Electronic Journal of Business and Management
ICDBSE Special Issue, 2025 pp. 55-66
e-ISSN:2550-1380-ISSN

Facilitating Intelligent Evaluation and Analysis of Teachers' Teaching Behaviors Based on Improved Yolov8 Algorithm

Liu Lihua
Jiangsu Open University, Information Construction Department
1511270375@qq.com

Li Fengxia
Jiangsu Open University, Information Construction Department
lifx@jsou.edu.cn

Feng Yujia
Jiangsu Open University, Information Construction Department
fengyj@jsou.edu.cn

Zhang Wei
Jiangsu Open University, Information Construction Department
zhangw@jsou.edu.cn

Abstract

Teachers' classroom behavior analysis is a crucial component of teaching analysis and evaluation. With the deep integration of Artificial Intelligence technology and the field of education and teaching, this paper investigates the application of object detection techniques to delve into teacher behavior data, optimize the traditional mode of teaching analysis, and ensure teaching evaluations are more objective and fairer. Integrating the ResT into YOLOv8 algorithm, ResT is an efficient multi-scale visual Transformer structure, which is based on the design structure of ResNet to capture the feature information of the object on different scales in a phased and fine-grained manner, and models the global dependence of the image relationship to enhance the feature extraction ability of the model on the target object, while reducing the model's attention to the background information of the image to achieve high-precision teachers' behavior detection for Remote Education. Extensive experiments demonstrate that, relative to the benchmark model YOLOv8, the algorithm presented in this paper achieves 93.8%, 92.8%, 95.3%, and 85.3% in Precision, Recall, mAP50, and mAP50-95 metrics respectively on the self-constructed teachers' classroom behavior dataset, which is an improvement of 3.5%, 2.9%, 3.1%, and 2.7%, respectively.

Keywords: *Smart Talk, Teaching Behavior, Deep Learning, Object Detection, Transformer*

1.0 Introduction

In July 2017, China's State Council issued the New Generation of Artificial Intelligence Development Plan, which explicitly utilized intelligent technology to accelerate the reform of talent cultivation mode and teaching methods. As a result, the innovative integration of

Artificial Intelligence technology and education has become one of the popular research fields. From supporting the rapid storage and transportation of learning resources and realizing the intelligent management of student information in the era of computational intelligence; to driving the intelligence of teaching and learning with Deep Learning methods based on big data in the stage of perceptual intelligence; to establishing learner-centered customized education in the era of cognitive intelligence (Wu & Liu, 2017). With the deep integration of Artificial intelligence technology and education and teaching, the digital education literacy and skills of students and teachers are improving, providing a theoretical basis and technical support for the realization of a more equitable, effective and humanized education system.

With the popularity of smart classroom, more and more students' and teachers' behavioral data were captured, and this type of massive data laid a good foundation for introducing Artificial Intelligence technology to promote the digitalization of education informatization. Teachers' classroom behavior analysis is a very important part of teaching analysis and evaluation, and common teachers' classroom behaviors include backboarding, pointing to the blackboard or classroom materials, interacting with students, and so on. The traditional teachers' classroom behavior analysis was based on manual evaluation, which is not only time-consuming and costly, but also more subjective in its conclusions. Therefore, researchers have gradually explored the utilize of computer vision technology to optimize the traditional mode of teaching analysis, to mine and analyze the teachers' behavioral data in the smart classroom in a more targeted manner, so as to make the method of teaching evaluation reach automation, real-time, and objectivity (Cheng & Song, 2021; Zhang & Wen, 2022). Thereby this will help teachers to understand their own teaching situation, reflect on and adjust their teaching methods in a timely manner, and promote the improvement of teaching quality.

However, there are also many challenges in the analysis of intelligent teacher classroom behavior. First, the collection of datasets, more research focuses on the analysis of students' behavior in the smart classroom, relatively few studies for the detection of teachers' classroom behavior, and there is no corresponding publicly available benchmark datasets, and there are multiple challenges such as illumination, deformation, complex backgrounds, scaling, etc. when collecting datasets; second, limited by the adaptive ability of the existing object detection technology in the face of real smart classroom scenarios and the speed, Therefore, how to effectively improve the accuracy of the teachers' classroom behavior detection on the basis of satisfying the real-time detection is the focus of this study.

