УДК 004.8

ENHANCED SMALL OBJECT DETECTION USING MODIFIED YOLOV11N MODEL

S. Zhu[1)], S. Ablameyko[1), 2)]

[1)] *Belarusian State University,*
Minsk, Republic of Belarus, zhushuaiyu1001@gmail.com
[2)] *United Institute of Informatics Problems, National Academy of Sciences of Belarus,*
Minsk, Republic of Belarus, ablameyko@yandex.by

The paper proposes an improved YOLOv11 model by introducing a Large Kernel Spatial Attention (LSKA) mechanism and a Gold-YOLO Neck structure, which significantly enhance small object detection performance. The LSKA module combines large-kernel convolution with channel attention to strengthen the model's ability to extract fine-grained features of tiny objects, while the Gold-YOLO Neck optimizes detection performance across different object scales through multi-scale feature fusion. Experiments on the DOTAv1 dataset demonstrate that the proposed method achieves a 1.3% improvement in small object detection accuracy (mAP@0.5) over the baseline YOLOv11, while maintaining real-time performance.

Keywords: Small Object Detection; YOLOv11; Attention Mechanism; LSKA; Gold-YOLO.

УЛУЧШЕННОЕ ОБНАРУЖЕНИЕ МЕЛКИХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ МОДИФИЦИРОВАННОЙ МОДЕЛИ YOLOv11n

С. Чжу[1)], С. Абламейко[1), 2)]

[1)] *Белорусский государственный университет,*
Минск, Беларусь, zhushuaiyu1001@gmail.com
[2)] *Объединённый институт проблем информатики НАН Беларуси,*
Минск, Беларусь, ablameyko@yandex.by

В статье предложена улучшенная модель YOLOv11 с внедрением механизма пространственного внимания на основе больших ядер (LSKA) и структуры шеи Gold-YOLO, что значительно повышает производительность обнаружения малых объектов. Модуль LSKA сочетает свертку с большими ядрами и канальное внимание для усиления способности модели извлекать мелкозернистые признаки tiny-объектов, в то время как шея Gold-YOLO оптимизирует производительность обнаружения на разных масштабах объектов посредством многомасштабного слияния признаков. Эксперименты на наборе данных DOTAv1 показали, что предложенный метод обеспечивает улучшение точности обнаружения малых объектов (mAP@0.5) на 1.3% по сравнению с базовой версией YOLOv11 при сохранении работы в реальном времени.

Ключевые слова: обнаружение малых объектов; YOLOv11; механизм внимания; LSKA; Gold-YOLO.

1. Introduction

With the synergistic advancement of remote sensing technology and deep learning, deep learning‑based object detection in remote sensing imagery [1] has demonstrated significant application value in fields such as urban management, precision agriculture, and national security. However, this technology still faces numerous challenges in real-world applications. First, due to the long-range imaging characteristics of spacecraft, objects usually occupy only a very small pixel area (often less than 32×32 pixels), which substantially increases the difficulty of feature extraction. Second, the complex and dynamic background environments (e.g., urban building clusters, natural landscapes) combined with the low contrast between objects and backgrounds further hinder detection performance. Moreover, the extreme diversity of object categories in remote sensing scenes (ranging from vehicles and ships to agricultural crops) and the intra-class variations in object morphology pose serious challenges to the generalization capability of detection models [2]. Although recent advances in network architecture design and training strategy optimization have continuously improved detection accuracy, problems such as missed detections and false alarms caused by insufficient feature representation remain a critical bottleneck for practical deployment. Therefore, breakthroughs in feature enhancement and multi-scale fusion are urgently required [3].

2. YOLOv11n-LSKA-GoldYOLO: Improved YOLOv11 model

2.1. YOLOv11n model

YOLOv11n is a next-generation lightweight object detection model released by Ultralytics in 2024 [12]. Building on the real-time advantages of the YOLO series, it significantly enhances feature extraction through multiple innovations. The model follows a backbone-neck-head architecture: the backbone integrates the C2PSA module, combining traditional Conv and SPPF with multi-head self-attention to improve global feature modeling; the neck uses an improved C3K2 module for multi-scale feature fusion, splitting features into two branches—one with standard convolutions, the other with configurable C3k or Bottleneck structures for multi-scale extraction. The head employs depthwise separable convolutions to reduce computational cost while maintaining accuracy.

