Student Classroom Behavior Image Recognition Based on YOLOv7

Siqing Chen[a]
School of Mathematics and Statistics, Hanshan Normal University, Chaozhou, China

Keywords: Intelligent Classroom Management System, Computer Vision, Automatic Recognition, Object Detection Algorithm.

Abstract: With the rapid development of educational informationization, intelligent classroom management systems have become the key methods to enhance the teaching quality and efficiency. However, the traditional classroom management methods mainly depend on teachers' observation and records, which is not only inefficient but also susceptible to subjective factors, leading to inaccuracy. In order to overcome these limitation, this paper uses the computer vision technology to automatically recognize students' classroom behaviors has become an important research direction in the field of educational technology. Among them, the You Only Look Once vision7 (YOLOv7) , which famous for its detection speed and accuracy, is well suited for real-time classroom action recognition and is a leading algorithm in this field. Using YOLOv7, educators can obtain objective analysis of the classroom conditions and teaching effectiveness to optimizing instructional strategies and providing personalized learning support. Moreover, the collection and analysis of student behaviour data do contribute to the school management and education methods optimization, promoting the development of educational management mode to a more scientific and fine direction.

1 INTRODUCTION

With the rapid development of educational informatization, intelligent classroom management systems have become a key method for improving teaching quality and efficiency. In the past, traditional classroom management mainly relied on teachers' observation and manual record-keeping. This method was not only time-consuming and labor-intensive, but also prone to errors due to subjective factors. Teachers might miss important behavioral cues or misunderstand them, thereby leading to potential biases in the assessment.

To overcome these limitations and promote the progress in the field of educational technology, researchers turned their attention to computer vision technology. This technology provides a promising solution for more objective and accurate automatic identification of students' classroom behaviors. By leveraging advanced algorithms and machine learning models, computer vision can analyze real-time video streams or images to detect and classify various classroom behaviors.

Among the various existing computer vision algorithms, You Only Look Once version 7 (YOLOv7) stands out as an efficient object detection algorithm. It combines fast detection speed and high accuracy, making it particularly suitable for real-time classroom behavior recognition. YOLOv7 can process a large amount of data quickly and accurately, enabling it to keep up with the dynamics of classroom interactions and provide teachers with timely and reliable information about students' behaviors.

In conclusion, integrating computer vision technology, especially YOLOv7, into intelligent classroom management systems marks an important step in leveraging advanced technologies to improve educational outcomes. Through the automatic identification of classroom behaviors, YOLOv7 enables teachers to focus more on teaching rather than manual observation, ultimately enhancing the overall quality and efficiency of the educational process.

2 RESEARCH OBJECTIVE

Utilizing the efficient detection capability of

[a] https://orcid.org/0009-0001-5038-6648

Chen, S.
Student Classroom Behavior Image Recognition Based on YOLOv7.
DOI: 10.5220/0013703500004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 649-653
ISBN: 978-989-758-765-8

YOLOv7 (Zhang， 2024), real-time monitoring of students' behavior in the classroom is conducted to promptly detect and correct negative behaviors (Redmon& Farhadi, 2018), such as lack of concentration and playfulness. By identifying students' behavior, objective data about classroom atmosphere and teaching effectiveness can be provided to teachers, helping them adjust teaching strategies and improve classroom quality. Analyze students' behavior patterns, understand their learning habits and needs, and provide personalized learning support and guidance for them.

Collect and analyze student behavior data to provide data support for school management and educational decision-making, and promote the scientific and refined management of education. Explore the application of YOLOv7 in the field of education, promote the development of computer vision technology in educational research, and provide new ideas and methods for research in related fields.

Classification and Monitoring of Behaviors in Educational Environments:

Classification and monitoring of behaviors have become a crucial aspect in the field of educational technology, and YOLOv7 has emerged as a powerful tool for achieving this goal. Numerous studies have utilized the capabilities of YOLOv7 to classify and detect students' behaviors in classroom settings (Li, 2025). Typically, these studies employ datasets containing multiple behavior labels to conduct comprehensive training of the models. The main focus of these tasks lies in designing a reasonable and comprehensive labeling system and implementing effective data augmentation strategies to enhance the robustness of the models.

Construction of Behavior Recognition Datasets:

A fundamental aspect of any research work in this field is the construction of high-quality behavior recognition datasets. These datasets serve as the cornerstone for training and evaluating models. Recognizing this, some researchers have embarked on ambitious projects aimed at collecting behavior data in diverse classroom environments. These efforts aim to provide a comprehensive and representative datasets for the development and improvement of behavior recognition models.