To address the above two challenges, the main contributions of this paper are as follows:

(1) Teaching video data from the demonstration classroom display of ideological and political theory courses in Chinese colleges were selected as the main research object, and the teachers' behavior detection dataset was constructed on its own, which contains the following seven categories such as board writing, guiding, looking around, one-handed gesture, open two-handed gesture, conservative two-handed gesture, and no obvious activities, to provide data support for the teachers' teaching analysis and evaluation.

(2) Taking YOLOv8 as the benchmark model, the ResT model is used to replace the original convolutional block structure in the feature extraction stage, which enhances the long time dependency of the model, and at the same time improves the model's ability to extract

features from the input image, and also reduces the computational amount of the model effectively.

(3) The teachers' teaching analysis detection algorithm proposed in this paper has outstanding performance on the self-constructed dataset, with several indexes substantially better than the benchmark model, realizing the accurate identification of various types of teacher behaviors in the whole process of teaching.

2.0 Literature Review

In this section, we will summarize the methods for analyzing classroom teaching behavior and introduce the YOLOv8 algorithm.

2.1 Methods of Analyzing Teachers' Teaching Behavior in Classroom Scenario

The traditional method of analyzing teachers' classroom teaching behavior is to organize experts, students and teachers to rate teachers' performance in the form of rating scales, and then manually organize, summarize and analyze the data manually. This way of assessing and analyzing teachers' classroom teaching behaviors based on relevant evaluation scales consumes a lot of human, material, and financial resources, and cannot guarantee the objectivity of the evaluation results, and the representative methods are Flanders Interaction Analysis System (FIAS) (Flanders, 1963), Teacher-Student (S-T) Behavioral Analysis Method (Fu & Zhang, 2011) and so on.

In addition to the above defects, the traditional method of analyzing teachers' classroom teaching behavior has no systematic evaluation standard and system due to the different evaluation subjects each time, and it is not possible to evaluate the process of the same teacher's teaching performance in the long term. At the same time, this kind of method is also limited by the workload and can not be promoted on a large scale. Therefore, the utilize of new technology to help teaching behavior analysis is imminent.

With the rapid development of Artificial Intelligence technology, the teachers' teaching behavior analysis methods have gradually become intelligent. Initially, researchers developed and applied classroom teaching behavior analysis software to automate the analysis of the teachers' teaching behavioral data, representative methods include Information Technology-Based Interaction Analysis System (ITIAS) (Gu & Wang, 2004), Classroom Teaching Behavior Analysis System (TBAS) (Mu & Zuo, 2015) and so on, but such systems still need to collect and process teachers' classroom behavior data manually. With the popularization of smart classroom and Deep Learning technologies, researchers have gradually introduced Computer Vision related technologies to realize fully automated teachers' teaching analysis.

Existing teachers' classroom behavior analysis is dominated by techniques such as image recognition, human key point detection, and object detection, focusing on a series of analyses of the teacher's face, head, gestures, and behavioral movements. Zheng (2021) assessed the key point information of teachers' nose, left hand and right hand based on HRNet and constructed teachers teaching behavior evaluation metrics based on it, aiming at intelligent analysis and automated evaluation. Pang (2022) used the Openpose model to obtain the information of teachers' skeleton data for the purpose of classifying teachers' non-verbal behaviors. Peng (2022) introduced the object detection algorithm YOLOv3 into the

classroom to recognize teachers' gesture movements, classifying gestures into three categories: symbolic gestures, descriptive gestures, and emotional gestures, which helps to subsequently correlate teachers' verbal thinking and cognition. Jiang (2023) implemented quantitative computation of teachers' beat gesture recognition and constructed a gesture dataset, and the success rate of detection in real classroom environments can reach 95.6%. These methods of deep integration with Deep Learning technology realize the automation, scale and normalization of the analysis of teachers' classroom teaching behaviors, and also provide strong support for the optimization of teachers' teaching quality and strategies in the future.