2.2. GOLD-YOLO

a) GOLD-YOLO is an advanced object detection model [6], whose core innovation lies in the aggregate-distribute (GD) mechanism. By employing convolution and self-attention operations, it efficiently fuses multi-level

features, achieving a balance between low latency and high accuracy (as shown in fig. 1). The backbone is responsible for feature extraction; the LOW-GD branch aligns and fuses large-scale features, while the HIGH-GD branch processes small-scale features. The Inject module integrates multi-scale information and passes it to the detection head, thereby enhancing detection performance.

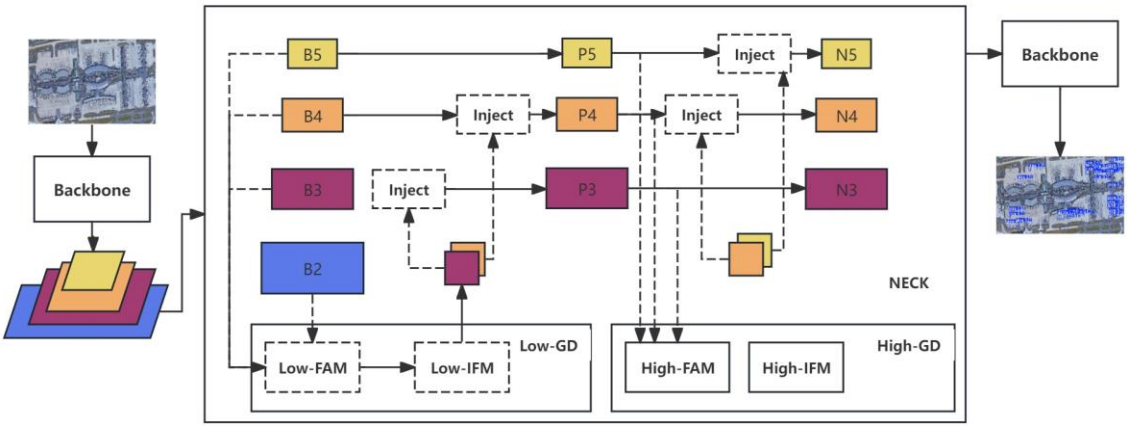

Fig. 1. Gold-YOLO architecture

b) The aggregate-distribute (GD) mechanism is the core of Gold-YOLO, designed for efficient integration of multi-level features. This mechanism aggregates features through the Feature Alignment Module (FAM) and Information Fusion Module (IFM), and then redistributes the fused results back to each layer via the Inject module [8]. In this way, the model fully exploits multi-scale information, achieving a balance between low latency and high accuracy.

In the Gold-YOLO architecture, two key modules are:

- Inject Module: Enhances feature representation by combining global and local features through convolution and a Sigmoid activation function;
- Lightweight Adjacent Feature Fusion (LAF) Module: Aligns and fuses adjacent layer features using average pooling and bilinear up/down-sampling, thereby improving the effectiveness of multi-layer feature interaction.

c) Multi-scale feature fusion [7] is a commonly used technique in object detection, aimed at improving the detection of objects at different scales. Low-level features provide detailed information for small objects, while high-level features capture semantic information for large objects. Fusing these features enhances the model's representational capacity, enabling more accurate detection and localization. In Gold-YOLO, this is realized through the

aggregate-distribute (GD) structure: the Low-GD branch, composed of Low-FAM and Low-IFM, handles large-scale features, while the High-GD branch, consisting of High-FAM and High-IFM, processes small-scale features.

