ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

Classroom Behavior Recognition and Research Based on DLKAS-YOLO8n

Jiangfeng Sun[1,a,*], Songyan Li[1,b], Jianchun Zhang[1,c]

[1]*College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454000, China*
[a]*sunjiangfeng@hpu.edu.cn*, [b]*lisongyan7050@163.com*, [c]*zhangjc@hpu.edu.cn*
[*]*Corresponding author*

Abstract: *As intelligent educational technologies advance rapidly, recognizing classroom behavior has become essential for enhancing teaching effectiveness and enabling personalized learning. Traditional recognition methods, however, encounter issues like sparse datasets, occlusions, and challenges in identifying small objects within complex classroom settings. This study introduces an enhanced YOLOv8 model, referred to as DLKAS-YOLOv8, which integrates the C2f-Faster module. First, the C2f-Faster module is incorporated to replace YOLOv8's original Bottleneck module, aiming to minimize redundant computations and boost feature extraction capabilities. Additionally, the Deformable Large Kernel Attention (DLKA) mechanism is applied to capture both detailed and broader contextual information, enhancing the model's versatility in challenging scenes. The Slide loss function is also employed to improve the model's handling of imbalanced data. Through comparative testing on a classroom behavior dataset, evaluated using mean Average Precision (mAP), Precision, and Recall metrics, DLKAS-YOLOv8 demonstrates superior performance in detection accuracy and processing speed, particularly for small and overlapping objects. Future research will integrate object tracking and facial recognition to further optimize the model's real-time capabilities and accuracy.*

Keywords: *C2f-Faster module, Deformable Large Kernel Attention (DLKA), Slide loss function, YOLOv8, Classroom behavior recognition, Deep learning*

1. Introduction

Classroom behavior recognition technology utilizes advanced deep learning algorithms to automatically capture students' actions and postures, such as raising hands, reading, sleeping, and participating in group discussions. Through the precise recognition of these behaviors, educators can objectively evaluate the effectiveness of classroom teaching and provide data-driven insights for optimizing teaching methods [1]. However, the complexity of classroom environments, diverse lighting conditions, and the variety of student behaviors present significant challenges for accurate behavior recognition [2].

With the rapid development of deep learning technology, studies [3]-[7] have explored using the classical two-stage recognition algorithm, Faster R-CNN, for classroom behavior recognition. While this algorithm achieves a certain level of accuracy, it processes the detection and classification of regions of interest in two separate steps, resulting in slower recognition speeds that fail to meet real-time requirements. To address this issue, studies [8]-[13] employed SSD, a single-stage detection algorithm, for classroom behavior recognition. However, this algorithm only predicts using the extracted feature maps without leveraging the advantages of different levels of feature maps, leading to suboptimal detection accuracy. In response, study [14] improved the YOLOv5 model by introducing the Coordinate Attention (CA) mechanism, enhancing the model's robustness without compromising detection speed. Nevertheless, it still suffers from high missed detection rates and low recognition accuracy for small and less distinguishable targets [15].

Given the challenges of low recognition speed, insufficient accuracy, and high missed detection rates for small objects in existing methods, this paper proposes an improved DLKAS-YOLO model based on YOLOv8. To suppress the background noise present in the network and enhance the focus on important student behavior features, the Deformable Large Kernel Attention (DLKA) mechanism is introduced, strengthening the model's ability to recognize student behaviors. Additionally, since the original loss function does not adequately account for the relationship between the width and height of the bounding box and their confidence scores, the Slide loss function is employed to replace the original one, improving

ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

both training speed and inference accuracy. This improved model is named DLKAS-YOLO.

2. YOLOv8 Network Model

YOLOv8 [16], known as the latest state-of-the-art (SOTA) object detection model, is celebrated for its fast detection speed and high accuracy. This model is composed of four key sections: the input layer, backbone, neck, and detection head. The input layer integrates features such as Mosaic augmentation, data enhancement, and adaptive anchor computation. The backbone utilizes the CSPDarknet architecture with convolution modules, C2f, and SPPF components to extract visual features effectively. The neck network incorporates the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN), enabling efficient fusion of high- and low-level features. For detection, three decoupled heads operate at various scales to separate detection and classification, thereby enhancing detection precision [17]. Furthermore, the model transitions from an anchor-based to an anchor-free approach, boosting localization and classification accuracy. A Task-Aligned Assigner is employed for positive-negative sample matching, combined with Distribution Focal Loss (DFL) to further refine performance. YOLOv8 offers five versions based on network depth and width: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, with YOLOv8n having the smallest parameter count. Given the computational requirements of this research, YOLOv8n is chosen as the foundational model.