2.2 The YOLOv8 Algorithm

The YOLOv8 algorithm is the eighth generation of the YOLO series of improved algorithms, which contains image feature extraction, feature sampling and fusion, and prediction module (Varghese & Sambath, 2024). Holistically speaking, the YOLOv8 algorithm has four main innovations. Firstly, it replaces the C3 module in YOLOv5 (Jocher & Chaurasia, 2022) with a new C2f module in the network structure, and the enriched gradient flow makes the target features extracted by the model more adequate. Secondly, in the prediction module, YOLOv8 is designed with a decoupled head structure, i.e., object category prediction is separated from object position prediction, and two independent convolutional layers are utilized to predict the target category and positional coordinates respectively. Thirdly, the algorithm abandons the traditional Anchor-Based design and utilizes the Anchor-Free prediction strategy, which does not require the fine design of the anchor hyper-parameters (e.g., the anchor's size, scale, etc.), and also enhances the model's regression ability for predicting objects of different sizes. Finally, in terms of the loss computation strategy, it discards the original box-matching strategy, and adopts the Task Aligned Assigner strategy, which makes the algorithm more demanding and strict in selecting positive and negative samples, and also, in the calculation of the loss of object margin deviation, it utilizes the Distribution Focal Loss(Li & Wang, 2020), which is calculated by weighting the samples of each category, so that the model can better learn the features of different categories of samples, effectively solving the category imbalance problem in the training process.

Although YOLOv8 has a certain magnitude of improvement in detection accuracy and speed compared with the previous generations of algorithms, it still has some obvious shortcomings. In the image feature extraction stage, YOLOv8 adopts stacked CBS and C2f modules to extract image features at different scales, but it is difficult to complete the fine-grained feature extraction of the target in complex scenes, which leads to the subsequent false and missed detections.

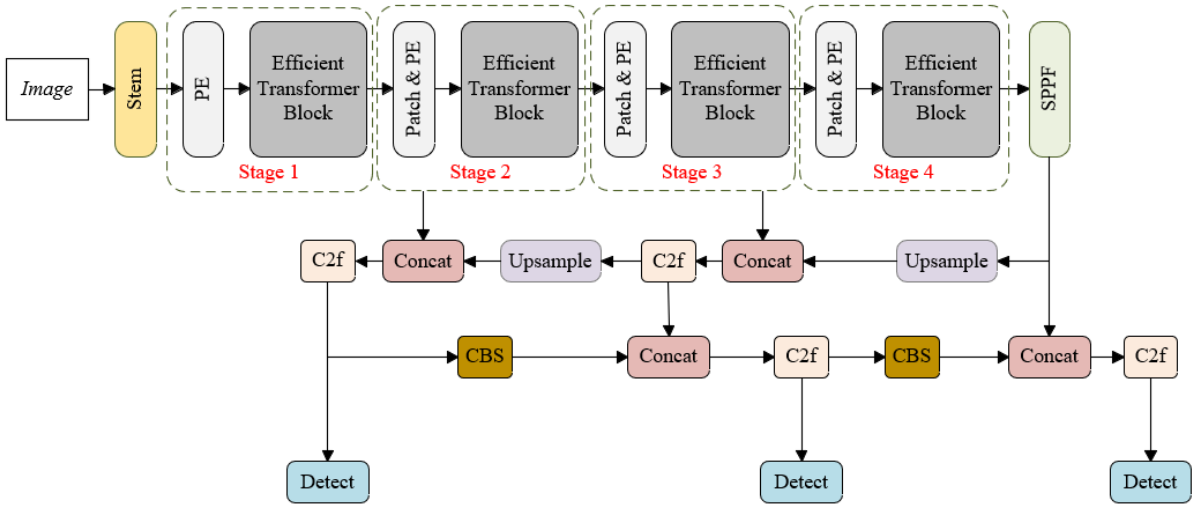
3.0 The Proposed Algorithm

Since the YOLOv8 algorithm contains models with different sizes of n, s, m, l, and x, for the actual smart classroom application scenarios and to measure the comprehensive performance and detection speed, the YOLOv8s model achieves a better balance between detection accuracy and speed, and thus YOLOv8s is selected as the benchmark detection model in this paper.

This paper proposes an improved YOLOv8s model incorporating the ResT (ResNet-like Vision Transformer) model and verifies that it can be successfully applied to a real smart classroom to achieve the task of recognizing teachers' teaching behaviors.

In the image feature extraction stage, in order to enhance the feature extraction and expression ability of the model on the input image, the ResT model is used to replace the original CBS and C2f modules, and the Transformer model is used to model the global dependency information of the image, to better extract the multi-scale feature expression of the image, and to complete the model's fine-grained detection of complex images. Figure 1 shows the overall structure of the proposed model in this paper.