Real-time Application of YOLOv7 in Classrooms:

Thanks to the real-time detection capabilities of YOLOv7, its potential for application in real classroom environments has attracted widespread attention. Researchers are currently exploring methods to integrate YOLOv7 into classroom management systems, with the ultimate goal of enhancing teaching effectiveness and student engagement. Additionally, the real-time feedback provided by YOLOv7 is invaluable to teachers, enabling them to adjust teaching strategies in real time based on students' behavior and engagement levels (Wang et al., 2022).

3 RESEARCH METHOD

The aim of this experiment is to use the YOLOv7 object detection algorithm(Zhang et al. , 2021) to achieve image recognition of three typical behaviors among students in the classroom: "raising hands", "reading", and "writing". The research content mainly includes the following aspects:

Dataset construction: Collect and annotate classroom images containing behaviors such as "raising hands," "reading," and "writing," and construct high-quality training and validation datasets.

Model training: Use YOLOv7 algorithm to train the dataset, optimize model parameters, and improve the accuracy and robustness of behavior recognition (Zhang & Li, 2025).

Performance evaluation: Evaluate the trained model through a test set and analyze its recognition performance in classroom scenarios, including metrics such as accuracy, recall, and real-time performance.

Application validation: Deploy the model in an actual classroom environment to verify its feasibility and practicality in practical applications.

The research objective is to develop an efficient and accurate student classroom behavior recognition program, providing technical support for intelligent classroom management, helping teachers to grasp students' learning status in real time, optimize teaching management strategies, and improve classroom efficiency and learning outcomes.

Created by Chengdu Neusoft College, it contains 5686 images and 45578 tags, covering six behaviors: raising hands, reading, writing, using mobile phones, lowering heads, and lying on the table. This experiment only tests three behaviors: raising hands, reading, and writing. The dataset covers different scenarios from kindergarten to university, and was evaluated using the YOLOv7 algorithm with an average accuracy of 80.3%, as shown in Figure 1. This dataset aims to provide a solid foundation for research on student behavior detection.

Original address: https://github.com/Whiffe /SCB-dataset?tab=readme -ov-file

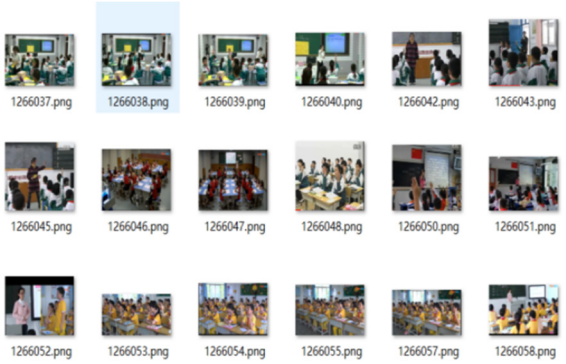

Figure 1: algorithm evaluation diagram for YOLOv7 (Picture credit: Original)

4 RESEARCH RESULT

4.1 Model Structure

The YOLOv7 model used in this experiment is an efficient single-stage object detection algorithm, whose core structure includes a backbone network, a feature pyramid network, and a detection head. YOLOv7 has undergone multiple optimizations based on the YOLO series, resulting in higher detection accuracy and faster inference speed. The specific structure is as follows:

a. Backbone (Chen et al., 2021): Using CSPDarknet as the backbone network, multi-level features are extracted through a cross stage local connection (CSP) structure to enhance the model's feature extraction capability.

b. Feature Pyramid Network (Neck) (Chen, 2024): Using Path Aggregation Network (PANet) and multi-scale feature fusion techniques, combining shallow and deep feature information to enhance the model's detection ability for small and large targets.

c. Head(Zhou, 2024): Based on the anchor mechanism, the classification branch and regression branch are used to predict the category and bounding box position of the target, respectively. At the same time, a dynamic label allocation strategy is introduced to optimize the training process.