2.3. Large Separable Kernel Attention (LSKA) Attention Mechanism

The mechanism principle of LSKA is an improvement on the application of traditional large kernel attention (LKA) modules in visual attention networks (VAN) [10]. By leveraging large, separable convolutional kernels and spatial dilation convolutions to capture extensive contextual information from images, the mechanism generates attention maps and weights the original features using these maps, thereby enhancing the network's focus on important features and improving model performance. A K×K convolutional operation is decomposed into a (2d-1)×(2d-1) depth convolutional (DW-Conv) operation, K/d×K/d deep dilated convolution (DW-D-Conv), and 1×1 convolution (Conv). The deep convolution and deep dilated convolution are further decomposed into horizontal and vertical layers, and the convolution kernels are connected. The LSKA structure diagram is shown in fig. 2 [9].

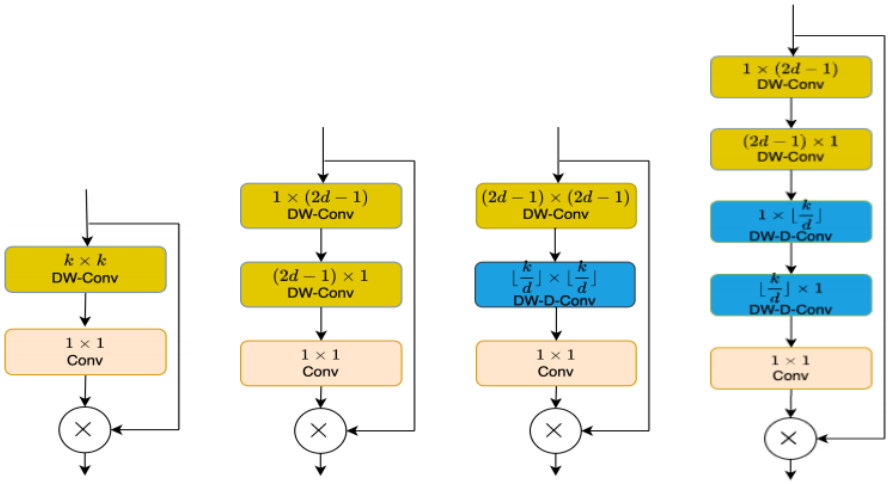

Fig. 2. LSKA Structure Diagram [10]

2.4. The proposed YOLOv11n-LSKA-GoldYOLO model

In this study, we propose an improved lightweight object detection model, termed YOLOv11n-LSKA-GoldYOLO, which is developed on the basis of the YOLOv11n framework (fig. 3). To enhance the representation ability of the network under complex scenarios, we integrate the Large Kernel Separation Attention (LSKA) mechanism into the backbone. By employing a decomposed large-kernel design, LSKA effectively enlarges the receptive field while

maintaining computational efficiency, enabling the model to better capture global contextual information and suppress redundant features. This design is particularly beneficial for detecting small objects and densely distributed targets. Furthermore, in the feature fusion stage, we adopt the Gold-YOLO Neck structure, which facilitates more efficient multi-scale information interaction and adaptive feature integration. This improves the effectiveness of feature propagation and strengthens detection accuracy across objects of different sizes. Through the combination of these two strategies, YOLOv11n-LSKA-GoldYOLO achieves superior detection performance while preserving low computational cost and real-time inference capability, thereby providing a practical and effective solution for lightweight object detection tasks.

