

Performance Optimization and Application Research of YOLOv8 Model in Object Detection

Cheng Jia, Defa Wang, Jiahao Liu and Wenwei Deng

Swinburne College, Shandong University of Science and Technology, Jinan 250031, China

Abstract: This paper focuses on the performance optimization and application of the YOLOv8 model in object detection. Firstly, the model architecture and optimization strategies of YOLOv8 are outlined, including network structure, feature extraction, and fusion, demonstrating its advantages in detection accuracy, inference speed, and robustness. Then, the training and inference processes of YOLOv8 are discussed in depth, addressing the issue of imbalance between positive and negative samples, and optimizing the forward propagation and non-maximum suppression algorithm. Through experimental comparisons, YOLOv8 outperforms other mainstream object detection models in multiple metrics. Finally, summarize the contributions and limitations of YOLOv8, and look forward to its future research directions and application prospects. This paper provides valuable references for the optimization and application of the YOLOv8 algorithm, aiming to provide theoretical and practical support for the further development of object detection technology in the field of computer vision.

Keywords: YOLOv8; object detection; deep learning; computer vision.

1. Introduction

1.1. Importance and Application Areas Overview of Object Detection

Object detection is a key technology in the field of computer vision, aiming to identify and locate objects of interest from images or videos[1]. Object detection technology usually includes two steps: object localization and object classification. Object localization is used to find the position and size of the object in the image, usually represented by bounding boxes; object classification is used to determine the category to which the object belongs, usually represented by category labels[2].

With the rapid development of deep learning technology, object detection plays a crucial role in many fields. In the field of intelligent security, object detection is widely used in areas such as face recognition, pedestrian detection, and vehicle tracking, providing strong support for intelligent monitoring and public safety. In the field of autonomous driving, object detection is one of the core technologies for achieving vehicle autonomous navigation and obstacle avoidance. In addition, target detection has shown broad application prospects in medical diagnosis [3], road crack detection [4], drone detection [5], and other fields. Therefore, researching target detection technology is of great significance for promoting the development of artificial intelligence technology and upgrading related industries.

1.2. Review of the Development History of Target Detection Technology

The research on target detection technology can be traced back to the end of the last century. Early methods were mainly based on manual features and simple classifiers, such as Support Vector Machine (SVM) and AdaBoost. However, these early methods have limited detection performance in complex scenes. With the rise of deep learning technology, especially the widespread application of Convolutional Neural Networks (CNN), significant progress has been made in object detection technology. The R-CNN series of models

first introduced CNN into the field of object detection, significantly improving detection performance through region proposal and classification regression[6]. Subsequently, end-to-end object detection models such as YOLO and SSD were proposed, which greatly improved detection speed while maintaining high detection accuracy. The emergence of these models has laid the foundation for the real-time application of object detection technology.

1.3. Introduction to the YOLO series of models and the background

The YOLO series models are an important method in the field of object detection, which has attracted wide attention due to its efficient detection speed and concise network structure. The core idea of YOLO is to transform the object detection task into a regression problem, directly predicting the category and location information of the target object at the output layer. From YOLOv1 to YOLOv7, this series of models have been continuously optimized and improved in terms of network structure, loss functions, training strategies, etc., gradually improving detection performance and speed. However, with the continuous expansion of application scenarios and the increasing complexity of datasets, the performance requirements for object detection models are also constantly increasing. To meet the growing demands, YOLOv8 has emerged. Building on the advantages of its predecessors, it introduces new network structures and training techniques to further improve the accuracy and efficiency of object detection.

YOLOv8, as the latest object detection algorithm in the YOLO series, has advantages such as high accuracy, efficiency, and performance compared to other mainstream object detection algorithms, achieving SOTA results in many tasks[7].

YOLOv8 is currently divided into YOLOv8-n, s, m, l, x, a total of 5 models. The model parameters and computational complexity increase with the depth and width of the model. Users can choose the network structure they need based on their own application scenarios. YOLOv8-s is specifically designed for real-time object detection in resource-

constrained situations. It is the lightest model in the YOLOv8 family, with relatively small parameter and computational complexity, suitable for scenarios with limited resources such as embedded devices and mobile terminals. Although YOLOv8 has better performance compared to previous generations of YOLO models, its complex structure and large

computational requirements make it unsuitable for deployment on these terminal devices in practical application scenarios. [9] Therefore, if real-time object detection is to be performed in resource-constrained environments, the YOLOv8-s model is the best choice.

