Liu Shijun
Аспірант(122 Комп’ютерні науки)
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID ID:`0000-0003-4882-559X
Email: bbxylsj@163.com
Vadym Shkarupylo
Dr.Sc., Professor
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
ORCID ID:0000-0002-0523-8910
Email: shkarupylo.vadym@nubip.edu.ua

SMALL TARGET DETECTION METHOD BASED ON YOLOV8S UAV PERSPECTIVE

Abstract. To address the challenges in detecting objects with varying scales, dense small targets, and complex backgrounds in drone aerial imagery, the YOLOv8s model incorporates a series of targeted improvements[1]. Specifically: 1) The backbone network introduces RFAConv to overcome feature extraction bottlenecks in complex scenes; 2) The neck network is restructured using BiFPN-GLSA to enhance multi-scale information fusion and spatial feature utilization; 3) A dual-layer detection architecture is introduced to specifically enhance feature representation for small objects; 4) The Inner-EIoU loss function is adopted to optimize bounding box regression accuracy. Validation on the VisDrone2019 dataset demonstrates that these enhancements comprehensively improve core metrics including precision, recall, and mAP, while simultaneously reducing model parameters. This achieves an outstanding balance between detection performance and computational efficiency.

Keywords: YOLOv8s; UAVs; small object detection; images; loss function

1.INTRODUCTION

1.1 YOLOv8s Model Algorithm Principle

YOLOv8s is a single-stage object detection algorithm based on deep learning, developed by Ultralytics. It is one of the most advanced models in the YOLO series. Building upon YOLOv5, it makes several improvements to the network architecture, feature extraction, prediction strategy, and training methods. YOLOv8s achieves an optimal balance between speed and accuracy, making it ideal for applications with high real-time requirements, such as intelligent video surveillance and autonomous driving[2]. The YOLOv8s model consists of a backbone network, a neck network, and a prediction head. The convolutional modules in the model consist of Conv2d (two-dimensional convolution), BatchNorm2d (two-dimensional batch normalization), and SILU (activation function)[3].

1.2 YOLOv8s Network Structure

YOLOv8s retains the backbone network architecture of YOLOv5. By parallelizing more gradient flow branches, it can acquire richer gradient information, thereby improving the model's accuracy and performance. YOLOv8s employs a PANet (Path Aggregation Network) feature pyramid structure in its neck module. This architecture fuses high-level semantic features with spatial details, enriching feature representations and enhancing the model's ability to detect objects of different sizes. The Concat operation upsamples small-scale feature maps and connects them to downsampled large-scale feature maps. This allows the model to utilize multi-scale feature information simultaneously, thereby enhancing detection capabilities[4].

2. THE THEORETICAL BACKGROUNDS

2.1 Machine Learning and Object Detection

Machine learning is a core branch of artificial intelligence focused on developing algorithms that can automatically learn patterns from data to make predictions or decisions. Based on learning paradigms, it is primarily categorized into supervised learning, unsupervised learning, and reinforcement learning. During the early development of object detection tasks, traditional machine learning methods (such as classifiers based on handcrafted features) were widely adopted. These approaches typically decomposed detection into two steps: region proposal and feature classification. Their performance heavily relied on the effectiveness of manually designed features (e.g., SIFT, HOG). However, due to limitations in feature representation capabilities under complex scenes, the accuracy and generalization ability of such methods quickly reached a bottleneck[5].

2.2 Deep Learning and Object Detection

As a pivotal branch of machine learning, deep learning constructs deep neural networks to mimic the brain's hierarchical information processing mechanisms. This enables end-to-end learning of powerful high-level feature representations directly from raw data, fundamentally transforming the technical paradigm in object detection. Deep learning models, exemplified by convolutional neural networks (CNNs), have replaced handcrafted features by achieving direct, precise mappings from image pixels to object locations and categories. Among these, the YOLO series models stand as representative algorithms. They innovatively reframed object detection as a single regression problem, directly regressing bounding boxes and class probabilities at the output layer. This approach significantly enhanced both training and inference efficiency. The success of deep learning laid a solid foundation for achieving high-precision, high-efficiency object detection in complex scenarios, such as those viewed from a drone's perspective[6].