3. DLKAS-YOLOv8 Algorithm for Student Classroom Behavior Recognition

Although the YOLOv8n algorithm offers high detection accuracy and speed, the accuracy of recognizing student behaviors in classrooms may decline due to local occlusion caused by densely seated and overlapping students, lighting issues, video angles, and motion blur [18]. To address these challenges, this paper improves the model in three aspects, resulting in the DLKAS-YOLOv8 model. This improved version outperforms the original model in detecting small and overlapping objects while maintaining a balance between lightweight design and performance. The effectiveness of the model has been validated on multiple datasets. The structure of the DLKAS-YOLOv8 network is shown in Figure 1

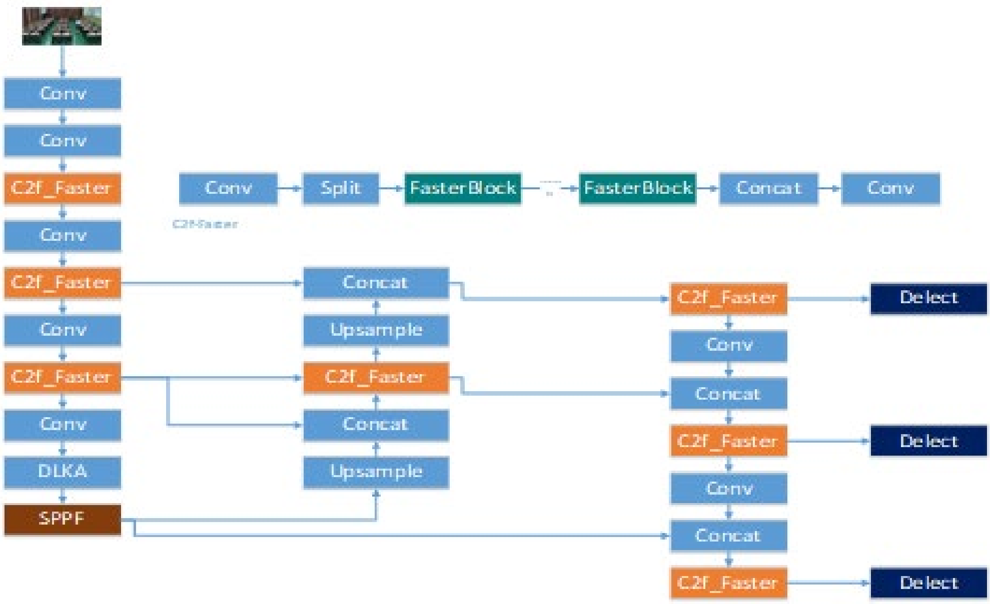

Figure 1: Structure of DLKAS-YOLOv8

1) Introduction of DLKA [19] - Deformable Large Kernel Convolutional Attention: This attention mechanism captures both fine details and broader context within images, allowing for a more thorough comprehension of image content. It adaptively emphasizes essential information while disregarding irrelevant parts, ensuring efficient use of computational resources. This functionality strengthens the model's adaptability and resilience across various scenarios, enhancing its ability to identify key features effectively.

2) Design of the C2f-Faster Module: The C2f-Faster module integrates partial convolutions, reducing the computational load during the feature fusion process. This design improves the efficiency of feature

ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

extraction while maintaining accuracy.

3) Use of the Slide Loss Function [20]: The original loss function is replaced with the Slide loss function to accelerate training and improve inference accuracy. This adjustment enhances the model's ability to handle imbalanced data and ensures more precise predictions.

3.1. Deformable Large Kernel Attention (DLKA) Mechanism

In classroom scenarios, the compact arrangement of students results in limited local feature information, making it more challenging to extract local features [21]. Consequently, global contextual information becomes crucial. By establishing long-range relationships between pixels [22], the model can capture the overall layout of the image and the relative positions of different regions. This global information helps the model better understand the image content and compensates for the insufficiency of local details to some extent [23]. To address these challenges, this paper proposes an efficient aggregation structure called Deformable Large Kernel Attention (DLKA), which leverages large kernel attention to infer target information through surrounding contextual data. The DLKA mechanism first applies a large convolutional kernel to the input feature map for initial feature extraction, integrating and restoring scattered and blurred features by capturing a broader receptive field and richer contextual information.

