Journal of Electronic Research and Application, 2026, Volume 10, Issue 4
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Improved Exam Paper Score Detection Algorithm Based on YOLO11n

Ruilin Mu*, Pengyuan Zhu, Peijie Yang

College of Mechanical Engineering, Tianjin University of Science & Technology, Tianjin 300222, China

***Corresponding author:** Ruilin Mu, mrl3667@tust.edu.cn

Abstract: To address the challenges in detecting scores and related information on test papers, such as complex backgrounds and diverse handwriting styles, this paper proposes an improved algorithm based on YOLO11n. A Difference-of-Gaussians downsampling module, DOG-Stem, is designed to enhance edge feature extraction. Moreover, a lightweight grouped detection head, EfficientHead, is constructed, reducing parameters and computational complexity by 10.5% and 15.9%, respectively, while maintaining high performance. Finally, the WIoU loss function is introduced to accelerate model convergence. Experimental results demonstrate that the improved model achieves an mAP50 of 96.3% and an mAP50-95 of 68.6% on the test set, representing increases of 1.3% and 1.6% over the original YOLO11n. The proposed model exhibits superior precision and robustness.

Keywords: YOLO11; Complex backgrounds; Text detection; Lightweight

Online publication: May 21, 2026

1. Introduction

As a vital link between computer vision and natural language processing, text detection is foundational to recognition tasks [1]. In examination grading, traditional manual review is labor-intensive and prone to human error, particularly under time constraints. With the advancement of educational digitalization, automatically extracting structured information, such as grading columns, annotated areas, and itemized scores, has become a core component of automated verification [2]. Deep learning-based object detection algorithms efficiently and accurately locate and extract scoring areas within exam papers, providing robust support for subsequent tasks.

Relevant research has demonstrated the efficacy of such frameworks. Wijaya *et al.* developed a "two-stage YOLOv7" license plate recognition system that first localizes the plate and then segments characters for end-to-end detection [3]. Sun *et al.* proposed RA-YOLOv8 to tackle background interference and curved text in electronic seals; their approach integrates an RFEMA module (combining RFAConv and EMA) into the backbone, employs AKConv in the neck for enhanced feature extraction, and utilizes the MPDIoU loss

function to refine bounding box regression [4].

2. Proposed methodology

2.1. Model improvement

Although YOLO11 performs excellently on general-purpose datasets, it still exhibits limitations in detecting dense small text and separating foreground targets from complex backgrounds. This study proposes an improved YOLO model to enhance text detection performance in challenging scenarios.

As illustrated in **Figure 1**, the proposed architecture consists of a backbone, a neck, and a detection head. Compared to the baseline YOLO11n, we introduce the DOG-Stem module to replace the initial convolutional layers. Furthermore, a lightweight EfficientHead is constructed using shallow grouped convolutions combined with 1×1 mapping, significantly boosting small-target and multi-scale detection capabilities at a low computational cost. Additionally, the original CIoU loss is replaced with WIoU loss to accelerate convergence and refine detection precision.