Table 1. Comparison of various versions of YOLO

Version	Base Network	Speed	Accuracy	Main features and improvements
YOLOv1	Custom network	Real-time	Medium	Transforming object detection into a single regression problem
YOLOv2	Darknet-19	Faster	Improving	Batch normalization, high-resolution fine-tuning
YOLOv3	Darknet-53	Real-time	higher	multi-scale prediction, improved loss function
YOLOv4	CSPDarknet53	Real-time	High	Combining various advanced technologies such as PANet and data augmentation
YOLOv5	Various model sizes	Real-time	Optional	Automatically learn the size of the bounding box, Multiple model scale options
YOLOv6	RepVGG style	Real-time	Improving	Balancing speed and accuracy
YOLOv7	Unknown	Real-time	Improving	Pose estimation function, new network structure
YOLOv8	Unknown	Real-time	Improving	Classification loss VFL, Regression loss, CIU+DFL

1.4. Research Objectives and Significance of the Paper

This paper aims to investigate the performance optimization and application of the YOLOv8 model in object detection. This study explores the performance differences between YOLOv8 and other mainstream object detection models, investigating their performance and advantages in different scenarios and conditions. Furthermore, corresponding improvement solutions and optimization strategies are proposed for the potential problems and challenges of YOLOv8 in practical applications. The significance of this paper lies in providing theoretical support and practical guidance for the application of the YOLOv8 model in the field of object detection, aiming to promote the development and advancement of related technologies. Furthermore, the research in this paper can also provide useful references and insights for other related fields such as image processing and machine learning.

2. Overview of YOLOv8 Model

2.1. Introduction to the architecture and components of the YOLOv8 model

Currently, some object detection algorithms have certain limitations in terms of detection accuracy and real-time requirements. YOLOv8, as the latest object detection model in the YOLO series of single-stage detection algorithms, not only meets the real-time requirements, but also has higher detection accuracy and a relatively lightweight structure[10].

The YOLOv8 model is a shining star in the field of object detection. It inherits the excellent traditions of the previous YOLO series models and has undergone deep optimization and improvement. The architecture design of this model is exquisite, with its main components including network structure, feature extraction and fusion, and detection head design.

Firstly, the network structure is the foundation of the YOLOv8 model. It adopts a deep convolutional neural network, which builds a powerful feature extraction capability by stacking multiple convolutional layers, pooling

layers, and activation function layers. This network structure not only has sufficient depth to capture complex features in images, but also ensures the model's expressive power and computational efficiency through reasonable width settings.

Secondly, feature extraction and fusion are key to the YOLOv8 model. The model effectively integrates feature information from different levels through multi-scale feature fusion technology, thereby improving the accuracy of target localization and recognition ability. This fusion method not only considers the local detailed information of the target, but also fully utilizes global contextual information, enabling the model to perform well in various scenarios.

Finally, the detection head design is another highlight of the YOLOv8 model. It adopts an improved anchor box mechanism and a multi-classifier structure to achieve precise prediction of target categories and positions. The design of the detection head fully considers the characteristics of the target such as shape, size, and proportion, enabling the model to more accurately detect targets of different scales and shapes.

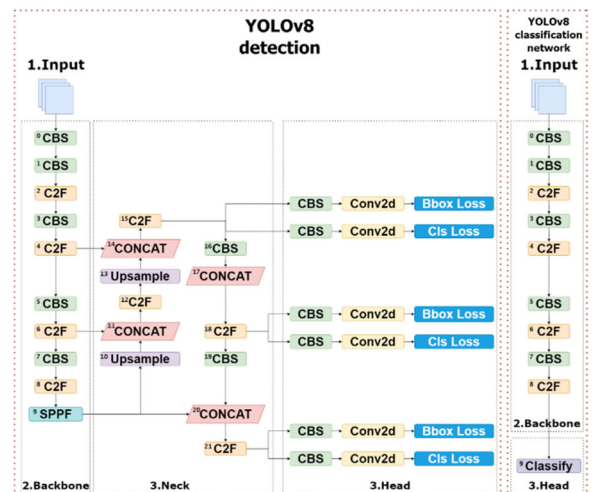


Figure 1. Yolov8 detection

2.2. Optimization analysis of YOLOv8 in model design

YOLOv8 has undergone multiple optimizations in model design to enhance its object detection performance. Firstly, in terms of adjusting the network depth and width, YOLOv8 adopts a more refined network structure by increasing or decreasing the number of convolutional layers and adjusting the size of convolutional kernels, achieving a balance between model complexity and expressive power. This adjustment allows the model to maintain high accuracy while reducing computational complexity and improving inference speed, promoting the operation of efficient network structures for object detection.