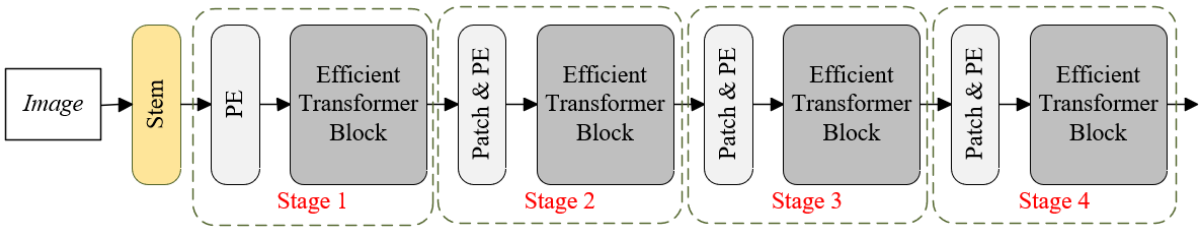
Figure 1: Structure of the Algorithm



3.1 ResT

The overall structure of ResT model is similar to that of ResNet. Firstly, the input image goes through a stem module to extract the low-dimensional features, and then goes through four sub-stage modules to obtain the feature maps under different sizes. The first Stage contains a Patch Embedding and several Efficient Transformer Blocks; in the subsequent three Stages, they contain a Patch Embedding, a Position Encoding layer and several Efficient Transformer Blocks. These modules are similar to the structure of ViT (Dosovitskiy & Beyer, 2020), the difference is that the ResT (Zhang & Yang, 2021) model proposes Efficient Transformer Block to reduce the amount of computation, the overall structure of the module is shown in Figure 2.

Figure 2: Structure of the ResT model



The input Token, i.e., X, first goes through a linear layer to get the query, i.e., Q in Figure2, and at the same time, in order to reduce the amount of computation, X will be transformed from 2-dimensional to 3-dimensional, and after that, it is fed into a depth-separable convolution and layer normalization for compression of the features, and then transformed from the 3-dimensional feature map to the 2-dimensional feature map, and at the same time, it is fed respectively into two independent linear layers to extract the features to get the key and value, i.e., K and V; and then through the following formula (1) to calculate the attention to get the result M. Finally a linear layer and a residual connection are utilized to get the final feature F. In the MSA, the input tokens are divided into n parts at the same time, and the final result will be the n parts of the results are connected. Compared with the MSA module in ordinary Transformer, Efficient Transformer Block utilizes depth-separable convolution to reduce the dimension and resolution of the input tokens, which reduces the computational complexity of the whole module significantly, and in addition, layer normalization is introduced in the module, which is used to enhance the information interaction of multi-heads. It is also due to these two major improvements that the ResT model exhibits stronger feature extraction capability than the Transformer model while significantly reducing the computational overhead.

$$\mathrm{EMSA}(Q,K,V) = \mathrm{LN}\left(\mathrm{Softmax}\left(\mathrm{Conv}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)\right)\right)V \quad (1)$$

Where EMSA is the Efficient Multi-Headed Self-Attention Module of Efficient Transformer Block, Conv is the standard 1×1 convolution operation. In order to model the ability of interactions between different Multi-Headed Attention Modules, EMSA adds an additional Softmax and layer normalization operation after the attention computation to statistical features.

Furthermore, in the Position Encoding module, the ResT model proposes a simple and efficient Pixel-wise Attention (PA) to encode positions. Specifically, PA employs a depth-separable convolution with a convolution kernel of 3×3 to obtain pixel-level weights, which are then activated using Sigmoid. The overall steps of location encoding obtained using PA are shown in equation (2) below:

$$x^{a} = \mathrm{PA}(x) = x * \sigma(\mathrm{DWConv}(x)) \quad (2)$$

where x is the input feature, x^a is the output feature, PA is the pixel-level attention, σ is the Sigmoid function, and DWConv is the convolution kernel is a 3×3 depth-separable convolution.