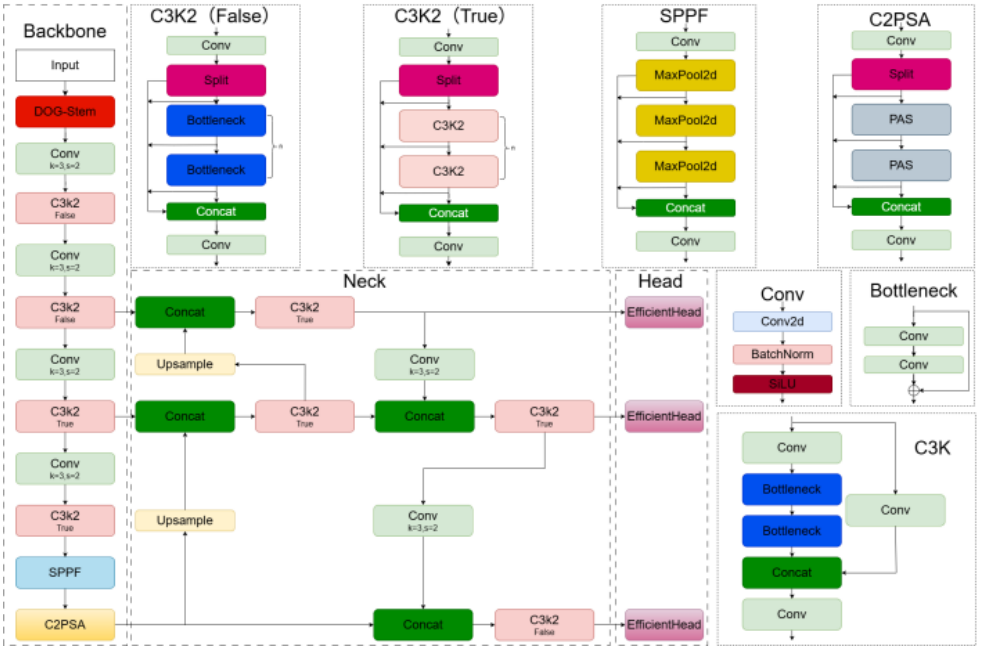

Figure 1. Network architecture of the improved model.

2.2. DOG-Stem

To mitigate the impact of complex backgrounds and noise interference, this study proposes the DOG-Stem module (**Figure 2**), which integrates a DOGFilter, convolutional downsampling, and DSConv components. Adhering to a "sharpening-before-denoising" strategy, the module first employs the DOGFilter to enhance edge details, followed by two 3×3 convolutional layers to downsample the feature map to H/2×W/2. Subsequently, parallel DSConv branches are utilized for feature summation and normalization, effectively suppressing noise and aggregating contextual information within a large receptive field. The DOGFilter provides dual capabilities in noise reduction and edge enhancement, where it leverages DSConv to suppress initial noise and then applies the DOG operator to highlight regions with significant intensity transitions [5].

As a lightweight denoising component, DSConv implements learnable depthwise convolutions for Gaussian-like smoothing, filtering out noise while preserving fine-grained details. Its formal definition is as follows:

$$DOG(x,y,\sigma,k) = G(x,y,k\sigma) - G(x,y,\sigma) \tag{1}$$

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \tag{2}$$

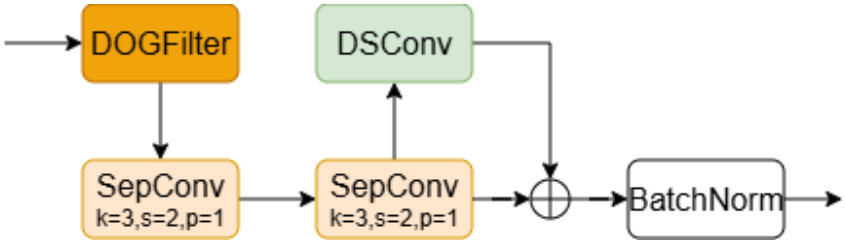

Figure 2. The structure of DOG-Stem.

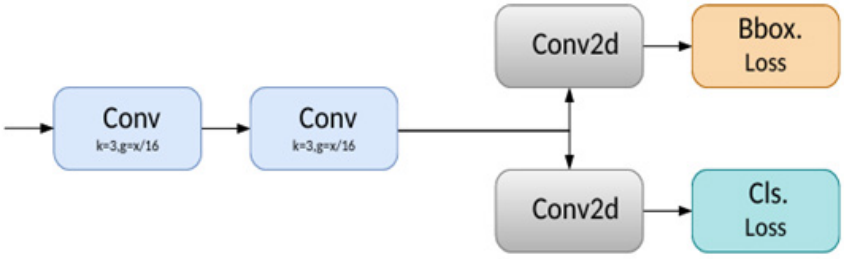

Figure 3. The structure of EfficientHead.

2.3. EfficientHead (EHead)

EfficientHead leverages shared backbone features (**Figure 3**), utilizing two independent 1×1 convolutions for bounding box regression and classification, respectively, to replace the deep dual-branch structure in conventional detection heads [6]. Its primary advantages are two-fold as follows:

(1) Lightweight Architecture: By replacing complex parallel convolutional blocks across scales with a combination of two-layer 3×3 grouped convolutions and a 1×1 projection convolution, it significantly reduces the number of parameters and computational overhead (FLOPs);

(2) Synergistic Feature Learning: The regression and classification branches share features and undergo joint gradient backpropagation, which strengthens feature representation and accelerates convergence.

