Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-8, August 2026
DOI: 10.66069/ojspub.16560803

Teacher Behavior Intelligent Recognition based on YOLO Framework and SPPF-Attention

Liqiong Lu[1,2], Dong Wu[1,2], Yongheng Chen[1], Zhongyan Liu[1], Ziwei Tang[3,*]

[1]School of Computer Science and Intelligence Education, Lingnan Normal University, Zhanjiang 524048, China
[2]Guangdong Provincial Key Laboratory of Development and Education for Special Needs Children, Lingnan Normal University, Zhanjiang 524048, China
[3]Taizhou University, College of Information Engineering, Taizhou, 225300, China
Correspondence Author, tangziwei@tzu.edu.cn

Abstract: ***The behavior of teachers in the classroom has a crucial impact on students. In the past, experts mainly evaluated teachers' performance in the classroom by observing their behaviors and providing ratings. This method is usually inefficient, time-consuming, and subjective. A teacher behavior intelligent recognition method was proposed in this paper. Based on the framework of YOLO V11, Attention mechanism and SPPF module was combined to form a new module SPPF-Attention which used to explore key features from feature layers of different sizes to recognize teacher behaviors in the classroom. Experimental results on SBC5_Teacher_Behavior dataset showed the proposed method obtained mAP value of 0.940 and had better recognition performance compared with other classic methods such as YOLO V8, YOLO V11, YOLO V12 and YOLO V13.***

Keywords: Teacher behavior recognition, YOLO Framework, SPPF-Attention, CNN.

1. Introduction

The evaluation of teachers' classroom performance is an important issue in the field of education. By comprehensively evaluating the performance of teachers in the teaching process, Teachers' teaching level, teaching effectiveness and impact on students can be better understood. Evaluating the performance of teachers also can provide effective reference for improving teaching quality and promoting students' comprehensive development. However, evaluating teachers' classroom performance has always been a challenging task. In the past, experts mainly evaluated teachers' performance in the classroom by observing their behaviors and providing ratings. This method has some limitations. Firstly, this requires evaluators to have rich relevant background knowledge and professional experience, which may not be easy to achieve in some cases, especially in large-scale evaluations. Secondly, this method is usually inefficient, time-consuming, and subjective, as the evaluators' subjective judgment may be influenced by factors such as personal preferences, observation perspectives, and experience. Therefore, seeking more effective and objective evaluation methods is one of the important challenges facing the current education field.

The development of artificial intelligence, especially CNN technology, has brought new developments to the intelligent evaluation of teacher classrooms. The use of artificial intelligence technology to automatically recognize teacher behaviors in the classroom and subsequently evaluate the quality of classroom teaching using teacher behavior has become a new trend [1-3]. This method can not only improve the objectivity and efficiency of evaluation, but also provide teachers with more timely and accurate feedback, promoting their professional growth.

There are currently two main types of intelligent recognition methods for teacher behaviors. The first type is based on human pose recognition methods, and the second type is a direct recognition method using CNN. Li et al. [4] used OpenPose to extract teacher skeleton information and features. Then, the skeleton data was input into the DenseNet network to classify the teacher's actions. They adopted a behavioral activity evaluation model to process body vectors formed by human keypoint coordinates, introducing the concept of limb weights, then rated teachers' activities based on the angles between body vectors. Nida et al. [5] observed the spatiotemporal relationships within video frames and used graph cuts to segment the contours of teachers from input videos, generating motion contours. These motion contours were represented as feature maps using deep convolutional neural networks. Then machine learning classifiers were used to classify the contours. Wu et al. [6] fused video information and skeletal information for recognizing teacher behaviors. Peng et al. [7] proposed an improved gesture recognition framework based on keypoint statistical transformation features, using a deep convolutional neural network model with multi-scale feature fusion to recognize teacher gestures. Li et al. [9] proposed a new method based on skeleton behavior recognition and lagged sequence analysis to identify teachers' classrooms behavior.