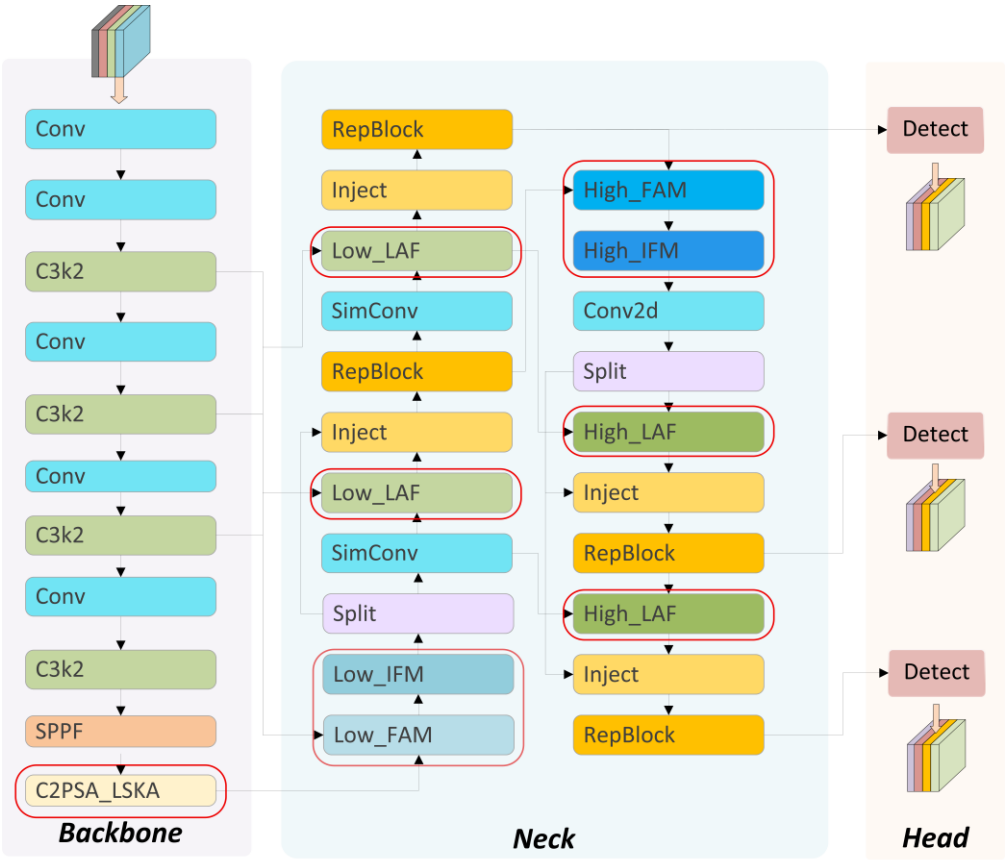

Fig. 3. Improved model structure diagram

3. Experimental Procedure and Analysis

The experiments in this study were conducted on the publicly available DOTAV1 dataset (Dataset for Object deTection in Aerial Images, version 1) [11], which is a large-scale dataset specifically designed for aerial image object detection. The dataset contains aerial and satellite images collected from various regions worldwide, covering diverse scenes such as urban areas, farmland,

ports, and forests. It consists of 2,806 high-resolution images annotated with 15 common object categories, totaling 188,282 instances. The dataset was divided into training (train), testing (test), and validation (val) subsets in an 8:1:1 ratio.

The experiments were carried out in a high-performance hardware and stable software environment. The hardware configuration included: GPU: NVIDIA RTX 4090 (24GB) for accelerated model training; CPU: Intel Core i9-14900KF; RAM: 60GB. The main framework was PyTorch 2.1.0 with CUDA 12.1 support. The programming language was Python 3.10. Training was conducted for 300 epochs with batch size automatically adjusted according to GPU memory (batch=-1). Stochastic Gradient Descent (SGD) was used as the optimizer, while all other parameters were kept at default settings.

Comparative experiments were conducted with the original YOLOv11n model, and the performance of the new model on the DOTAv1 dataset was evaluated in detail. Key performance metrics, including mAP50% and mAP50-95%, are summarized in table 1.

Table 1

Experimental results

Модель	mAP50%	mAP50-95%
YOLOv11n	42	25.7
YOLOv11n-LSKA	42.9	26
YOLOv11n-goldyolo	42.2	25.8
YOLOv11n-LSKA-goldyolo	43.3	26

Detection examples are provided in fig. 8. The experimental results demonstrate that the YOLOv11n-LSKA-GoldYOLO model achieves a 1.3% improvement in mAP50% and a 0.3% increase in mAP50-95% compared to the original YOLOv11n. This significant enhancement stems from the integration of the Large Kernel Separation Attention (LSKA) mechanism and the Gold-YOLO Neck structure, enabling the model to fully leverage multi-scale feature information in aerial images. Results of object detection with our model is shown in fig. 4.