Next, a series of small-kernel convolution layers produce an attention map from these features. This attention map is then merged with the original input feature map, allowing the model to capture both fine edges and larger structures by utilizing convolutional kernels of varying sizes. This enhances the model’s capability to identify targets in crowded classroom environments. Finally, the output feature map is weighted according to attention scores, enabling the network to concentrate on distinctive image features, effectively compensating for any local information deficits.

To further address the challenges posed by overlapping contours and details in high-density images—where traditional convolution methods struggle to accurately capture geometric variations—Deformable Convolution[24](DCNv3) is introduced. This replaces conventional convolutions in depthwise and dilated convolutions, forming Deformable Depthwise Convolutions (DDW-Conv) and Deformable Depthwise Dilated Convolutions (DDW-D-Conv). These enhancements enable the network to better focus on the target areas.

The DLKA module utilizes deformable convolutions to adaptively adjust the sampling grid’s offsets based on the feature map, creating a dynamic rather than fixed attention distribution. This enables the network to focus more effectively on essential features of target objects while minimizing redundant computational load. Figure 2 illustrates the structure of the Deformable Large Kernel Attention (DLKA), and its mathematical formulation is provided in Equation (1).

$$\begin{gathered} Attention = Conv1\times1(DW-D(DW(F'))) \\ Output = Conv1\times1(Attention\otimes F') + F \end{gathered} \tag{1}$$

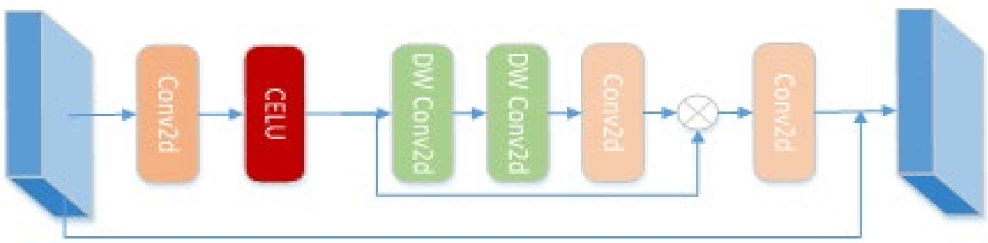

Figure 2: Structure of the DLKA Module

3.2. Design of the C2f-Faster Module

This study incorporates the partial convolution (PConv) concept from FasterNet [25] to create the Faster Block structure, replacing the Bottleneck module in C2f. The modified C2f module, named C2f-Faster, is shown in Figure 3. In the PConv layer, only 1/4 of the input channels undergo convolution, while the remaining 3/4 remain unaltered. The resulting feature map is produced by concatenating these two parts, maintaining the input's channel count and feature map size. A 1×1 pointwise convolution then

ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

doubles the channel count in the output feature map, utilizing the unaltered channels from the PConv layer to avoid information loss. Finally, a 1×1 convolution restores the channel count to its initial size, aligning the shortcut branch X dimensions with the feature map processed by the main path, ensuring seamless feature fusion between the shortcut and main branches.