The second method use different CNN structure and feature fused design to identify teacher behaviors. Ma et al. [1] proposed an automatic classroom analysis strategy for analyzing teaching methods in classroom environments. They utilized a three-dimensional convolutional neural network with time distribution layers and Conv2DLSTM to detect teacher behaviors in instructional videos. Chen et al. [2] improved YOLOv4 [9] by constructing and embedding an embedded connection component at the network end to detect teacher and student behaviors in the classroom. Based on the YOLO V8 [10] framework, Jing et al. [11] improved the backbone network and detection head to recognize teacher behaviors.

The YOLO series methods stand out in the field of object recognition due to their excellent recognition performance, especially YOLO V11[12], which achieves an excellent balance between recognition accuracy and recognition speed.

Teacher classroom behavior recognition not only requires accuracy in recognition, but also has certain requirements for recognition time efficiency. Therefore, based on the framework of YOLO V11 and designing a method for mining teacher behavior features according to the characteristics of teacher classroom behavior, is a good choice. In this paper, a new teacher behavior recognition was proposed based on the framework of YOLO V11 and SPPF-Attention. Teachers exist independently in the classroom, with significant differences from students in terms of clothing, behavior, and positioning. Additionally, teachers may appear in different positions within the classroom. Therefore, this paper embedded attention mechanism into the SPPF module to explore key features of feature layers with different sizes and use them to recognize teacher behaviors. Our proposed method was used to compare with other excellent methods in the field of object recognition such as YOLO V8, YOLO V11, YOLO V12[13] and YOLO V13[14] on SBC5_Teacher_Behavior dataset [15]. Our method achieved mAP value of 0.941, was better than 0.931 in YOLO V11, 0.930 in YOLO V12, 0.928 in YOLO V8 and 0.893 in YOLO V13.

2. Teacher Behavior Recognition Based on YOLO Framework and SPPF-Attention

The framework of our proposed teacher behavior recognition method has been shown in Figure 1. The left of backbone network consisting of Conv, C3K2, C2PSA and SPPF-Attention modules is used to mine teacher behavior features. The right of backbone network consisting of Upsample, Concat and C3K2 modules is utilized to fuse feature layers of different sizes. Then the fused features are sent to three different detection heads for performing classification and regression tasks.

Compared to the framework of YOLO V11, the biggest difference is SPPF-Attention module. SPPF-Attention module is specially designed to mine the key teacher behavior features of different size. Considering that the identity of a teacher in a classroom is very special. Firstly, generally speaking, there is only one teacher; Secondly, teachers have essential differences from students in terms of clothing, behavior, demeanor, and posture. Finally, the differences in teachers' behavior are often reflected in some subtle aspects, such as standing and guiding, blackboard writing, and backboard. Attention map is designed to added to SPPF module, and the value of attention map is related to the variance of feature maps of different sizes.

As shown in Figure 2, compared to SPPF, an attention map in SPPF-Attention module is used to perform dot multiplication with the convolutional feature layer to obtain a new feature layer which can mine the key features of teacher behaviors. Attention map is used twice, which is performed with two feature layers with different sizes.