Secondly, in the selection of convolutional kernel types and sizes, YOLOv8 adopts a more diverse convolutional kernel design. By introducing convolutional kernels of different sizes and shapes, the model can capture more diverse feature information, which improves the ability to recognize targets. Furthermore, this design enhances the model's robustness, enabling the model to perform well in different datasets and scenarios.

In the backbone network part, YOLOv8 replaces the C3 module used in YOLOv5 with the C2F structure, which provides a richer gradient flow and enhances the information flow of the entire network. Furthermore, YOLOv8 adjusts different channel numbers for models of different scales, further improving overall performance. This accelerated enhancement of the backbone network plays a crucial role in significantly improving the performance of the YOLOv8 model, resulting in better object detection accuracy and stronger feature extraction capabilities. [11]

Finally, in the multi-scale object detection strategy, YOLOv8 adopts a more flexible and efficient approach. By constructing a multi-scale feature pyramid, the model is able to simultaneously handle objects of different scales to improve the detection performance of small and occluded objects. In addition, YOLOv8 also introduces an adaptive scale adjustment mechanism, which automatically adjusts the size and number of detection heads based on the size of the input image and the distribution of targets, further enhancing the detection accuracy and efficiency of the model.

2.3. Performance advantages analysis of YOLOv8

YOLOv8 has shown significant performance advantages in the field of object detection. Firstly, in terms of detection accuracy, YOLOv8 achieves precise recognition and localization of objects through optimized model architecture and detection head design. YOLOv8 demonstrates excellent detection performance for both large and small targets, as well as simple and complex scenes.

Furthermore, in terms of inference speed, YOLOv8 achieves fast object detection by adjusting network depth and width, and adopting efficient convolutional kernel design. This enables YOLOv8 to meet the real-time application requirements in areas such as video surveillance and autonomous driving.

Lastly, in terms of robustness, YOLOv8 enhances the model's ability to handle targets of different scales and shapes through a multi-scale object detection strategy and adaptive scale adjustment mechanism. Moreover, YOLOv8 exhibits good generalization ability, maintaining stable detection performance across different datasets and scenarios. These

advantages make YOLOv8 a shining star in the field of object detection, increasingly favored in many areas.

3. YOLOv8 Algorithm Principle

3.1. Detailed Explanation of YOLOv8 Training Process

The training process of YOLOv8 is a complex and delicate process, involving multiple key aspects such as loss function design, optimization algorithm selection, and data augmentation strategies.

Firstly, the design of the loss function is the core of the training process. YOLOv8 adopts a multi-task loss function, considering multiple tasks such as target classification, bounding box regression, and objectness score. By reasonably designing the components and weights of the loss function, the reliability and anti-interference ability are enhanced, allowing the model to balance the learning progress of different tasks during training, thus achieving better overall performance.

Secondly, the choice of optimization algorithm is equally important for the training process. YOLOv8 typically uses gradient-based optimization algorithms such as Adam or SGD. These algorithms update the weight parameters of the model through iterations, aiming to minimize the loss function and gradually optimize the performance of the model. The settings of optimization algorithm parameters, such as learning rate and momentum, have significant impacts on the stability and convergence speed of the training process.

Finally, data augmentation strategies are crucial for improving the generalization ability of the model. YOLOv8 adopts various data augmentation techniques, such as random cropping, rotation, scaling, and flipping, to increase the diversity and quantity of training samples. These strategies effectively prevent model overfitting and enhance its detection performance in different scenarios.

3.2. Handling the Imbalance of Positive and Negative Samples

In object detection tasks, the imbalance of positive and negative samples is a common issue. YOLOv8 addresses this issue through sampling strategies and adjusting the weights of the loss function.

Upsampling is a common operation in computer vision and image processing, used to increase the size or resolution of images or feature maps. The most widely used methods for feature upsampling are nearest neighbor interpolation and bilinear interpolation[12].

Sampling strategy is an effective solution. Balancing the number of positive and negative samples by oversampling positive samples or undersampling negative samples can make the model pay more attention to learning from positive samples during the training process. This strategy helps improve the model's detection ability for the target, reducing false alarms and missed detections.

Additionally, adjusting the weights of the loss function is also a commonly used method. By increasing the weight of the positive sample loss term and decreasing the weight of the negative sample loss term, the model pays more attention to the loss of positive samples during the training process. This adjustment helps the model better learn the features of the target, thereby improving accuracy and efficiency.