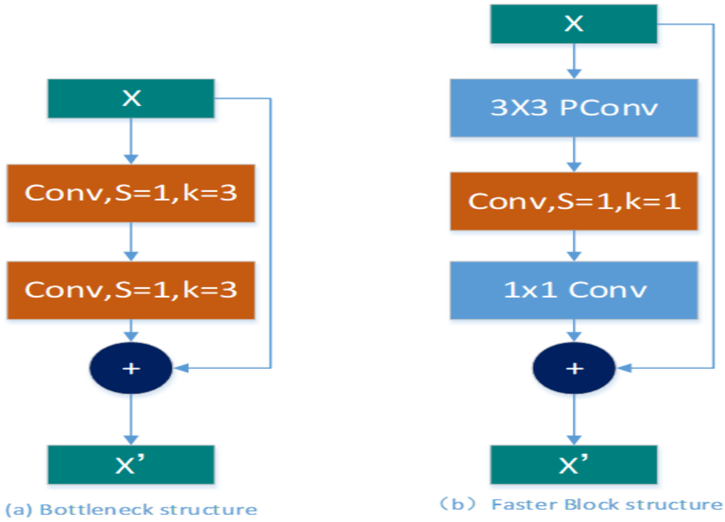

Figure 3: C2f-Faster Module

3.3. Slide Loss Function

In YOLOv8n, cross-entropy is predominantly used for classification tasks. However, with imbalanced datasets—where difficult samples are significantly fewer than easy ones—the model may overly optimize for the abundant easy samples, overlooking the challenging ones. To counter this, the Slide loss function is implemented. Slide loss offers the benefit of dynamically adjusting sample weights, particularly by assigning higher weights to difficult samples. This encourages the model to focus more on learning from these cases, enhancing detection performance for small and blurry objects. Consequently, the model achieves a more balanced and accurate detection in practical applications. Specifically, Slide loss adjusts sample weights by comparing the Intersection over Union (IoU) between the predicted and ground truth boxes. Samples with an IoU below the average are considered difficult and are assigned increased weights. This approach effectively intensifies the model's focus on boundary samples, especially those hard to classify, boosting adaptability and detection accuracy in complex scenarios. The mathematical formulation of Slide loss is provided in Equation (2).

$$f(x)=\begin{cases}1, & \text{if } x\le \mu-0.1\\ e^{1-\mu}, & \text{if } \mu<x<\mu-0.1\\ e^{1-x}, & \text{if } x\end{cases} \tag{2}$$

In this equation, xxx denotes the IoU (Intersection over Union) value between the predicted box and the ground truth box, while μ\muμ serves as the threshold to differentiate between positive and negative samples.

4. Experiments

The data for this study were gathered from real classroom environments, where student behaviors were classified into eight categories: writing with head down, reading with head down, attentive listening, head turning, hand raising, standing, group discussion, and teacher guidance. The images were annotated using the LabelImg tool and saved in YOLO format. A total of 8,431 images were collected and split into training and testing sets in a 4:1. The behavioral standards are shown in Table 1.

ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

Table 1: Behavior standard table

name	standard
write	Face down, pen in hand
Read book	Face down, book in front of you
Lecture	Face the blackboard
Turned	Turn the head.
hand	Raise arms
stand	Stand with arms down or hold a textbook
Panel discussion	Multi-person connection
Lectures	Teachers are in touch with students

4.1. Evaluation Metrics

This paper adopts the commonly used mAP (mean Average Precision) in object detection to evaluate the model's performance, while also comparing metrics such as precision, recall, and the number of model parameters. mAP evaluates the algorithm by calculating the average precision across different confidence thresholds, as shown in Equation (3).

$$
\begin{aligned}
mAP &= \frac{\sum_{i=1}^{n} Pr(i)\mathrm{d}(Re(\mathrm{i}))}{m} \\
Pr &= \frac{TP^2}{TP+FP} \\
Re &= \frac{TP}{TP+FN}
\end{aligned}
\tag{3}
$$

4.2. Comparative Experiments

In the comparative experiments, five classic object detection algorithms and the proposed method were evaluated, including RTDETR, EfficientDet, YOLOv5s, YOLOv6n, YOLOv8n, and the newly proposed DLKAS-YOLOv8. Additionally, three recently developed pose recognition algorithms—GS-YOLOv5n, SNSS-YOLOv7, and BCE-YOLOv8—were included for comparison. The performance of these models was assessed using metrics such as mAP50, precision, recall, and parameters. The experimental results are shown in the table 2.

Table 2: Comparative Experimental Results on the Custom Dataset.