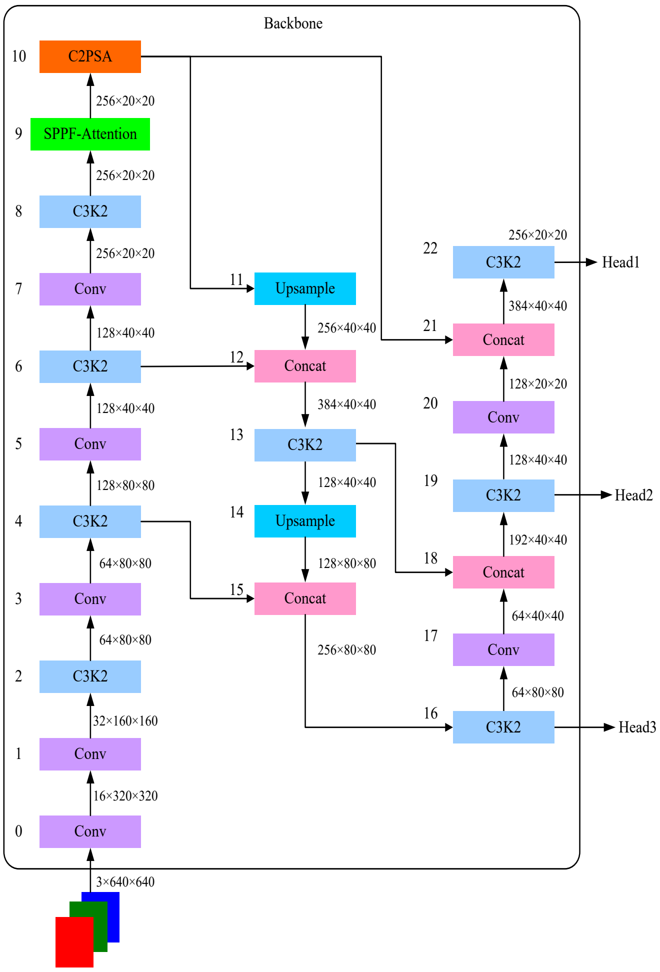

Figure 1: The framework of our proposed method

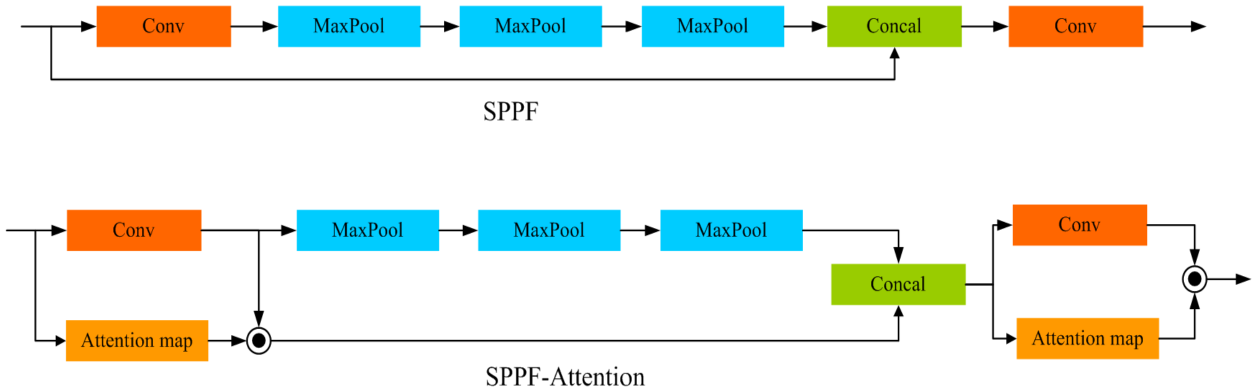

Figure 2: The design of SPPF and SPPF-Attention

Figure 3: Different teacher behaviors on SBC5_Teacher_Behavior dataset.

3. Experimental Analysis

3.1 Experimental Set

In this paper, all methods are implemented based on Pytorch 2.2 and are all trained and tested using Geforce 3060 graphics card. Some parameters in the experiment process, such as learning rate, are consistent with the initial setting of the method, and the value of bath size is adjusted according to the size of GPU memory. YOLO V8, YOLO V11, our proposed method, YOLO V12 and YOLO V13 use yolo8s.pt, yolo11s.pt, yolo11s.pt, yolo12s.pt and yolo13s.pt to initialize the parameters respectively.

3.2 Performance Evaluation Indicators for Teacher Behavior Recognition

Accuracy (Precision, *P*), recall (Recall, *R*) and comprehensive indicators *F* and mean average precision (*mAP*) are used to evaluate the recognition performance of classroom student behaviors.

$$P = \frac{TP}{FP+TP};$$

$$R = \frac{TP}{TP+FN};$$

$$F = \frac{2 \times P \times R}{P+R};$$

Among them, *TP* (true positive) represents the number of correctly recognized samples; *FP* (false positive) represents the number of error recognized samples; *FN* (false negative) represents the number of undetected samples. Here, when the IOU between the predicted box and the ground truth box exceeds 50%, it is considered correct recognition. *mAP* is the average of the average precision values for all categories, calculated based on the precision recall curve.