YOLOv7 also introduces Model Scaling technology(Tu, 2023), which achieves a balance between model performance and computational efficiency by adjusting the depth, width, and resolution of the network, making it more suitable for practical application scenarios. This experiment utilizes the advantage of YOLOv7 to fine tune the model for classroom behavior recognition tasks, in order to achieve accurate detection of "raising hands", "reading", and "writing" behaviors

4.2 Experimental Result

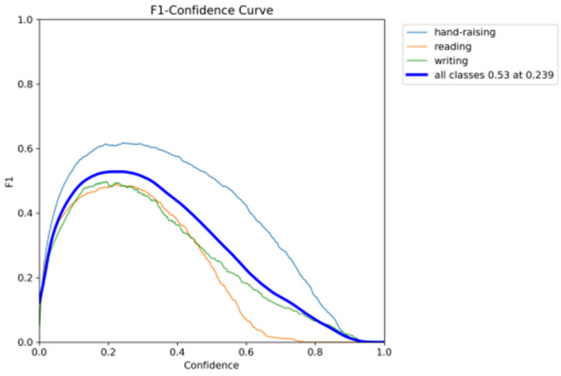

Figure 2: F1-Confidence Curve (Picture credit: Original)

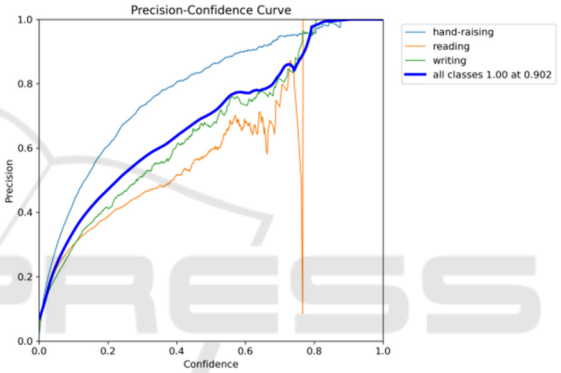

Figure 3: Precision-Confidence Curve (Picture credit: Original)

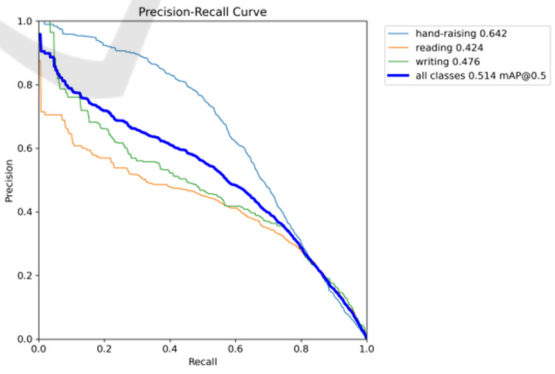

Figure 4: Precision-Recall Curve (Picture credit: Original)

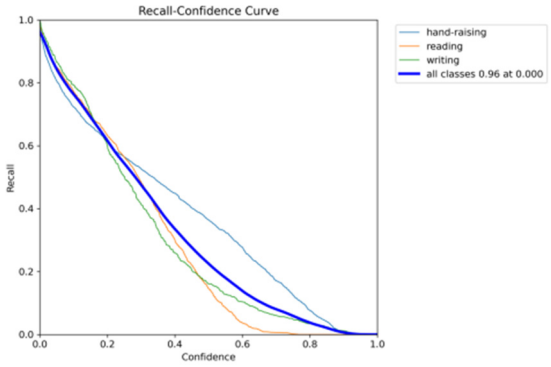

Figure 5: Recall-Confidence Curve (Picture credit: Original)

F1-Confidence Curve: Figure 2 shows the F1 scores of the model at different confidence thresholds. The F1 score for 'hand raising' behavior is the highest, indicating that the model performs the best in recognizing hand raising behavior.

Precision-Confidence Curve:

Figure 3 shows the accuracy of the model at different confidence thresholds. At high confidence (0.902), "all classes" achieved the highest accuracy of 1.00, indicating that the model's predictions were very accurate at high confidence.

Precision-Recall Curve:

Figure 4 shows the accuracy of the model at different recall rates. The accuracy of the "hand raising" behavior is the highest (0.662), indicating that the model not only has high accuracy in recognizing hand raising behavior, but also has a relatively good recall rate.

Recall-Confidence Curve:

Figure 5 shows the recall rate of the model at different confidence thresholds. The recall rate of 'all classes' is highest at a medium confidence level (0.606), indicating that the model can recognize most behavior instances at this confidence level.

Based on these charts, this article can draw the following conclusions:

The model performs the best in identifying "raising hands" behavior, whether in terms of F1 score, accuracy, or recall.

The accuracy of the model is very high under high confidence, which may mean that the model can make accurate predictions in very certain situations.

The recall rate decreases with increasing confidence, which is expected because higher confidence means that the model will only predict more certain instances, which may result in missing some correct behavior instances.

These charts provide a comprehensive perspective on the performance of the model, helping to understand its performance at different confidence thresholds and providing a basis for further optimizing the model. As shown in Figure 6.

