Proceedings of The 6th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2026.31570

Research on the Innovation and Development of Ideological and Political Education Based on Artificial Intelligence Technology

Siqi Wang

School of Marxism, Wanjiang University of Technology, Maanshan, China
wang18895556001@163.com

Abstract. In education digital transformation, classroom behavior analysis is core of educational evaluation innovation; traditional teaching lack comprehensive learning feedback, deep learning provide new solutions. This paper address YOLO's insufficient accuracy, efficiency and real-time performance in complex classroom scenarios to support accurate monitoring system construction. Due to dense/small targets and occlusion causing low accuracy, accuracy-efficiency imbalance, and real-time shortage leading to missed/false judgments, this paper propose three algorithms: MRD-YOLO (high precision, MFFN module) improves mAP50/mAP50-95 by 3.2%/6.1% on POCO and 3.2%/4.8% on SCB-Dataset3 vs YOLOv8n; DAM-YOLO (lightweight, DGCST module) balances accuracy-efficiency, with slightly lower accuracy than MRD-YOLO but 1.1M fewer parameters, 2.3B less FLOPs and 2.2MB smaller size; FME-YOLO (lightweight/real-time, FEAM module) optimizes accuracy and lightweight vs the two, achieving 178.3/177.6fps; Web first loading <3s, single frame <1s, meeting real-time requirements.

Keywords: Artificial Intelligence, Deep Learning, and Behavior, Recognition in Ideological and Political Education Classrooms, YOLO

1. Introduction

AI shifts education to "data-based"; students' classroom behavior affect teaching quality [1]. Traditional management: teachers lack timely learning situation, causing slow feedback. Deep learning behavior recognition: use computer vision to monitor classroom videos, detect features and feedback dynamically for teaching adjustment. Universities explore AI/big data in teaching; BNU and Xidian's projects applied successfully, showing AI's potential [2]. Policies promote education intelligence: Ministry of Education's plan requires "smart classrooms"; 2023 report shows AI improves teaching efficiency. AI in classrooms faces challenges like small target detection and occlusion, needing algorithm optimization.

DOI: 10.54254/2755-2721/2026.31570

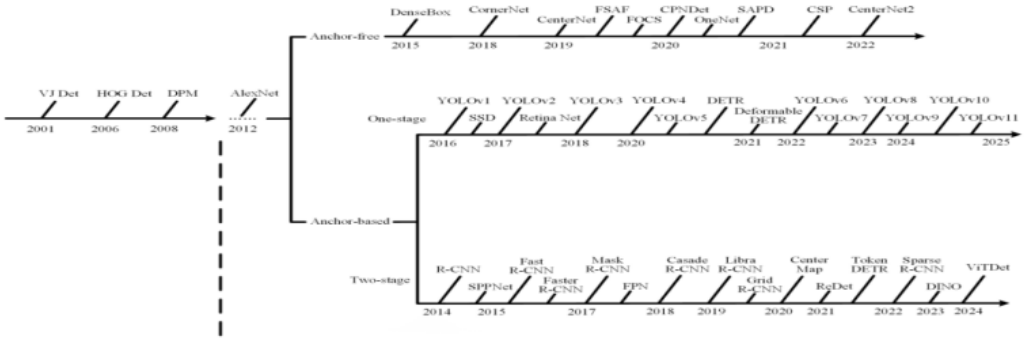


Figure 1. Evolution of target detection

2. Basic assumptions

Deep learning, key to ML, mimic human brain neural network to auto-learn big data features. Unlike traditional ML, it avoid manual extraction, using multi-layer networks for efficient image, speech and NLP task handling. Its rapid development drive by large datasets, GPUs and optimized algorithms [3]. Past decade, it advance significantly; DNN, CNN, RNN are key for complex problems. CNN, widely used in deep learning, excel in computer vision, auto-extracting features via convolution/pooling, crucial for classroom behavior recognition.

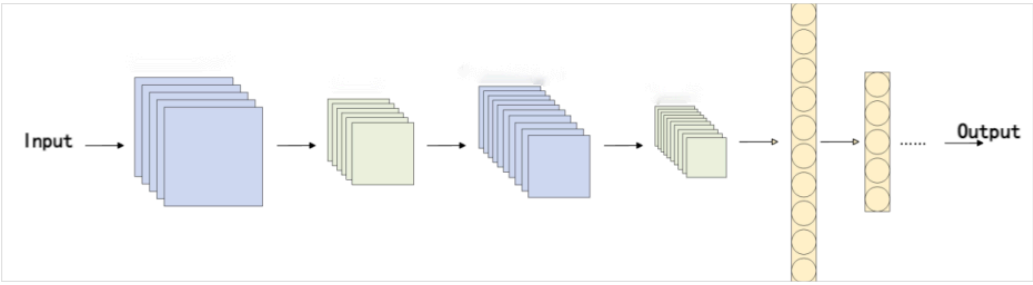


Figure 2. Hierarchical structure diagram of convolutional neural networks

$$Y(i,j) = \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} X(i+m, j+n) \cdot W(m,n) \tag{1}$$