2.4. WIoU loss function

The original YOLO11n framework employs CIoU as the bounding box regression loss [7]. Although CIoU incorporates aspect ratio constraints based on IoU, it suffers from inconsistent gradients for width and height, leading to convergence difficulties. Furthermore, CIoU focuses primarily on anchor box shape optimization while neglecting the imbalance in anchor box quality, which limits the model's generalization. To address these issues, this study replaces CIoU with WIoUv3 (the third version of Wise-IoU) to accelerate convergence and enhance detection precision [8]. WIoUv3 introduces a dynamic non-monotonic focusing mechanism. Its formal definition is as follows:

$$L_{\text{WIoUv3}} = rL_{\text{WIoUv1}}, \quad r = \frac{\beta}{\delta\alpha^{(\beta-\delta)}}, \beta = \frac{L_{\text{IoU}}^{*}}{\overline{L_{\text{IoU}}}} \tag{3}$$

L_{IoU}^{*} denotes the IoU loss of the current anchor box, $\overline{L_{IoU}}$ denotes the mean IoU loss of anchor boxes. α and δ are hyperparameters for WIoU loss.

L_{WIoUv1} is defined as follows:

$$L_{\text{WIoUv1}} = R_{\text{WIoU}} L_{\text{IoU}}, L_{\text{IoU}} = 1 - IoU \tag{4}$$

$$R_{\text{WIoU}} = \exp\left(\frac{(x - x^{\text{gt}})^2 + (y - y^{\text{gt}})^2}{w_g^2 + h_g^2}\right) \tag{5}$$

3. Experimental settings and datasets

Experiments were performed on an Ubuntu 24.04 platform (PyTorch 2.1.0, CUDA 12.1) using an AMD Ryzen 5 5600X CPU and an NVIDIA GeForce RTX 4070 SUPER GPU. We trained the model for 100 epochs with a batch size of 16. The SGD optimizer was employed with an initial learning rate of 0.01, a momentum of 0.937, and a weight decay of 0.0005.

This dataset comprises 2,768 scanned undergraduate examination papers across eight distinct subjects. As shown in **Figure 4**, we annotated four categories: Student ID, header scoring, total score, and itemized deduction areas, color-coded as red, yellow, green, and blue, respectively. The data, reflecting diverse grading styles, is split into training, validation, and test sets in an 8:1:1 ratio.

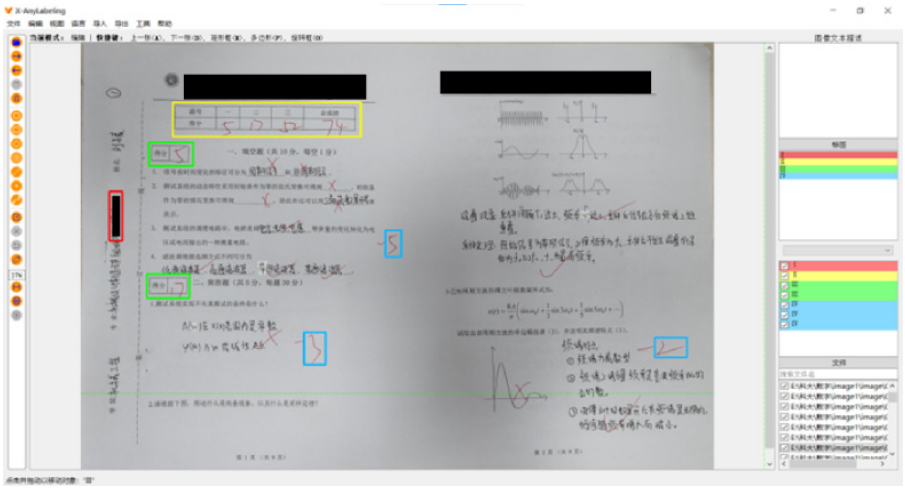

Figure 4. Data annotation.

4. Experimental results and analysis

4.1. Detection results

To visually demonstrate the model's improvement, we present a comparative analysis of YOLO11n and the enhanced model using detection results and heatmaps, as shown in **Figure 5** and **Figure 6**. In **Figure 5**, red boxes denote false positives, duplicate detections, or incomplete detections, while yellow boxes indicate missed detections; other colors represent correctly detected instances. The comparison shows that under

complex backgrounds and interference from grading annotations, YOLO11n is prone to false positives, false negatives, and duplicate detections due to its limited feature extraction capability and insufficient edge feature capture, which allows key features to be overwhelmed by noise. By introducing the DOG-Stem module, the model's edge feature extraction ability is strengthened, effectively reducing these errors and improving detection accuracy. The corresponding heatmaps in **Figure 6** further validate the effectiveness of the proposed enhancement module.