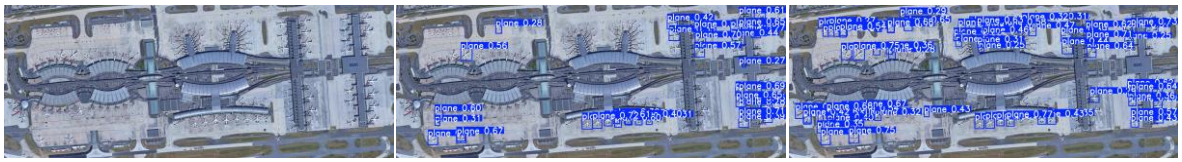

Fig. 4. Original image, result of YOLOv11n and our method

4. Conclusion

In this study, the proposed YOLOv11n-LSKA-GoldYOLO model demonstrated significant improvements in detection performance on the DOTAV1 dataset. The LSKA mechanism effectively captures long-range

dependencies among widely distributed objects in remote sensing images through large receptive field separable convolution kernels, making it particularly suitable for handling the expansive scenes and small-object clusters commonly found in the DOTAV1 dataset. Meanwhile, the Gold-YOLO Neck's gather-and-distribute architecture enhances the model's representational capability across objects of different sizes through multi-level feature fusion. Specifically, the low-order GD branch improves detection accuracy for small objects such as vehicles and ships, while the high-order GD branch optimizes the localization of large-scale structures such as airports and ports. Experimental results demonstrate that this improved approach significantly enhances the robustness of aerial image object detection while maintaining real-time performance, providing a more effective solution for remote sensing vision tasks.

References

1. A review of deep learning target detection methods / Y. Zhao [et al.] // Chinese Journal of Graphics. 2020. Vol. 25, № 4. P. 629–654.

2. Effective small object detection in remote sensing images based on improved YOLOv8 network / Z. Li [et al.] // Nonlinear Phenomena in Complex Systems. 2024. Vol. 27, № 3. P. 278–291.

3. Remote sensing small object detection algorithm based on dual-layer attention mechanism / X. Chen [et al.] // Journal of Rocket Force University of Engineering. 2025. Vol. 39, № 1. P. 60–66.

4. *Li Y., Zhang X.* Object detection for UAV images based on improved YOLOv6 // IAENG International Journal of Computer Science. 2023. Vol. 50, № 2. P. 62–66.

5. *Varghese R., Sambath M.* YOLOv8: A novel object detection algorithm with enhanced performance and robustness // Int. Conf. on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India, 2024. P. 1–6. DOI: 10.1109/ADICS58448.2024.10533619.

6. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism / C. Wang [et al.] // Advances in Neural Information Processing Systems. 2023. Vol. 36. P. 51094–51112.

7. SeaFormer++: Squeeze-enhanced axial transformer for mobile visual recognition / Q. Wan [et al.] // International Journal of Computer Vision. 2025. Vol. 133, № 6, P. 3645–3666.

8. Shuffle transformer: Rethinking spatial shuffle for vision transformer / Z. Huang [et al.]. arXiv preprint arXiv:2106.03650, 2021.

9. Research on greenhouse tomato maturity detection algorithm based on YOLOv10n / J. Li [et al.] // Journal of Inner Mongolia Agricultural University (Natural Science Edition). 2024. URL: https://link.cnki.net/urlid/15.1209.S.20250409.1333.008 (date of access: 22.07.2025).

10. *Lau K. W., Po L. M., Rehman Y. A. U.* Large separable kernel attention: Rethinking the large kernel attention design in CNN // Expert Systems with Applications. 2024. Vol. 236. Article no. 121352.

11. DOTA: A large-scale dataset for object detection in aerial images / G. S. Xia [et al.] // IEEE Conf. on Computer Vision and Pattern Recognition. 2018. P. 3974–3983.

12. *Zhou K., Jiang S.* Forest fire detection algorithm based on improved YOLOv11n // Sensors. 2025. Vol. 25, iss. 10. Article no. 2989.