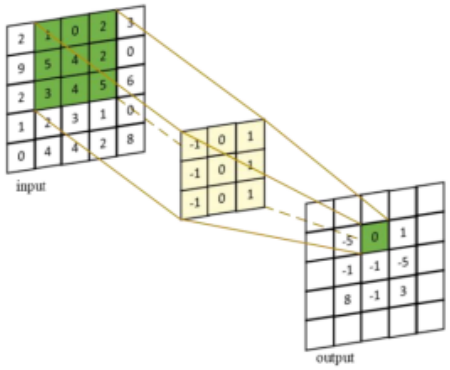


Figure 3. Convolutional computation process

DOI: 10.54254/2755-2721/2026.31570

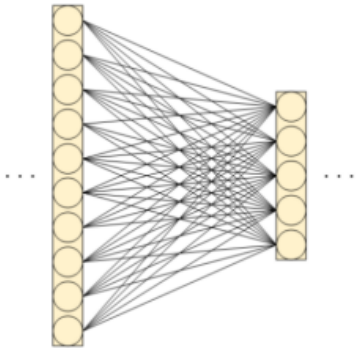


Figure 4. Schematic diagram of fully connected structure

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$Softmax(x) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{3}$$

$$f(x) = max(0, x) \tag{4}$$

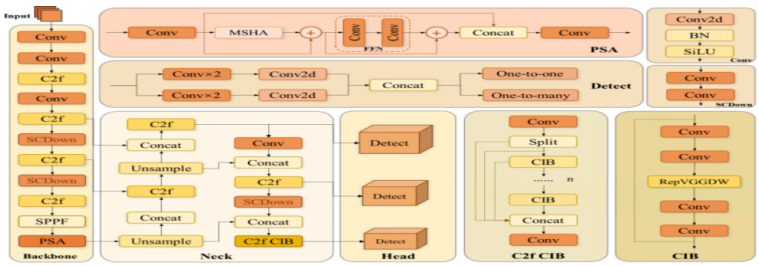


Figure 5. YOLOv10n architecture diagram

Further analysis find YOLOv10 show better accuracy and efficiency in classroom behavior recognition. Thus, we focus on YOLOv10n for detailed optimization [4]. This paper mainly show its research results, especially good performance in accuracy, speed and real-time.

$$P = \frac{TP}{TP+FP} \times 100\% \tag{5}$$

$$P = \frac{TP}{TP+FN} \times 100\% \tag{6}$$

$$AP = \int_0^1 P(R)dR \times 100\% \tag{7}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{8}$$

$$FPS = \frac{1000}{\text{Preprocess}+\text{Inference}+\text{Postprocess}} \tag{9}$$

DOI: 10.54254/2755-2721/2026.31570

3. Lightweight high precision real-time classroom behavior recognition algorithm based on YOLO10n

This chapter focus on YOLOv10n, optimizing design to boost accuracy, lightness and real-time for classroom behavior recognition task. Based on it, improved FME-YOLO propose with better performance, no consider complex scene adaptability [1]. Figure below show its structure; yellow dashed box indicate the improvements part.

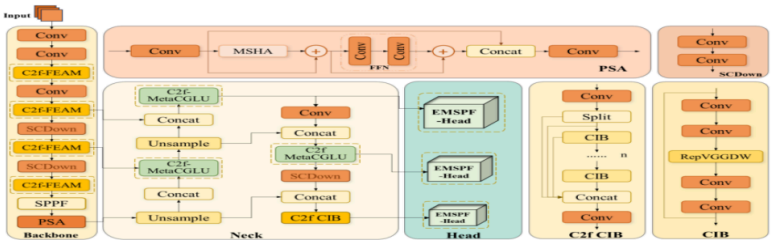
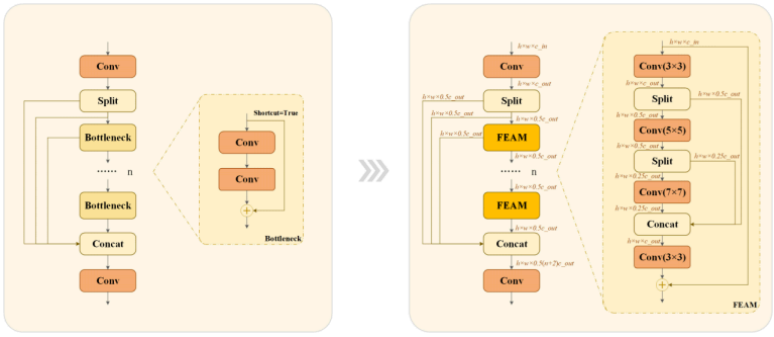


Figure 6. FME-YOLO structural diagram

In classroom scenes, because students have various postures and are densely distributed, the original feature extraction module's ability to adjust receptive field is limited, making it hard to effectively capture students' features of different postures and scales. To solve this problem, this chapter introduce FEAM module, replacing the Bottleneck of traditional C2f module in Backbone.

