EAI Endorsed Transactions
on AI and Robotics
Review Article EAI.EU

Computer vision recognition in the teaching classroom: A Review

Hui Jiang[1], Wentao Fu[2,*]

[1]College of Foreign Languages, Northeast Forestry University, Harbin 150040, China
[2]College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China

Abstract

Artificial intelligence introduces computer vision recognition into the teaching classroom, and computer vision recognition technology lays a solid foundation for the intelligent teaching classroom. Through the classroom camera video stream to the classroom student information data collection, voice, posture, facial, physiological signal data recognition analysis processing to extract and define the characteristics of student behaviour, automatic classification behaviour and then record and display student behaviour, thus effectively help teachers to grasp the students learning state and emotions, to promote the quality of teaching has far-reaching significance. At the same time, the challenges and problems of the effective application of computer vision in the teaching classroom and the corresponding solutions are discussed.

Keywords: Artificial Intelligence, Deep Learning, Computer Vision, Teaching classroom behaviour recognition

Received on 05 October 2023, accepted on 12 December 2023, published on 08 January 2024

doi: 10.4108/airo.4079

*Corresponding author. Email: micoc@foxmail.com

1. Introduction

Advances in AI technology have enabled our machines to no longer rely solely on human intelligence, but to have more perceptive capabilities, more accurate speech recognition, faster decision-making, and more powerful language processing [1]. Deep learning, machine vision, natural language processing, and data mining are some of the key methods on which artificial intelligence technologies are based. Classroom action behavior analysis is gradually emerging as a hot topic of scientific inquiry with the growth and popularity of computer information technology in teaching and learning reform. These methods are used to create algorithms and models that let computers discover patterns in data, learn from them, and make judgments [2-4].

The blending of the Internet and education has sped up the development of smart teaching, combining deep learning and computer vision technology, using video and audio to analyze and analyze classroom information to automatically detect and identify students' classroom behavior and evaluate the analysis, which is crucial to help teachers accurately grasp students' learning status and timely adjust teaching methods and support the improvement of teaching.

This paper reviews deep learning-based target detection methods and analyses their advantages and disadvantages when applied to behavioural pattern recognition in the teaching classroom, and integrates and introduces the four feature extraction and processing methods based on attitudinal data, face data, voice data and physiological signals, as well as the various types of behavioural feature classification methods, which are the mainstays of behavioural pattern recognition in the teaching classroom at present. Meanwhile, the challenges and problems faced by the effective application of computer vision in teaching classroom and the corresponding solutions are discussed.

2. Classroom Behavior and Recognition

The acts and demeanor of students and teachers in the classroom are referred to as classroom behavior. It encompasses all audible and visible actions including listening, respect, collaboration, involvement, and conformity to norms and laws. A healthy learning environment and the promotion of efficient teaching and learning depend on effective classroom behavior. The method teachers educate and the degree of student focus in the classroom have the greatest immediate influence on how well students are taught and learned, despite the fact that a range of circumstances can affect this. Teachers often use a variety of strategies and techniques to promote positive classroom behavior, such as establishing clear expectations, providing positive reinforcement, giving corrective feedback, and modeling appropriate behavior [5-6].

The efficiency of teaching and learning is greatly influenced by the behavior of the students in the classroom. The meaning of teaching is to make students gain something, and whether students gain something or not is an important criterion to assess the effectiveness of teaching [7-8]. Even the best instructional design cannot achieve good teaching effectiveness if students are not focused on the lesson. There is a deeper connection between a teacher's instructional approach, the behaviors of the students in the classroom, and the overall teaching outcome. while teachers' efficient teaching style is often able to effectively guide students' classroom behaviors and produce better learning results, while teachers' good teaching effect is also a manifestation of an excellent teaching style [9]. With the aid of artificial intelligence technology, the video and voice data on the teaching screen in the classroom is instantly gathered using cameras and eye-tracking meters installed in the space. Through the automatic calculation of face recognition and technologies such as voice recognition and gesture recognition, data analysis is used to evaluate the expressive information of teachers and learners in the classroom.

Shao Y C et al [10] tried to reform the lesson on "Information Technology" in Natural Science and formed the classroom behavior curves of four teachers to obtain the optimal and the worst teaching plan sequences of the lesson. In the teaching reform practice, the teachers improved the classroom behavior curves to varying degrees by optimizing their lesson plans based on the best teaching plan sequence, the worst teaching plan sequence, the classroom behavior curve, and the students' feedback on the teaching plan.

The introduction of computer vision into the teaching classroom allows teachers to have real-time access to student performance and also helps teachers to adjust their teaching strategies at any time to ensure that they can reach all students, so that they can adjust their teaching content and methods in a timely manner to alert students who have wandered off, and students can also see their respective learning performance for the whole class after the lesson in a timely manner. AI-based classroom assessments have the potential to accelerate the evaluation process and improve feedback between teachers and students as compared to more conventional feedback techniques like tests, questionnaires, or interviews. This strategy has successfully aided in the development of classroom assessment, increasing its effectiveness and precision. Table 1 gives some typical classroom behaviors and representational actions of students.

Table 1. Typical student classroom behaviors and representational actions

Student Classroom Behavior	Characterizing behavioral actions
Listen to the lecture	Sitting upright with eyes level in front of you
Read the book	Small head bow, flipping through textbooks
Writing	Small head bowing, hand holding pen to write
whisper to each other's ear	Face deflected, lips moving slightly to talk to people around
Play Mobile	Head down sharply, hands on the table or under the table
Sleeping	Head down, supported by one hand or hands on the table
be stunned	Dull eyes, staring at something for a long time
Raise your hand	Body sitting upright, one hand rising straight up

Figure 1 shows the information collection and processing flow of classroom behavior recognition.