3.3 SBC5_Teacher_Behavior Dataset

SCB-Dataset5 is a public dataset constructed by Jinan University.SCB5_Teacher_behavior_Stand_BlackBoard_Screen is a sub dataset that contains teacher behavior and objects related to teacher behavior, such as blackboard and screen. Specifically, the dataset contains eight categories, namely guide, answer, On-stage interaction, blackboard-writing, teacher, stand, screen and blackboard. There are 8983 images in the training set and 3240 images in the testing set. The dataset can be downloaded on two websites including https://github.com/Whiffe/SCB-dataset and https://pan.baidu.com/s/1keysWqSPnVCftQqDogKb4w?pwd=SCB5. Figure 3 has shown different teacher behaviors on this dataset.

3.4 Comparative Analysis of Recognition Methods of Teacher Behaviors

Table 1 lists the recognition results of different methods on SBC5_Teacher_Behavior dataset. From the experimental results listed in Table 1, Our method has the highest value of *mAP*, reaching 0.940, followed by YOLO V11 and YOLO V12 with *mAP* value of 0.931 and 0.930, followed by YOLO V8 and V13 with *mAP* value of 0.923 and 0.893. From the F indicator, our method still achieved the highest value of 0.914. Especially, compared with YOLO V11, our method has improved both accuracy(*P*) and recall(*R*). Experimental results has shown that SPPF-Attention module can improve the accuracy and regression rate of target recognition by mining key features from feature layers of different sizes.

Table 1: Comparison of recognition performance

Method	*P*	*R*	*F*	*mAP-50*
YOLO V8	0.911	0.876	0.893	0.923
YOLO V11	0.898	0.9	0.899	0.931
YOLO V12	0.908	0.886	0.897	0.930
YOLO V13	0.849	0.862	0.855	0.893
Our method	**0.918**	**0.910**	**0.914**	**0.940**

Table 2 lists the detailed AP values for each teacher behavior category. Our proposed method has the highest value in *AP-guide*, *AP-answer*, *AP-On-stage-interaction* and *AP-teacher*. YOLO V11 has the best performance in *AP-stand* and *AP-blackBoard*. YOLOV8 has the best performance in *AP-blackboard-writing* and *AP-screen*. These experimental results revealed that our method has significant advantages in object recognition with large scale variations, such as guide, answer, and On-stage interaction. Experimental results once again proved that our strategy of adding attention maps into SPPF module is correct, and mining key features on different scale feature layers is indeed beneficial for improving the performance of teacher behavior recognition.

3.5 Teacher Behavior Recognition Samples

Teacher behavior recognition samples has shown in Figure 4. From Figure 4, it can be seen that two behaviors of teacher and stand often occur together, and teacher and blackboard, or teacher and screen, often appear together. And our method can basically recognize them normally. In addition, due to the inclusion of both teacher and student objects in the guide behavior, some images mistakenly recognize the guide behavior as teacher. In the future, we will consider introducing transformer mechanism to improve this situation.

Table 2: Comparison of AP for each category

Method	*AP-guide*	*AP-answer*	*AP-On-stage-interaction*	*AP-blackboard-writing*	*AP-teacher*	*AP-stand*	*AP-screen*	*AP-blackBoard*	*mAP-50*
YOLO V8	0.741	0.898	0.889	**0.982**	0.963	0.953	**0.98**	0.982	0.923
YOLO V11	0.779	0.921	0.888	0.973	0.969	**0.959**	0.978	**0.984**	0.931
YOLO V12	0.772	0.913	0.896	0.977	0.972	0.95	0.979	0.983	0.930
YOLO V13	0.71	0.848	0.782	0.966	0.948	0.931	**0.98**	0.979	0.893
Our method	**0.821**	**0.933**	**0.914**	0.974	**0.975**	0.948	0.976	0.98	**0.940**