Method	Dataset				
	mAP50	Precision	Recall	FLOPs(G)	Parameters
RTDETR	0.89	0.879	0.89	108.3	3.28×10^7
Efficientdet	0.651	0.71	0.706	2.391	3.83×10^6
YOLOv5s	0.887	0.8	0.863	16.0	7.04×10^6
YOLOv6n	0.848	0.848	0.823	11.9	4.24×10^6
YOLOv8n	0.861	0.871	0.836	8.2	3.01×10^6
GS-YOLOv5n	0.886	0.882	0.86	12.6	5.91×10^6
SNSS-YOLOv7	0.867	0.868	0.853	86.9	2.726×10^7
BCE-YOLOv8	0.884	0.875	0.878	18.8	3.28×10^6

According to the data in Table 2, DLKAS-YOLOv8n demonstrates significant improvements over the baseline model YOLOv8n on the custom student behavior dataset. Specifically, it achieves increases of 4.6%, 1.1%, and 6.2% in mAP50, precision, and recall, respectively. Notably, DLKAS-YOLOv8 also outperforms the second-best model, RTDETR, in these three metrics. When compared with recent advancements in the pose recognition field—GS-YOLOv5n, SNSS-YOLOv7, and BCE-YOLOv8—DLKAS-YOLOv8 achieves higher mAP50 scores by 2.1%, 4%, and 2.3%, respectively. In terms of precision, although GS-YOLOv5n performs similarly to DLKAS-YOLOv8, the parameter count of GS-YOLOv5n is 1.69 times larger. For recall, DLKAS-YOLOv8 surpasses GS-YOLOv5n, SNSS-YOLOv7, and BCE-YOLOv8 by 3.8%, 4.5%, and 2%, respectively. Based on the five metrics in Table 2, DLKAS-YOLOv8n exhibits excellent performance and accuracy in the task of student classroom behavior recognition, with the added advantage of a relatively lightweight model.

ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

4.3. Ablation Experiments

To evaluate the algorithm's effectiveness, this section carries out ablation experiments using the following configurations: YOLOv8n, YOLOv8n_C2f_Faster, YOLOv8n_DLKA, YOLOv8n_Slide, and DLKAS-YOLOv8.YOLOv8n_C2f_Faster: This model replaces the original C2f module in YOLOv8 with the enhanced C2f_Faster module, which reduces redundant computations and memory access while improving spatial feature extraction.YOLOv8n_DLKA: This configuration introduces the deformable large-kernel attention (DLKA) module at the end of the YOLOv8 backbone to enhance feature extraction.YOLOv8n_Slide: In this model, the original loss function in YOLOv8n is swapped for the Slide loss function, aimed at increasing training speed and inference accuracy.DLKAS-YOLOv8: This is the proposed algorithm, which integrates the C2f_Faster module, DLKA mechanism, and Slide loss function. These experiments comprehensively assess the impact of various enhancements, allowing for an evaluation of each modification's contribution to the model's performance.

These experiments comprehensively evaluate the effects of different enhancements. The comparative results help identify the individual contributions of each modification to the model's performance, offering a deeper insight into the advantages of the DLKAS-YOLOv8 model. The outcomes of the ablation experiments are summarized in the table. For the task of recognizing student behaviors in the classroom, the improved algorithms show notable performance improvements at every stage.

Table 3: Ablation Experiment Results

Method	C2f_Faster	DLAK	SLide	mAP50	mAP50-90	Precision	Recall	Parameters
YOLOv8n	×	×	×	0.861	0.675	0.871	0.836	3.01×10^6
YOLOv8n_ C2f_Faster	√	×	×	0.873	0.688	0.847	0.851	3.02×10^6
YOLOv8n_DLAK	×	√	×	0.876	0.709	0.851	0.872	3.4×10^6
YOLOv8n_SLid	×	×	√	0.870	0.687	0.845	0.863	3.01×10^6
DLKAS_yolov8	√	√	√	0.907	0.739	0.882	0.898	3.48×10^6

As shown in Table 3, substituting the C2f module with the enhanced C2f_Faster module leads to increases of 1.2% and 13% in mAP50 and mAP50-90, respectively, compared to the baseline YOLOv8n model. Incorporating the DLKA attention mechanism further boosts the mAP50 and mAP50-90 scores by 2.1% and 1.6%, highlighting its role in enhancing the model's ability to detect targets across different scales and to extract features from cluttered backgrounds. The addition of the Slide loss function in the YOLOv8n_Slide configuration results in gains of 0.9% and 1.2% in mAP50 and mAP50-90, respectively, indicating its effectiveness in mitigating sample imbalance between easy and difficult examples in the dataset.