3.RESEARCH METHODS

3.1 Introduction of Receptive Field Attention Convolution (RFAConv)

Standard convolution operations form the core of convolutional neural networks, extracting features through sliding windows with shared parameters to reduce excessive parameters and computational costs associated with fully connected layers. However, standard convolutions fail to adequately account for spatial information differences, thereby limiting network performance. Attention mechanisms enhance network performance by focusing models on key features, yet existing theoretical frameworks do not specifically address spatial feature issues within receptive fields. RFA (Receptive Field Attention Convolution), as a spatial attention mechanism, not only highlights the importance of different features within receptive fields but also enhances attention to spatial characteristics. RFA effectively resolves parameter sharing challenges within convolutional kernels. Spatially receptive field features dynamically generated based on kernel size achieve performance gains through RFA, giving rise to the RFAConv model. This transform model processes features through channel-wise layered processing and dynamically adjusts structures to generate attention weights that avoid feature redundancy, thereby preserving global information integrity. RFA convolution integrates spatial receptive field features with attention mechanisms to optimize feature weight allocation. Specifically, RFA processes input feature X through dual pathways: Path 1 employs average pooling to extract global features, then uses convolution and activation functions to enhance spatial feature importance, generating weight and attention maps; Path 2 directly processes input features via convolution and activation functions to generate weight and attention maps[7].

3.2 Designing a Dual-Layer Small Object Detection Architecture

The YOLOv8s detector head processes large-scale features from high-level feature maps. To address the challenge of capturing small object features, a dual-layer small object detection architecture is designed to better detect small targets. The specific architecture of the dual-layer small object detection structure is as follows: A C2f module is introduced after the initial convolutional layer of the backbone network to capture shallow feature information P1. Subsequently, a convolutional layer is applied to the feature map at the P2 layer, downsampling it to the P3 size. Finally, the downsampled P2 feature map is fused with the P3 feature map to form an updated P3 feature map[8]. This fused feature map undergoes a Conv layer to reduce its size, matching the feature map size of other detection layers. Subsequently, the fused feature map is output through the P1+P2 small object detection layer, enhancing robustness and accuracy when detecting objects of varying scales.

4. THE RESULTS AND DISCUSSION

To enhance the detection robustness of drone aerial images in complex scenes and enable their application on edge computing devices, this paper proposes an improved model based on YOLOv8s. The model incorporates four core enhancements: First, it introduces receptive field attention convolutions (RFAConv) into the backbone network to strengthen feature extraction capabilities; Second, the neck network adopts BiFPN and incorporates a Global-Local Spatial Aggregation (GLSA) module to optimize multi-scale feature fusion and specifically mitigate small object detection failures[9]. Third, a dedicated two-layer small object detection architecture is designed to strengthen feature capture for small targets. Finally, the Inner-EIoU loss function is employed to enhance localization accuracy by optimizing the bounding box regression process. Experiments on the VisDrone2019 dataset demonstrate that our model exhibits significant advantages over baselines and other YOLO variants. Future work will focus on further optimizing performance without increasing model complexity to better serve resource-constrained real-time drone detection tasks.

5. CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

Based on the theoretical analysis, architectural improvements, and experimental validation presented, this study concludes that the proposed enhancements effectively address key challenges in UAV-based small target detection[10]. The integration of RFAConv, BiFPN-GLSA, a dual-layer detection head, and Inner-EIoU loss collectively contributes to a model that is both more accurate and computationally efficient. Future work will explore avenues for further lightweight model design and investigate the integration of temporal information from video sequences to improve detection stability and accuracy in dynamic scenarios.

REFERENCES

[1]Wu T, Dong Y. 2023. Yolo-se: improved yolov8 for remote sensing object detection and recognition. Appl Sci. 13(24):12977.