The whole feature extraction module of YOLOv8s utilizes stacked convolutional blocks to extract the feature, which is limited by the disadvantage of the small Receptive Field, which makes its extracted feature maps less correlated and the local features of the object are not described finely enough. In order to solve this problem, the method in this paper proposes to utilize the ResT model instead of the original convolutional block and C2f module in the image feature extraction stage of the YOLOv8s. On the one hand, the long time dependency of the image is modeled by the Transformer, which makes the feature correlation between the different sizes of the feature maps extracted by the model closer and more complementary,

and on the other hand, the Efficient Transformer Block proposed by ResT reduces the resolution and dimension of the input feature maps by introducing depth-separable convolution, so that the computation of the overall model does not increase significantly compared to the convolution block, which also ensures the speed advantage of the model when actually deployed. Meanwhile, for the task of detecting teachers' classroom behaviors under the smart classroom studied in this paper, the ResT model can better capture the fine-grained information of the video frames by using its efficient feature extraction capability, and learn the fine-grained features of teachers' behavioral actions based on the long time dependency.

4.0 The Teachers' Teaching Behaviors Dataset

There are more public and category-rich behavior detection and recognition datasets, but due to the specificity of the classroom teaching scene, such as walking, jumping and other behavioral categories rarely appear in the classroom, taking into account the teachers' classroom behavior needs to make some sense. Therefore, this paper constructs a teachers' teaching behavior detection dataset in the following four steps: data collection, data cleaning, manual labeling, and format conversion.

Data collection: Through the research and analysis of existing literature related to teachers' classroom teaching behaviors, this paper mainly selects teachers' teaching video data from public platforms such as the Ideological and Political Theory Course Demonstration Classroom Showcase in Chinese Colleges and Universities, and classifies teachers' behaviors under the smart classroom into seven categories, which are board writing, guiding, ringing, single gesture, open two-handed gesture, conservative two-handed gesture, and no obvious activities.

Data cleaning: Pre-processing of teachers' classroom teaching behavior data obtained from public platforms. Firstly, the video is randomly sampled images through the frequency of one frame per second to ensure the diversity of training data. Secondly, the sampled video frame dataset is cleaned to eliminate the invalid data with blurring, action ambiguity, etc. Finally, the teachers' classroom teaching behavior detection dataset is obtained, which contains 3495 training images, 891 validation images, and 1500 test images.

Manual labeling: The LabelImg tool is used to implement the manual annotation of the training dataset for the seven categories, noting board writing, guiding, ringing, single gesture, open two-handed gesture, conservative two-handed gesture, and no obvious activities as: 0, 1, 2, 3, 4, 5, and 6, respectively, and obtaining the standard xml-formatted annotation files, which contains information about the categories of the teacher's behaviors and the coordinates of the location where teacher is located in the images.

Format conversion: This paper converts the standard xml-formatted labels into txt-formatted files, i.e., the format of the file first is the object category index (0 for board writing, 1 for guiding, 2 for ringing, 3 for the single gesture, 4 for open two-handed gesture, 5 for the conservative two-handed gesture, and 6 for the no obvious activity), and then is the horizontal coordinates of the normalized object center point, the vertical coordinates of the normalized object center point, the width of the normalized object center point, and the height of the normalized object center point.

After completing the above steps, the self-constructed teachers' teaching behavior detection dataset is used for the training and validation of the proposed algorithm.

5.0 Experimental results

In this section, we provide a detailed introduction to the parameter settings and results of our experiment.

5.1 Hyper-Parameter Setting

In the training stage, the algorithm proposed by this paper is consistent with the YOLOv8s model and adopts the stochastic gradient descent algorithm, with the learning rate, momentum, weight decay parameter, batch, and the number of training epoch set to 0.01, 0.937, 0.0005, 64, and 100, respectively. In the aspect of data enhancement, the probabilities of image hue, saturation, and value are set to 0.015, 0.7, and 0.4, respectively. The image scale scaling factor, image left-right flip probability, and Mosaic probability are set to 0.5, 0.5, and 1.0, respectively, and Mosaic method is turned off in the last ten epochs. Meanwhile, in this paper, the cross-entropy loss is used to calculate the object category loss, and the weight is set to 0.5, and the DFL and CIoU Loss are used to calculate the box deviation, and the weights are set to 1.5 and 7.5.

5.2 Experimental results

Figure 3 shows the metrics results of YOLOv8s on the MS COCO2017 validation set reproduced in the paper. MS COCO2017 is a challenging object detection benchmark dataset containing 80 classes of common objects such as humans, animals, furniture, electronics, vehicles, etc., and the validation set contains 5000 images. As can be seen in Figure 8, YOLOv8s achieves 67.3%, 54.7%, 59.9%, and 43.9% for Precision, Recall, mAP50, and mAP50-95 metrics on the MS COCO2017 validation set, respectively.