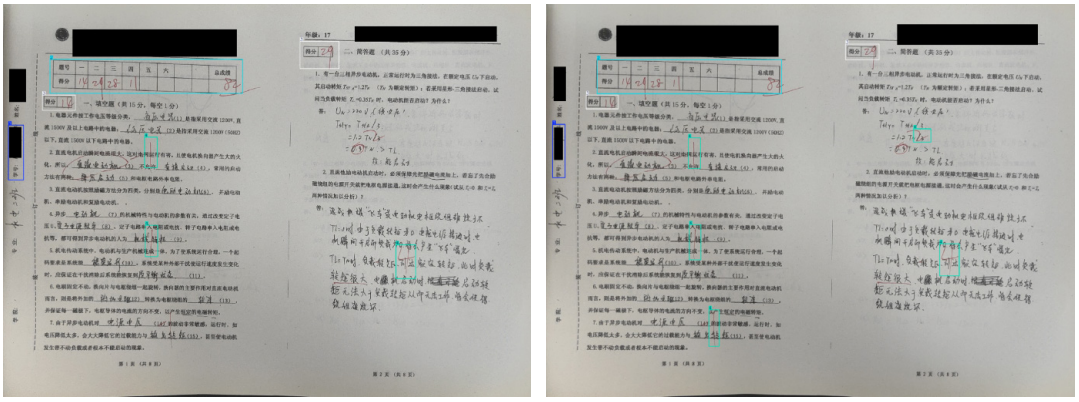

Figure 5. Comparison of detection results (Left: YOLO11n, Right: Improved model).

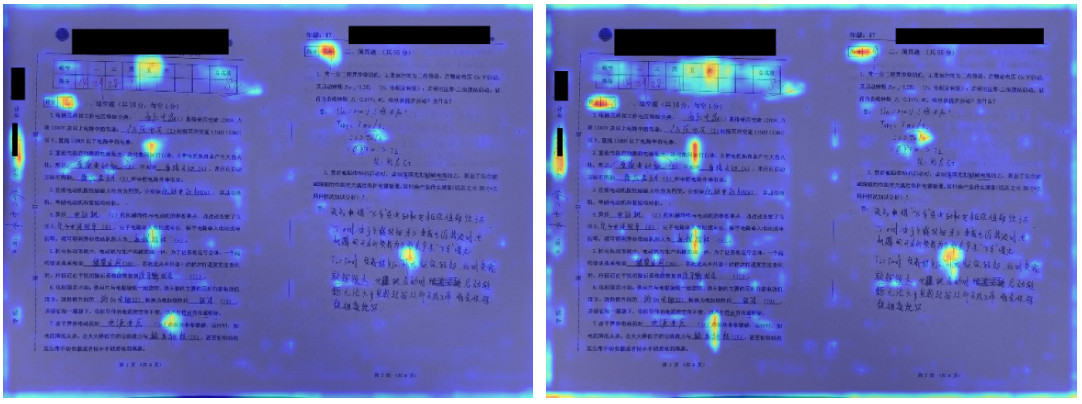

Figure 6.Heatmap comparison (Left: YOLO11n, Right: YOLO11n + DOG-Stem).

4.2. Ablation experiment

To clearly demonstrate the specific impact of integrating different improvement points on model performance, we also conducted ablation experiments. The results are shown in **Table 1**.

Table 1. Results of ablation experiments

Model	Module						
	EHead	WIoU	DOG-Stem	mAP50	$mAP_{50\text{-}95}$	Par(M)	GFLOPs
YOLO11	×	×	×	95.0	67.0	2.58	6.3
	√	×	×	95.1	67.4	2.31	5.0
	√	√	×	95.7	68.3	2.31	5.0
	√	√	√	96.3	68.6	2.31	5.3

The results show that introducing the EHead module reduces the parameter count of YOLO11n by about 10% and computational load by 20.6%, with no significant change in mAP. Further improvements, including the WIoU loss function and the DOG-Stem module, enhance overall accuracy, increasing mAP50 by 1.3%, mAP75 by 2.1%, and mAP50-95 by 1.6%. Meanwhile, integrating these modules does not noticeably increase the model's parameters or computational cost.

4.3. Comparative experiment

To validate the effectiveness of the WIoU loss function, we replaced the original YOLO11n loss with several mainstream alternatives and conducted comparative experiments. As shown in **Table 2** and **Figure 7**, WIoU not only improves overall detection accuracy but also accelerates network convergence, outperforming the other loss functions across multiple metrics.