The proposed DLKAS-YOLOv8 algorithm surpasses the baseline model, achieving improvements of 4.6% and 6.4% in mAP50 and mAP50-90, respectively. It also enhances precision and recall by 1.1% and 6.2%, respectively. Overall, the enhanced algorithm shows significant benefits for recognizing student behaviors in the classroom, demonstrating its effectiveness and robustness for real-world applications.

4.4. Visualization Analysis

To assess the proposed model's effectiveness in real classroom settings, two scenarios—group discussion and teacher guidance—were chosen for testing. Figure 4 displays the visualization results of various methods, revealing that the detection performance of the DLKAS-YOLOv8 model has significantly improved compared to the baseline YOLOv8 model. The DLKAS-YOLOv8 model enhances the C2f module by upgrading it to C2f_Faster, which reduces redundant calculations and memory access while fully extracting spatial features. Additionally, it incorporates the DLKA deformable large-kernel convolutional attention module to capture broader contextual information in images, addressing the accuracy issues caused by local occlusion. These enhancements enable DLKAS-YOLOv8 to perform better in scenarios involving small target dense occlusions in classroom behavior detection, effectively lowering both the false positive and false negative rates.

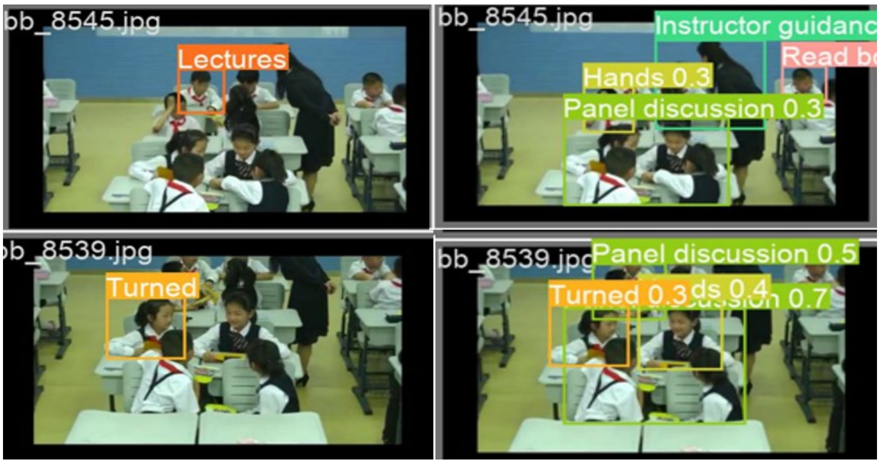

Figure 4: Comparison of Classroom Behavior Detection Results

5. Conclusion

This study proposes the DLKAS-YOLOv8 algorithm, based on the YOLOv8 framework, to efficiently recognize student behavior in the classroom. The model introduces the FasterBlock module from FasterNet to replace the Bottleneck module in C2F, along with the DLKA attention mechanism, enabling the model to adaptively focus on critical information in images while filtering out irrelevant data for efficient resource allocation. Additionally, the Slide Loss function addresses the issue of sample imbalance within the dataset. Experimental results confirm that DLKAS-YOLOv8 demonstrates excellent performance in student behavior recognition, with significant advantages in model lightweighting. While the model achieves high precision, there remains room for improvement in accuracy.

High-quality datasets and optimized network structures are essential for enhancing model accuracy, efficiency, and generalization ability while reducing complexity. Future work will focus on constructing high-quality datasets and further refining the network structure to improve the model's lightweight design and precision. Additionally, future research will explore integrating object tracking and facial recognition algorithms to establish real-time associations between student facial information and behavior. This study not only contributes an innovative model for student behavior analysis in the field of computer vision and behavioral recognition but also provides practical solutions for educational technology applications, demonstrating the potential of deep learning in real-world teaching scenarios.

References

[1] Chen Li. Research on "Students' Classroom Problem Behaviors" from the Perspective of Senior High School Teachers [D]. Nanchang University.,2024.DOI:10.27232/d.cnki.gnchu.2024.001111.

[2] Zhao Ke. Methods and Implementations for Human Skeleton Behavior Recognition in Classroom Scenarios [D]. Shenyang University,2024.DOI:10.27692/d.cnki.gsydx.2024.000125.

[3] Zhou Ye. Research on the Detection of Primary School Students' Classroom Behaviors Based on Faster R-CNN [D]. Sichuan Normal University,2021.DOI:10.27347/d.cnki.gssdu.2021.000962.