Figure 3: Plot of metrics results for YOLOv8s model on the MS COCO2017 validation set

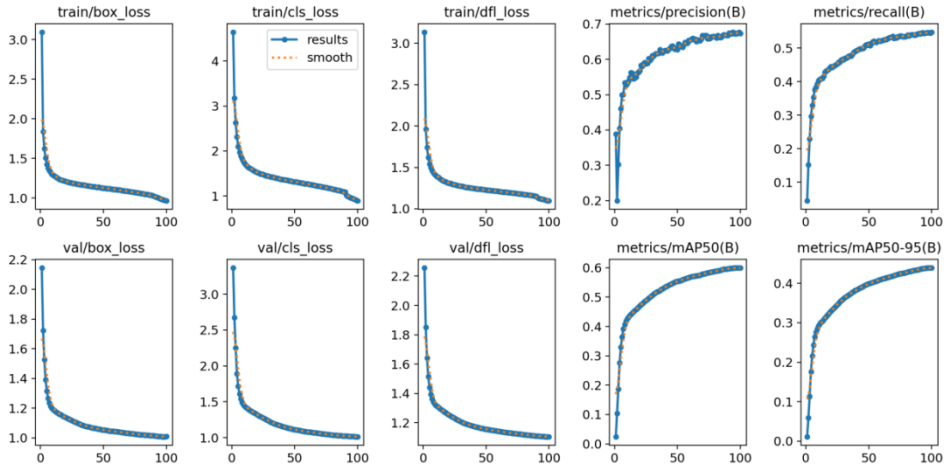


Figure 4 shows the results of the metrics of this proposed algorithm in the paper on the MS COCO2017 validation set. The metrics of the proposed algorithm in Precision, Recall,

mAP50, and mAP50-95 are 69.1%, 55.9%, 61.5%, and 44.9%, respectively, and the above metrics are improved compared to the benchmark YOLOv8s, which is 1.8%, 1.2%, 1.6%, and 1.0%, respectively. The overall performance is outstanding, thanks to the fact that the ResT model used in the feature extraction stage has a more powerful feature extraction capability compared to the original convolutional block, and the multi-scale feature extraction strategy allows the algorithm to better extract the features of the target in different scenes and sizes.

Figure 4: Plot of the metrics results of the model proposed in this paper on the MS COCO2017 validation set

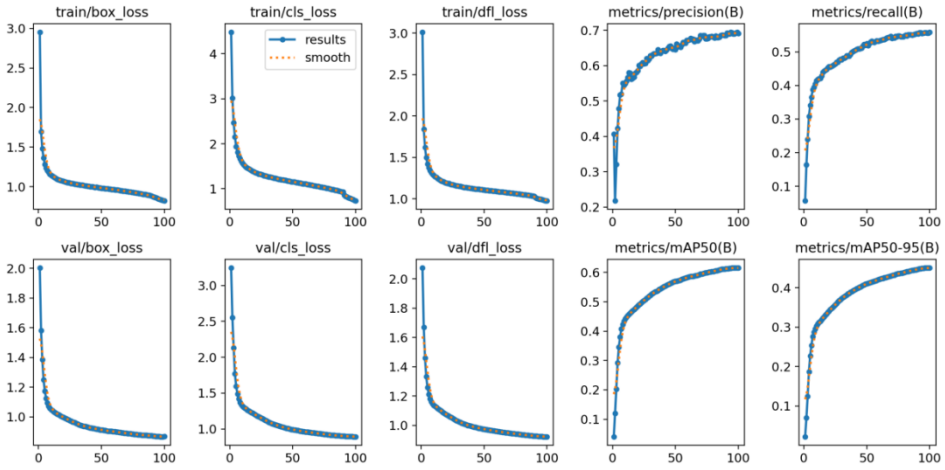


Figure 5 shows the result metrics based on the YOLOv8s model on the self-constructed teachers' teaching behavior detection dataset in this paper. As can be seen from the figure, with the increase of the number of training rounds, the loss of the model on the training set and the loss on the validation set both decrease synchronously, and finally the YOLOv8s is able to converge to 0.494, 0.433, and 0.928 for the box_loss, cls_loss and dfl_loss, respectively, on the validation set. The YOLOv8s is able to achieve 90.3%, 89.9%, 92.4%, and 82.6% of Precision, Recall, mAP50, and mAP50-95 metrics on the validation set of the self-constructed teachers' teaching behavior detection dataset respectively.