Table 2. Comparison of loss function results

Loss function	P	R	mAP50	mAP50-95
CIoU	94.6	93.7	94.9	67.1
DIoU	95.1	94.8	95.5	68.3
EIoU	95.2	94.3	95.5	68.4
GIoU	95.6	94.3	95.4	67.6
PIoU	94.9	94.6	95.4	68.2
WIoU	95.2	95.0	95.6	68.1

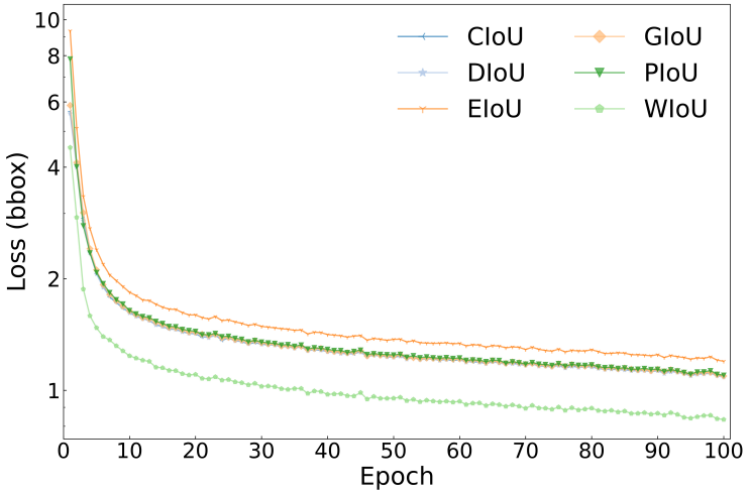

Figure 7. Loss curves for different loss functions.

To validate the model's superiority, we compared it with mainstream algorithms under identical experimental conditions, as shown in **Table 3**. The results indicate that YOLO11n achieves the best mAP50 while maintaining the lowest parameter count and computational complexity. After incorporating the proposed improvements, the model achieves a better balance between accuracy and efficiency, further demonstrating the effectiveness of the proposed enhancements.

Table 3. Comparison of experimental results across different algorithms

Method	mAP50	mAP50-95	Par(M)	GFLOPs
Faster R-CNN	88.1	57.6	41.36	90.91
Casde R-CNN	88.2	59.3	69.16	118.71
Retinanet	85.8	54.0	36.39	80.097
FCOS	85.4	50.7	32.12	78.64
DINO	91.1	60.8	46.55	118.72
YOLOv5n	94.6	67.1	2.50	7.1
YOLOv8n	94.9	67.1	3.01	8.1
YOLO11n	95.0	67.0	2.58	6.3
YOLO12n	94.0	65.5	2.56	6.3
Ours	96.3	68.6	2.31	5.3

5. Conclusion

This study proposes an improved YOLO11n-based model. The DOG-Stem module enhances edge representation and suppresses noise to generate high-quality initial features, while EHead reduces parameters and computational complexity without sacrificing accuracy. In addition, replacing the original loss with WIoU improves detection accuracy and accelerates convergence. Results show that the model performs well in complex environments, though occasional missed and false detections remain. Future work will focus on addressing these issues and optimizing the model for edge-device deployment.

Funding

The 2019 Ministry of Education Industry-University Cooperation Collaborative Education Project (Project No.: 201902077020)

Disclosure statement

The author declares no conflict of interest.

References

[1] Long S, He X, Ya H, 2018, Scene Text Detection and Recognition: The Deep Learning Era. International Journal of Computer Vision, 126(1): 1–24.

[2] Wijaya V, Soewito B, et al., 2024, Efficient License Plate Detection and Recognition with YOLOv7 and OCR. International Journal of Intelligent Systems and Applications in Engineering, 12(3): 1598–1605.

[3] Sun H, Tan C, Pang S, et al., 2024, RA-YOLOv8: An Improved YOLOv8 Seal Text Detection Method. Electronics, 13(15): 3001.

[4] Lu W, Chen S, Li H, et al., 2025, LEGNet: Lightweight Edge-Gaussian Driven Network for Low-Quality Remote Sensing Image Object Detection, arXiv, arXiv:2503.14012.

[5] Marr D, Hildreth E, 1980, Theory of Edge Detection. Proceedings of the Royal Society of London. Series B,

Biological Sciences, 207(1167): 187–217.

[6] Li C, Li L, Jiang H, et al., 2022, YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications, arXiv, arXiv:2209.02976.

[7] Zheng Z, Wang P, Liu W, et al., 2020, Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation, arXiv, arXiv:2005.03572.

[8] Tong Z, Chen Y, Xu Z, et al., 2023, Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism, arXiv, arXiv:2301.10051.