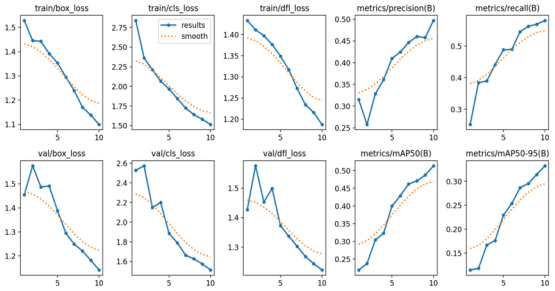

Figure 6: performance chances during model training (Picture credit: Original)

5 CONCLUSIONS

This experiment utilized the YOLOv7 algorithm to conduct image recognition for three types of classroom behaviors: "raising hands", "reading", and "writing". By constructing and annotating a high-quality datasets, we trained and optimized the model, thereby enhancing its accuracy and robustness in recognition. The experimental results demonstrated that the model performed best in recognizing the "raising hands" behavior, and exhibited extremely high accuracy even at high confidence levels. As the confidence level increased, the recall rate also improved.

And, YOLOv7 can effectively be applied to classroom behavior recognition, providing technical support for intelligent classroom management systems and helping teachers understand student dynamics in real time, thereby improving teaching efficiency. Future work will focus on further optimizing the model to enhance its ability to recognize more behaviors and exploring its applicability in different teaching environments to achieve wider application.

These findings not only confirmed the effectiveness of the YOLOv7 algorithm in the field of classroom behavior recognition but also provided solid technical support for intelligent classroom management systems. By capturing and analyzing students' behavior data in real time, teachers can more intuitively understand the classroom atmosphere and student dynamics, enabling them to adjust teaching strategies in a timely manner and improving teaching efficiency. In addition, intelligent classroom management systems can provide data support for personalized learning, helping students receive more precise learning guidance.

Looking ahead, researchers will continue to optimize the YOLOv7 model to enhance its ability. At the same time, researchers will explore the potential of this algorithm in different environments

to achieve wider application. Researchers believe that with the continuous progress and improvement of technology, intelligent classroom management systems will play an increasingly important role in the education field, creating a more efficient and convenient learning environment for teachers and students. Therefore, conducting in-depth research and exploration on the YOLOv7 algorithm and its application in classroom behavior recognition has important practical significance and broad application prospects.

REFERENCES

Chen, L., Yang, J., Cao, T., Zheng, Y., Wang, Y., Zhang, B., Lin, Z., & Li, W. 2025. A Self-distillation Object Segmentation Method Based on Transformer Feature Pyramid. Journal of Electronics & Information Technology.

Chen, Y., Li, J., & Liang, J. 2024. Multi-scale feature cross fusion for automatic grading network of sketch images. School of Software Engineering, South China Normal University.

Li, K. 2024. Research and application of Classroom Abnormal behavior Detection Algorithm based on YOLOv7 (Master's thesis, Xi 'an University of Architecture and Technology). Master's https://link.cnki.net/doi/10.27393/d.cnki.gxazu.2024.000163 doi: 10.27393 /, dc nki. Gxazu. 2024.000163.

Redmon, J., & Farhadi, A. 2018. YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Tu, C. 2023. Research on Lightweight Technology and Application based on YOLOv5 Target Detection Network (Master's Thesis, Southwest Jiaotong University). Master of https://link.cnki.net/doi/10.27414/d.cnki.gxnju.2023.000866.

Wang, C., et al. 2022. YOLOv7: A high-performance detector for real-time object detection. arXiv preprint arXiv:2207.08430.

Zhang X. 2024. Research on Multi-rotor UAV Recognition Algorithm and System based on YOLOv7 Series (Master's Thesis, Jiangxi University of Science and Technology). Master of https://link.cnki.net/doi/10.27176/d.cnki.gnfyc.2024.000293doi:10.27176/d.cnki.gnfyc.2024.000293.

Zhang, Y., & Li, J. 2025. Scale-free network Robust optimization algorithm against malicious attacks based on multi-granularity integration. Journal of kunming university of science and technology (natural science edition) 1-20. Doi: 10.16112 / j.carol carroll nki. 53-1223 / n. 2025.01.231.

Zhang, Y., et al. 2021. Classroom behavior recognition based on deep learning technology. Journal of Educational Technology.

Zhou, X., Wang, K., Zhou, X., & Han, J. 2024. An electric vehicle helmet wearing detection algorithm based on YOLOv10n. Electronic Measurement Technology.