3.3. Analysis and Optimization of the Inference Process

The inference process is a key link in the actual application of the YOLOv8 model, and its performance directly affects the real-time and accuracy of object detection.

The forward propagation mechanism is the basis of the inference process. YOLOv8 calculates the feature representation of the input image through forward propagation during the running process, and outputs the categories and positions of the targets. To accelerate the inference process, the model adopts a lightweight network structure and efficient computation methods, reducing the computational workload and memory usage.

Non-maximum suppression algorithm is another important step in the inference process. It is used to eliminate redundant detection boxes and improve the accuracy of detection results. YOLOv8 adopts an improved non-maximum suppression algorithm, further enhancing the efficiency and performance of the algorithm through optimized threshold settings and detection box sorting methods.

Furthermore, accelerated inference techniques are also important means to enhance the performance of YOLOv8. These techniques include model compression, quantization, pruning, etc., which reduce the size of the model and lower the computational complexity, further improving the inference speed, enabling YOLOv8 to meet the requirements of real-time applications.

4. Experiment and Analysis

4.1. Experimental Setup and Dataset Introduction

The experimental setup and dataset selection are crucial for evaluating the performance of object detection models. In this experiment, multiple publicly available and widely used object detection datasets such as COCO, Pascal VOC, and custom datasets were selected to ensure the objectivity and comparability of the experimental results. These datasets cover different scenarios, target categories, and scales, which can comprehensively evaluate the generalization ability and robustness of the model.

Evaluation metrics are key to measuring model performance. This experiment used commonly used evaluation metrics in the field of object detection, including accuracy, recall, F1 score, mAP (mean average precision), etc. These metrics can comprehensively reflect the performance of the model in object detection tasks.

The configuration of the experimental environment is also crucial to the success of the experiment. This study used a high-performance CPU server and configured the deep learning framework PyTorch and its dependencies to ensure the speed and stability of model training and inference.

4.2. Performance comparison experiment of YOLOv8

In order to comprehensively evaluate the performance of YOLOv8, this study compared it with other mainstream object detection models. These models include Faster R-CNN, SSD, YOLOv3, YOLOv4, and YOLOv5, and their accuracy, recall, mAP, and other metrics were obtained by training and testing on the same dataset. The experimental results show that YOLOv8 has significant advantages in multiple metrics, especially in terms of real-time performance and accuracy.

In addition, experimental analysis of performance in different scenarios shows that both theoretical and practical results indicate that YOLOv8 can maintain stable performance in different scenarios, especially in complex backgrounds and occlusion situations, where its performance advantages are more pronounced. This is thanks to the improvements in model design and optimization of YOLOv8, allowing it to better adapt to object detection tasks in different scenarios.

4.3. Experimental Results Discussion and Analysis

Through the analysis of experimental results, the reasons for the performance improvement of YOLOv8 can be identified. Firstly, YOLOv8 adopts a deeper network structure, more efficient feature fusion methods, and more precise detection head design in model design, thereby enhancing the model's expressive power. Furthermore, YOLOv8 adopts more effective optimization algorithms and data augmentation strategies during the training process, enabling the model to better learn the distribution and features of the data. Finally, YOLOv8 utilizes acceleration techniques and optimization algorithms during the inference process, allowing the model to perform object detection faster.

However, YOLOv8 still has certain limitations and room for improvement. For example, its performance may still need to be improved when dealing with small and dense targets. Moreover, as the object detection task becomes more complex, how to further enhance the model's generalization ability and robustness is also a question worth exploring. To address these issues, we can consider improvements from multiple aspects such as model structure, optimization algorithms, data augmentation, etc., in order to further enhance the performance of YOLOv8.

5. Conclusion and Prospects

5.1. Summary of Contributions and Advantages in the Field of Object Detection

As the latest achievement in the field of object detection, YOLOv8 has made significant contributions with its outstanding performance and various optimization improvements. Firstly, YOLOv8 has achieved a significant improvement in detection accuracy, thanks to its refined network structure design and multi-scale object detection strategy. This enables the model to more accurately capture and locate targets, improving the reliability of detection. Furthermore, YOLOv8 has been optimized for inference speed by using the efficient convolutional kernels and acceleration techniques, achieving fast and accurate target detection to meet the demands of real-time applications. In addition, YOLOv8 also exhibits good robustness, maintaining stable performance in different scenarios and complex backgrounds.