Figure 4: Teacher behavior recognition sample

4. Conclusion

This paper focuses on the recognition of teacher behaviors. In order to solve the problem of time-consuming and laborious in traditional classroom teaching behavior evaluation methods, this paper designed SPPF-Attention module based on the framework of YOLO V11 to mine the key characteristics of feature layers of different sizes to improve the performance of teacher behavior recognition. The proposed method was validated on the BC5_Teacher_Behavior dataset and compared with other classic object recognition methods. Experimental results has shown that our proposed method achieved the highest value 0.940 in mAP, and has better performance than YOLO V13 with 0.893, YOLO V8 with 0.923, YOLO V12 with 0.930 and YOLO V11 with 0.931.

Acknowledgments

The authors are grateful to the anonymous reviewers and the helpful suggestion given by the partners. The research was supported by the Guangdong Philosophy and Social Sciences Planning Project [No.GD24CJY21], the Guangdong Provincial Natural Science Foundation [No.2026A1515011947], the Natural Science Foundation of the Jiangsu Higher Education Institutions of China [NO.25KJB416005], the Education Model Innovation Team[No.LT2404] and "Computer Theory Teaching and Research Office" 2025 Guangdong Province Undergraduate University Teaching Quality and Teaching Reform Project Construction Project [Yue Jiao Gao Han [226] No. 4].

References

[1] Rafique M A, Khaskheli F, Hassan M T, et al. Employing automatic content recognition for teaching methodology analysis in classroom videos[J]. Plos one, 2022, 17(2): e0263448.
[2] Chen H, Guan J. Teacher–student behavior recognition in classroom teaching based on improved YOLO-v4 and Internet of Things technology[J]. Electronics, 2022, 11(23): 3998.
[3] ZHAO J, XU M L, WANG X Z. A novel dataset based on indoor teacher-student interactive mode using AIoT[J]. Internet of Things,2024(25):101044.
[4] Li S, Ding Z, Chen H. A Neural Network-Based Teaching Style Analysis Model[C]//2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE, 2019, 2: 154-157.
[5] Nida N, Yousaf M H, Irtaza A, et al. Instructor activity recognition through deep spatiotemporal features and feedforward extreme learning machines [J]. Mathematical Problems in Engineering, 2019.
[6] WU D L, JIA C, WEI D, et al. The recognition of teacher behavior based on multimodal information fusion [J]. Mathematical Problems in Engineering, 2020, 10(13): 1-8.
[7] PENG Z L, YANG Z D, XIA J B, et al. Recognizing teachers hand gestures for effective non - verbal interaction [J]. Applied Sciences, 2022,12(22):11717.
[8] Li JN, Li RY et al. Recognition and Analysis of Teaching Behavior Based on Multi-scale GCN[J]. Computer science, 2024, 51(10):135-143.
[9] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection [EB/OL]. [2020]. https://github.com/AlexeyAB/darknet.
[10] Ultralytics. YOLO V8 [EB/OL]. [2023-02-03]. https://github.com/ultralytics/ultralytics.
[11] Jing HW, Xu J. Analysis of Classroom Behaviors of International Chinese Language Teachers Based on Deep Learning [J]. Computer technology and development, 2025, 35(12):58-66.
[12] Khanam R, Hussain M. YOLOv11: An Overview of the Key Architectural Enhancements [EB/OK]. [2024-10-23]. https://www.arxiv.org/pdf/2410.17725.
[13] Tian Y, Ye Q, Doermann D. YOLOv12: Attention-Centric Real-Time Object Detectors [EB/OK]. [2025-02-18]. https://arxiv.org/pdf/2502.12524.
[14] Lei M, Li S, Wu Y, et al. YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception [EB/OL]. arXiv preprint arXiv:2506.17733, 2025.
[15] Fan Y. SCB-Dataset: A Dataset for Detecting Student Classroom Behavior [EB/OL]. [2025]. https://arxiv.org/html/2304.02488v5#S4.