[2]Li Z, Wang Y, Chen K, Yu Z. 2022. Channel Pruned YOLOv5-based Deep Learning Approach for Rapid and Accurate Outdoor Obstacles Detection. arXiv Preprint. arXiv:2204.13699.

[3]Yu Z, Liu Y, Yu S, Wang R, Song Z, Yan Y, Li F, Wang Z, Tian F. 2022. Automatic detection method of dairy cow feeding behaviour based on YOLO improved model and edge computing. Sensors (Basel). 22(9):3271.

[4]Alqarqaz, M., Bani Younes, M., & Qaddoura, R. (2023). An Object Classification Approach for Autonomous Vehicles Using Machine Learning Techniques. World Electric Vehicle Journal, 14(2), 41.

[5]Feng, J., & Yi, C. (2022). Lightweight Detection Network for Arbitrary-Oriented Vehicles in UAV Imagery via Global Attentive Relation and Multi-Path Fusion. Drones, 6(5), 108.

[6]Wang X, Liu J. Vegetable disease detection using an improved YOLOv8 algorithm in the greenhouse plant environment. Sci Rep. 2024 Feb 21;14(1):4261.

[7]Li, L., Liu, X., Chen, X., Yin, F., Chen, B., Wang, Y., & Meng, F. (2024). SDMSEAF-YOLOv8: a framework to significantly improve the detection performance of unmanned aerial vehicle images. Geocarto International, 39(1).

[8]Zhang H, Hao C, Song W, Jiang B, Li B. 2023. Adaptive Slicing-Aided Hyper Inference for Small Object Detection in High-Resolution Remote Sensing Images. Remote Sensing. 15(5):1249. doi: 10.3390/ rs15051249.

[9]Yang C, Huang Z, Wang N. 2022. QueryDet: cascaded sparse query for accelerating high-resolution small object detection. Paper Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[10]Zhanchao Huang, Jianlin Wang, Xuesong Fu, Tao Yu, Yongqi Guo, Rutong Wang,DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detectionInformation Sciences,Volume 522,2020,Pages 241-258.

MINISTRY OF EDUCATION
AND SCIENCE OF UKRAINE

NATIONAL UNIVERSITY
OF LIFE AND ENVIRONMENTAL
SCIENCES OF UKRAINE

FACULTY OF INFORMATION
TECHNOLOGY

PROCEEDINGS

XIII International scientific
and practical conference

GLOBAL AND REGIONAL PROBLEMS OF INFORMATIZATION IN SOCIETY AND NATURE USING '2025

13-14 November 2025

Kyiv, NULES of Ukraine

Kyiv 2025

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ

МАТЕРІАЛИ

XIII Міжнародної науково-
практичної конференції

ГЛОБАЛЬНІ ТА РЕГІОНАЛЬНІ ПРОБЛЕМИ ІНФОРМАТИЗАЦІЇ В СУСПІЛЬСТВІ І ПРИРОДОКОРИСТУВАННІ '2025

13-14 листопада 2025 року

Київ, НУБіП України

Київ 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

МАТЕРІАЛИ

XIII Міжнародної науково-практичної конференції

ГЛОБАЛЬНІ ТА РЕГІОНАЛЬНІ ПРОБЛЕМИ ІНФОРМАТИЗАЦІЇ В СУСПІЛЬСТВІ І ПРИРОДОКОРИСТУВАННІ '2025

13-14 листопада 2025 року

Київ, НУБіП України

Київ 2025

УДК 004

Рекомендовано до друку вченою радою факультету інформаційних технологій Національного університету біоресурсів і природокористування України (протокол № 4 від 18.12.2025).

Укладач: д.т.н., доцент Шкарупило В.В.

Збірник матеріалів XIII Міжнародної науково-практичної конференції "Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні '2025", 13–14 листопада 2025 року, НУБіП України, Київ. – К.: НУБіП України, 2025. – 206 с.