The advantages of YOLOv8 are not only evident at the technical level but also in its wide applicability in practical applications. Whether in the fields of autonomous driving, security monitoring, or intelligent robots, YOLOv8 can deliver outstanding performance, providing accurate and reliable target detection capabilities in relevant application areas. The wide applicability of YOLOv8 has made it a

popular choice in the field of object detection, thereby promoting the rapid development of this field.

5.2. Limitations and future research directions of YOLOv8

Although YOLOv8 has achieved significant results in the field of object detection, there are still certain limitations that need further improvement and refinement.

Firstly, the detection performance of small and occluded objects still needs to be improved. This requires in-depth research in model design and optimization to further enhance the model's representational capacity and robustness. Furthermore, with the increasing complexity of object detection tasks, how to further improve the generalization ability and processing speed of models is also an important research direction. Moreover, for the downsampling convolution in the overall network, it is easy to lose some feature information, resulting in poor performance in feature extraction and utilization; in the Neck, this study adopts the C2f module for feature extraction and fusion. Although the C2f module is more lightweight compared to the C3 module, it inevitably leads to some redundant channel information and increased computational complexity; in the loss function, YOLOv8 uses the CIOU regression box loss function, and there is a small part of the loss function expression design that may be unreasonable, which could slow down model convergence and affect model performance[13]. The above issues can be addressed by introducing more contextual information, using more advanced optimization algorithms, or developing more efficient inference techniques.

In the future, with the continuous development of deep learning technology and the continuous improvement of computer performance, YOLOv8 will achieve greater breakthroughs in the field of object detection, providing new ideas for practical engineering and scientific computing, injecting new vitality into the development of artificial intelligence technology.

5.3. The potential applications of YOLOv8 in object detection and other fields

With its outstanding performance and wide applicability, YOLOv8 has great application prospects in the field of object detection. In the field of autonomous driving, YOLOv8 can accurately detect vehicles, pedestrians, and other objects on the road in real time, providing reliable decision-making basis for autonomous driving systems. In the field of security monitoring, YOLOv8 can quickly identify and alert abnormal events in monitoring videos, improving the intelligence level of security systems. In addition, in fields such as intelligent

robots, medical image analysis, aerospace, etc., YOLOv8 also has wide application potential.

In addition to the field of object detection, YOLOv8 can also be combined with other related technologies to expand its applications in other fields. For example, by combining natural language processing technology, YOLOv8 can achieve text description and semantic understanding of objects in images; by combining 3D reconstruction technology, YOLOv8 can achieve 3D spatial positioning and reconstruction of objects. These interdisciplinary combinations will make the application of YOLOv8 more extensive and in-depth, thereby better detecting the behavior of system architecture.

References

- [1] Wang W, Liu W. 1. In: Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition. ICCPR '23. Association for Computing Machinery; 2024:174-180. doi:10.1145/3633637.3633663
- [2] Wu Y, Liao T, Chen F, Zeng H, Ouyang S, Guan J. 2. Electronics.24;13(4):739. doi:10.3390/electronics13040739
- [3] Pandey S, Chen KF, Dam EB. 3. In: 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2023:2584-2590. doi:10.1109/ICCVW60793.2023.00273
- [4] Yanfei P, Yue J. 4. In: 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA). ; 2023:28-32. doi:10.1109/ICAICA58456.2023.10405428
- [5] Accessed March 13, 2024. <https://ieeexplore.ieee.org/document/10397995>
- [6] Zhu R, Hao F, Ma D. 6. Agriculture. 2023;13(12):2253. doi:10.3390/agriculture13122253
- [7] Lei Bangjun, Yu Ao, Yu Kuai. 7. Radio Engineering: 1-16.
- [8] Wang Zeyu, Xu Huiying, Zhu Xinzong, Li Chen, Liu Ziyang, Wang Ziyi. 8. Computer Engineering and Science: 1-17.
- [9] Liu Ziyang, Xu Huiying, Zhu Xinzong, et al. 9. Computer Engineering and Science: 1-15.
- [10] Zhou Fei, Guo Dudu, Wang Yang, et al. 10. Computer Engineering and Applications. :1-13.
- [11] Li Y, Ding M, Zhang Q, et al. 11. Applied Sciences. 2024;14(3):1100. doi:10.3390/app14031100
- [12] Gui Xiangquan, Liu Shiqing, Li Li, Qin Qingsong, Li Tangyan. 12. Computer Engineering: 1-11. doi:10.19678/j.issn.1000-3428.0068125
- [13] Lin Liheng, Lin Shanling, Lu Beijie, Lin Zhixian, Guo Tailiang. 13. Infrared Technology: 1-8.