Figure 5: Plot of metrics results for YOLOv8s model on the validation set of the self-built dataset

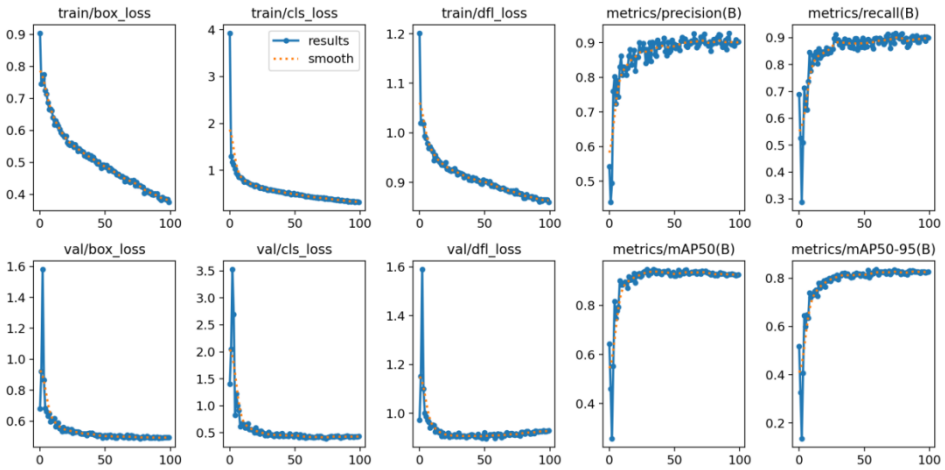


Figure 6 shows the metrics of the proposed algorithm in the paper on the validation set of the self-constructed teachers' teaching behavior detection dataset. As can be seen from the figure, compared to the benchmark model YOLOv8s, the metrics Precision, Recall, mAP50, and mAP50-95 are improved by 3.5%, 2.9%, 3.1%, and 2.7%, to 93.8%, 92.8%, 95.3%, and 85.3%, respectively.The finer-grained target feature extraction brought by the ResT leads to a decrease in the target loss, and the improvement in the performance metrics is largely attributed to the ability of the ResT model to capture both local and global features of the target object. The experimental results verify that the algorithm in this paper is effective for teacher behavior recognition and can meet the task requirements of teacher behavior detection for normal whole process teaching.

Figure 6: Plot of metrics results of the model proposed in this paper on the validation set of the self-constructed dataset

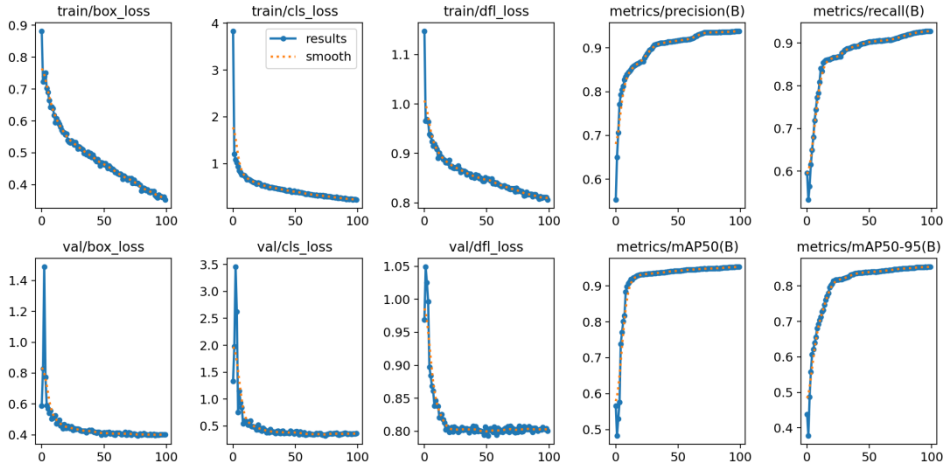


Table 1 shows the metrics comparison of multiple models on the validation set of the self-constructed dataset in the paper. It can be seen that the model proposed in this paper shows significant advantages. Specifically, the precision, recall, mAP50, and mAP50-95 metrics of the YOLOv8s model are 90.3%, 89.9%, 92.4%, and 82.6%, respectively; the YOLOv8+ResT, i.e., the variant model obtained by replacing the convolutional blocks in the original YOLOv8s with the ResT in the feature extraction stage of the model, shows significant advantages in precision, Recall, mAP50, and mAP50-95 metrics, which can reach 93.8%, 92.8%, 95.3%, and 85.3%, respectively. ResT utilizes the Transformer to model the global dependency information of the image, and better extracts the multi-scale feature representation of the image, so that it can ultimately achieve a better result.