[4] GUO M H, LU C Z, LIU Z N, et al. Visual attention network[J]. Computational Visual Media, 2023, 9(4): 733-752.

[5] HOIEM D, DIVVALA S K, Hays J H. Pascal VOC 2008 challenge[J]. World Literature Today, 2009, 24(1): 1-4.

[6] PADILLA R, NETTO S L, Da Silva E A B. A survey on performance metrics for object-detection algorithms [C]/ 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020: IEEE, 2020: 237-242.

[7] ZHAO Y, KV W, XU S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, 2024: IEEE, 2024: 16965- 16974.

[8] TAN M, PANG R, Le Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, 2020: IEEE, 2020: 10781-10790.

ISSN 2616-5775 Vol. 7, Issue 11: 109-116, DOI: 10.25236/AJCIS.2024.071115

[9] MUSHTAQ F, RAMESH K, DESHMUKH S, et al. Nuts&bolts: YOLO-v5 and image processing based component identification system[J]. Engineering Applications of Artificial Intelligence, 2023, 118: 105665

[10] ATREY J, REGUNATHAN R, RAJASEKARAN R. Realworld application of face mask detection system using YOLOv6 [J]. International Journal of Critical Infrastructures, 2024, 20(3): 216-240.

[11] PADILLA R, NETTO S L, Da Silva E A B. A survey on performance metrics for object-detection algorithms [C]/ 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020: IEEE, 2020: 237-242.

[12] ZHAO Y, KV W, XU S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, 2024: IEEE, 2024: 16965- 16974.

[13] TAN M, PANG R, Le Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, 2020: IEEE, 2020: 10781-10790.

[14] Cheng Wenhao. Research on Classroom Student Detection and Behavior Recognition Based on YOLOv5 Deep Learning [D]. Chang'an University,2023.DOI:10.26976/d.cnki.gchau.2023.002213.

[15] Gao Fei. Research on Human Pose Estimation Based on Multi-Network Fusion [D]. Dalian Jiaotong University,2024.DOI:10.26990/d.cnki.gsltc.2024.000012.

[16] JOCHE G, CHAURASIA A. (2023). Ultralytics/YOLOv8 in PyTorch. URL: https://github. com/ ultralytics/ultralytic

[17] Chen Yuling. Research on Small Target Feature Enhancement and Detection Methods Based on Convolutional Neural Network [D]. Southwest University of Science and Technology, 2024. DOI:10.27415/d.cnki.gxngc.2024.000164.

[18] Zeng Yuqi, Liu Bo, Zhong Baichang, et al. Student Classroom Behavior Detection Algorithm Based on Improved YOLOv8 in the Context of Smart Education [J]. Computer Engineering,2024,50(09):344-355. DOI:10.19678/j.issn.1000-3428.0069597.

[19] Azad R, Niggemeier L, Hüttemann M, et al. Beyond self-attention: Deformable large kernel attention for medical image segmentation [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 1287-1297.

[20] YU Z,HUANG H,CHEN W,et al.YOLO-faceV2:a scale and occlusion aware face detector[J].Pattern Recognition,2024,155:110714.

[21] WANG W, DAI J,CHEN Z, et al. Intern image: Exploring large-scale vision foundation models with deformable convolutions[C]// 2023 IEEE Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 14408-14419.

[22] Cui Shaoguo, Xu Song, Wang Mingyang, et al. Research Progress of Deep Learning Knowledge Tracing for Intelligent Education [J/OL]. Computer Engineering,1-18[2024-10-27]. https://doi.org/10.19678/j.issn.1000-3428.0069653.

[23] Liu Genghuan, Zeng Xiangjin, Dou Jiazhen, et al. Research Progress of Small Target Detection Technology Based on Deep Learning (Invited) [J]. Infrared and Laser Engineering,2024,53(09):196-228.

[24] Chen Chen, Bao Wenxing, Chen Xu, et al. Student Classroom Behavior Recognition Algorithm Based on Improved YOLOv8: DMS-YOLOv8 [J/OL]. Computer Engineering and Applications,1-17[2024-10-27].http://kns.cnki.net/kcms/detail/11.2127.tp.20240913.1601.008.html.

[25] Chen J, Kao S, He H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023:12021-12031.