10] Z. Zheng, Y. Xiang, Y. Qi, Y. Lin, and H. Zhang, "Fully convolutional neural networks for high-precision medical image analysis," **Transactions on Computational and Scientific Methods**, vol. 4, no. 12, 2024.

[11] X. Yan, J. Du, L. Wang, Y. Liang, J. Hu, and B. Wang, "The synergistic role of deep learning and neural architecture search in advancing artificial intelligence," **Proc. 2024 Int. Conf. Electron. Devices Comput. Sci. (ICEDCS)**, Sept. 2024, pp. 452-456.

[12] J. Hu, Z. Qi, J. Wei, J. Chen, R. Bao, and X. Qiu, "Few-shot learning with adaptive weight masking in conditional GANs," **Proc. 2024 Int. Conf. Electron. Devices Comput. Sci. (ICEDCS)**, Sept. 2024, pp. 435-439.

[13] Y. Liang, E. Gao, Y. Ma, Q. Zhan, D. Sun, and X. Gu, "Contextual analysis using deep learning for sensitive information detection," **Proc. 2024 Int. Conf. Comput. Inf. Process. Adv. Educ. (CIPAE)**, Aug. 2024, pp. 633-637.

[14] T. Mei, Z. Zheng, Z. Gao, Q. Wang, X. Cheng, and W. Yang, "Collaborative hypergraph networks for enhanced disease risk assessment," **Proc. 2024 Int. Conf. Electron. Devices Comput. Sci. (ICEDCS)**, Sept. 2024, pp. 416-420.

[15] Z. Gao, T. Mei, Z. Zheng, X. Cheng, Q. Wang, and W. Yang, "Multi-channel hypergraph-enhanced sequential visit prediction," **Proc. 2024 Int. Conf. Electron. Devices Comput. Sci. (ICEDCS)**, Sept. 2024, pp. 421-425.

[16] Z. Liu, M. Wu, B. Peng, Y. Liu, Q. Peng, and C. Zou, "Calibration learning for few-shot novel product description," **Proc. 46th Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval**, July 2023, pp. 1864-1868.

[17] J. Song and Z. Liu, "Comparison of norm-based feature selection methods on biological omics data," **Proc. 5th Int. Conf. Adv. Image Process.**, Nov. 2021, pp. 109-112.

[18] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," **Proc. Eur. Conf. Comput. Vis.**, 2018, pp. 734-750.

[19] K. Duan, S. Bai, L. Xie, et al., "CenterNet: Keypoint triplets for object detection," **Proc. IEEE/CVF Int. Conf. Comput. Vis.**, 2019, pp. 6569-6578.

[20] Z. Dong, G. Li, Y. Liao, et al., "CentripetalNet: Pursuing high-quality keypoint pairs for object detection," **Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.**, 2020, pp. 10519-10528.

[21] J. Redmon, S. Divvala, R. Girshick, et al., "You only look once: Unified, real-time object detection," **Proc. IEEE Conf. Comput. Vis. Pattern Recognit.**, 2016, pp. 779-788.

[22] L. Huang, Y. Yang, Y. Deng, et al., "DenseBox: Unifying landmark localization with end-to-end object detection," **arXiv preprint**, arXiv:1509.04874, 2015.

[23] J. Yu, Y. Jiang, Z. Wang, et al., "UnitBox: An advanced object detection network," **Proc. 24th ACM Int. Conf. Multimedia**, 2016, pp. 516-520.

[24] Z. Tian, C. Shen, H. Chen, et al., "FCOS: Fully convolutional one-stage object detection," **Proc. IEEE/CVF Int. Conf. Comput. Vis.**, 2019, pp. 9627-9636.

[25] R. Girshick, J. Donahue, T. Darrell, et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," **Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.**, 2014, pp. 580-587.

[26] S. Ren, K. He, R. Girshick, et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," **Adv. Neural Inf. Process. Syst.**, vol. 28, 2015.

[27] K. He, G. Gkioxari, P. Dollár, et al., "Mask R-CNN," **Proc. IEEE Int. Conf. Comput. Vis.**, 2017, pp. 2961-2969.

[28] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single shot multibox detector," **Proc. IEEE Eur. Conf. Comput. Vis.**, 2016, pp. 21-37.

[29] C. Y. Fu, W. Liu, A. Ranga, et al., "DSSD: Deconvolutional single shot detector," **arXiv preprint**, arXiv:1701.06659, 2017.

[30] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," **Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.**, 2017, pp. 7263-7271.

[31] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," **arXiv preprint**, arXiv:1804.02767, 2018.

[32] T. Y. Lin, P. Goyal, R. Girshick, et al., "Focal loss for dense object detection," **Proc. IEEE Int. Conf. Comput. Vis.**, 2017, pp. 2980-2988.

[33] M. Everingham, L. Gool, C. Williams, et al., "The Pascal visual object classes challenge 2007," **Int. J. Comput. Vis.**, vol. 88, pp. 303-338, 2010.

[34] T. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common objects in context," **Proc. IEEE Eur. Conf. Comput. Vis.**, 2014, pp. 740-755.

[35] K. Chen, J. Wang, J. Pang, et al., "MMDetection: Open mmlab detection toolbox and benchmark," **arXiv preprint**, arXiv:1906.07155, 2019.