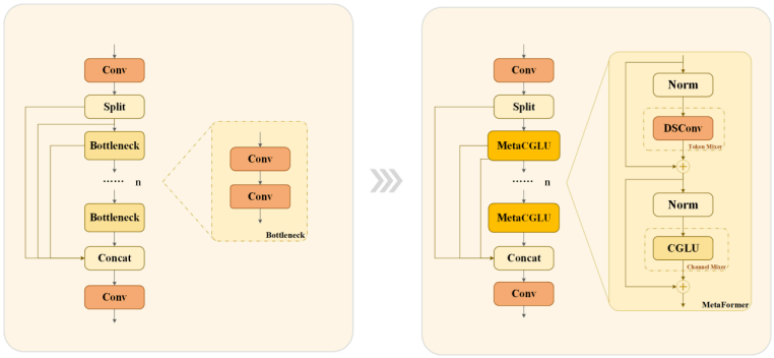


(a) Backbone's C2f module structure (b) C2f-FEAM module structure

Figure 7. Comparison before and after improvement

This chapter propose MetaCGLU, new feature fusion module replacing C2f's Bottleneck in Neck (structure Fig 5.3). Based on MetaFormer, it combine lightweight CGLU/DSConv, enhancing feature fusion depth/efficiency, improving detailed focus and semantic fusion, raising detection accuracy and computing efficiency.

DOI: 10.54254/2755-2721/2026.31570



(a) The C2f module structure of Neck (b)MetaCGLU module structure

Figure 8. Comparison before and after improvement

To evaluate the optimization degree of the proposed modules in this chapter and their different combination orders on algorithm performance, a series of ablation experiments are designed. As shown in Table 1, "√" mean that the optimization module is added based on the YOLOv10n network model.

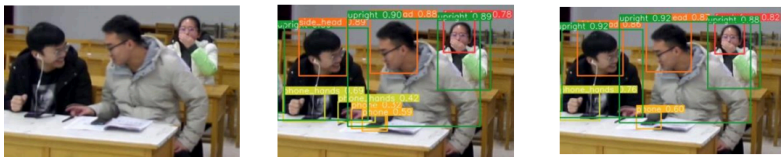
Table 1.

FEAM	Meta CGLU	EMSPF-Head	P/%	R/%	mAP50/%	mAP50-95/%	Params/M	GFLOPs/B	MB	FPS/fps
			88.0	82.4	88.5	62.3	2.3	6.5	5.8	165.7
√			89.0	83.3	89.6	63.6	2.0	6.0	5.4	181.6
	√		89.6	84.2	89.4	63.9	2.2	6.2	5.6	164.8
		√	90.8	83.9	91.3	65.4	1.9	6.0	4.3	169.3
	√	√	91.0	85.5	90.1	65.6	2.0	5.9	5.2	174.5
√	√	√	92.3	87.2	92.6	68.8	1.7	5.5	3.9	178.3

Results vs YOLOv10n: P/R/mAP50/mAP50-95 rise 1.0%-1.3%; params/computation/size down 0.3M/0.5B/0.4MB; FPS up 9.6% to 181.6 fps, verifying FEAM's role in boosting accuracy, speed and lightweight.

Actual detection performance

Real classroom image YOLOv10n detection FME-YOLO detection



(a) repeated detection correction

(b) Leak detection correction

DOI: 10.54254/2755-2721/2026.31570

(c) False Positive Correction

(d) Improved detection accuracy

Figure 9. Comparison of target detection

(a) YOLOv10n detection

(b) FME-YOLO detection

Figure 10. Comparison of detection performance for student-targeted dense detection

In sparse classrooms, YOLOv10n has repeated, missed and false detection with low accuracy, misjudge behaviors especially in inter-student occlusion and small targets (Fig. 9). FME-YOLO

DOI: 10.54254/2755-2721/2026.31570

reduce such errors, improve accuracy and handle simple scenes well. In dense complex scenes, YOLOv10n fail to capture details, causing serious missed detection of long-distance targets (Fig. 10). FME-YOLO improve significantly, recognize long-distance targets and details effectively, show superiority in handling dense students and background interference.

4. Conclusion

Under educational digital transformation, classroom behavior recognition is key to education reform and "data-driven" teaching. This paper optimize YOLO via deep learning for high-precision, low-cost real-time monitoring. Experiments: FME-YOLO outperform YOLOv10n (precision/recall up, parameters/computation down, FPS +7.6%); SCB-Dataset3 (mAP50/lightweight up, FPS +7.1%), verifying effectiveness. Gradio Web app on NVIDIA RTX 4050: first load <3s, single-frame <1s, meet real-time needs. FME-YOLO balance accuracy/lightweight, aiding behavior capture, learning feedback and teaching. Current research rely on video, need teacher's active check, lack automatic intervention. Future work enhance classroom behavior analysis diversification and intelligence.

References

[1] Shi D. Transnext: Robust foveal visual perception for vision transformers [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 17773- 17783.
[2] Xu W, Wan Y. ELA: Efficient local attention for deep convolutional neural networks [J]. arXiv preprint arXiv: 2403.01123, 2024.
[3] Wang W, Dai J, Chen Z, et al. Internimage: Exploring large-scale vision foundation models with deformable convolutions [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 14408-14419.
[4] Zhu X, Hu H, Lin S, et al. Deformable convnets v2: More deformable, better results [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9308-9316.