Table 1: Comparison of metrics of different variant models on the validation set of the self-built dataset

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLOv8s	90.3	89.9	92.4	82.6
YOLOv8s+ResT	93.8	92.8	95.3	85.3

The finer-grained target feature extraction brought about by the ResT results in a decrease in target loss, and the improvement in performance metrics is largely attributed to the ability of the ResT to capture both local and global features of the target. The experimental results verify that the proposed algorithm in the paper has good results for teachers' teaching behavior detection.

6.0 Conclusion

In this paper, taking the single-stage object detection model YOLOv8s as the benchmark, and targeting the problem of insufficient long-time dependency brought by CNN due to the limited and fixed receptive field, we propose to use the ResT model to replace the CBS and C2f modules in YOLOv8s, which effectively enhances the model's ability of feature extraction for targets in the stage of image feature extraction. Facing the complex scenario of teachers' classroom behavior detection, which requires long time dependence, the proposed algorithm of this paper improves several metrics in the self-constructed teachers' classroom behavior detection dataset significantly compared with the YOLOv8s, which can provide a strong support for the subsequent teaching analysis and evaluation.

7.0 Funding

This research was supported by the 2022 Jiangsu Higher Education Ideological and Political Course Education and Teaching Reform and Innovation Demonstration Project Digital Evaluation Reform and Innovation of Ideological and Political Education and Teaching in Higher Education (Project Number: 12), and the 2023 Jiangsu Provincial Education Science Planning Special Project Research on Lifelong Digital Learning Adaptability and Learning Preferences for the Construction of a Learning Society in Jiangsu Province (C/2023/01/62).

8.0 References

Cheng, X., Song, C., Shi, J. G., et al, 2021. A Survey of Generic Object Detection Methods Based on Deep Learning. Acta Electronica Sinica, 49(07), 1428-1438.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al, 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Flanders, N. A., 1963. Intent, Action and Feedback: A Preparation for Teaching. Journal of Teacher Education,14(3), 251-260.

Fu, D. R., Zhang, H. M., 2011. Educational Information Processing[M].Beijing Normal University Publishing House.

Gu, X. Q., Wang, W., 2004. New Exploration of Classroom Analysis Technology to Support Teachers' Professional Development. China Education Technology, (7), 18-21.

Jiang, S. N., 2023. Intelligent Recognition Method for Teacher's Beat Gesture Based on Yolov5. Fujian Computer, 39(09), 8-13.

Jocher, G., Chaurasia, A., Stoken, A., et al, 2022. ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations.

Li, X., Wang, W., Wu, L., et al, 2020. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. Advances in Neural Information Processing Systems, 33, 21002-21012.

Mu, S., Zuo, P. P., 2015. Research on the Analysis Method of Classroom Teaching Behavior in the Information Technology Teaching Environment. E-education Research, 36 (09), 62-69.

Pang, S. Y., Zhang, A. R., Lai S. H., et al, 2022. Automatic Recognition of Teachers' Nonverbal Behavior Based on Dilated Convolution. IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), 162-167.

Peng, Z. L., Yang, Z. D., Xiahou, J. B., et al, 2022. Recognizing Teachers' Hand Gestures for Effective Non Verbal Interaction. Applied Sciences, 12(22), 11717.

Varghese, R., Sambath, M., 2024. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS).

Wu, Y. B., Liu, B. W., Ma, X. L., 2017. Building An Ecosystem of "Artificial Intelligence + Education".Journal of Distance Education, 35(5), 13.

Zhang, Y., Wen, G. Z., Mi, S. Y., et al, 2022. Overview on 2D Human Pose Estimation Based on Deep Learning. Journal of Software, 33(11), 4173−4191.

Zhang, Q., Yang, Y. B., 2021 . Rest: An efficient transformer for visual recognition. Advances in neural information processing systems, 34, 15475-15485.

Zheng, Y. H., 2021. An Evaluation Method for Teaching Behaviors based on Posture Recognition Algorithm. Software Engineering, 24(04), 6-9.