(i) Preprocessing of the video surveillance feed. The raw data set is obtained after pre-processing based on the raw data that the classroom cameras gathered and submitted.

(ii) Target behavior detection during information acquisition. using tools like deep learning, image processing, and other methods to extract and recognize items from the data.

(iii) Processing and analysis of features. Following behavior detection, the features are identified and processed into four categories, including voice, posture, facial, and physiological signal features. The relevant behavioral features are then extracted to define the typical behavioral features in the classroom.

(iv) Classifying behavior feature types. The behavioral categories of students can be determined by examining the behavioral traits that were retrieved after the recognition processing, and pertinent data is recorded and displayed.

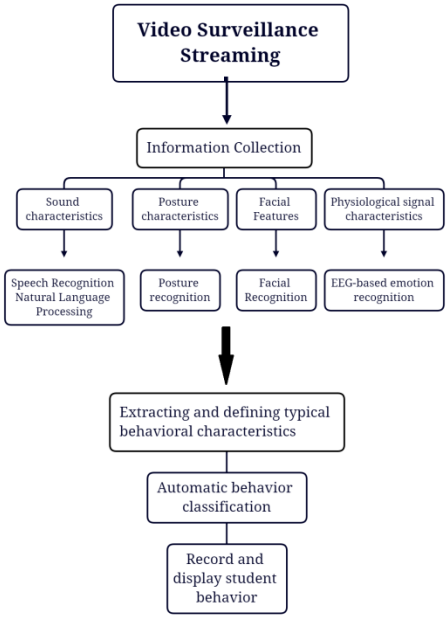

Figure 1. Classroom behavior recognition flow chart

3 Deep learning-based target detection method

SSD (Single Shot Multi-Box Detector) detection algorithm is an object detection algorithm proposed by Wei Liu in 2016. Figure 2 shows the SSD network structure. The algorithm is based on deep convolutional neural networks and is capable of detecting multiple objects in a single image in real-time. The fundamental benefit of the SSD approach over conventional object identification techniques, such as rapid R-CNN, is that it performs region suggestion and classification using a single network, which results in a large speedup. The SSD algorithm works by dividing the input image into a grid of fixed-sized boxes and predicting the presence of objects of different proportions and aspect ratios in each box. To increase the accuracy of object localization, the system also forecasts the offset value of each box. A set of bounding boxes and the matching class probabilities of the discovered items are the algorithm's final outputs. The SSD algorithm has been widely adopted in various computer vision applications such as autonomous driving, robotics, and security surveillance. Additionally, it has been expanded and enhanced in later research, such as SSD with MobileNet architecture, which increases the algorithm's accuracy and performance [11]. Wei Zhang et al.[12] optimized the SSD target detection algorithm to improve the teaching efficiency of English teachers in the teaching classroom, designed the optimized Mobilenet-Single Shot MultiBox Detector (Mobilenet-SSD), and through the relevant experiments on the analysis of students' behaviours in the classroom, the optimized algorithm achieved an average detection accuracy of 82.13%, which improves the detection efficiency and provides support for teachers to understand students' classroom learning status.

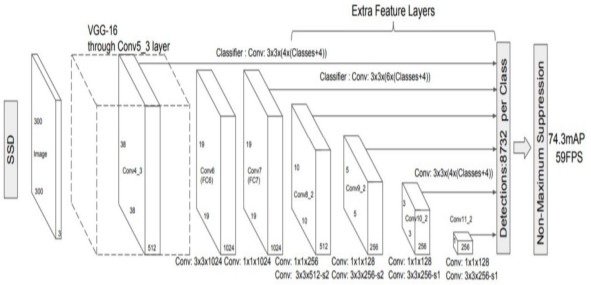

Figure 2. SSD Network Architecture

The YOLO (You Only Look Once) algorithm is an object detection algorithm proposed by Joseph Redmon et al. in 2016. Figure 3 shows the Yolo network structure. The main advantage of the YOLO algorithm is that it can detect multiple objects in an image in real-time and with high accuracy. It works by dividing the input image into a cell and predicting the bounding box and class probability of each cell. The algorithm uses a deep convolutional neural network for prediction, and the final output is a set of bounding boxes and the corresponding class probabilities of the detected objects. The YOLO approach uses a single neural network to carry out object localization and classification, in contrast to other object detection algorithms that employ area suggestion techniques, such as R-CNN. Because each image only needs to be processed once, the YOLO technique is much faster than other conventional object detection algorithms. The YOLO algorithm has been widely adopted for various computer vision applications such as robotics, surveillance, and autonomous driving. Additionally, it has been enhanced and expanded in the following studies, such as the YOLOv3 algorithm, which increases the method's accuracy and robustness.

Based on the YOLOX deep learning network, Wang et al. [13] examined the status of the classroom. Equation (1) represents the confidence score, and Equation (2) represents the deep learning network's loss function when applied to analyze student classroom status identification.

$$C_{\mathrm{j}}^{i} = P_{i,j} \times IOU_{pred}^{truth} \tag{1}$$

$$Loss = \mathrm{loss}_{\mathrm{Reg}} + loss_{Obj} + loss_{Cls} \tag{2}$$

The loss function is divided into 3 parts: bounding box regression loss loss_{Reg}, confidence loss, loss_{Obj} and category loss $loss_{Cls}$. During the training process, the loss function decreases and converges to the minimum value. During this process, the loss function decreases and converges to the mean. Once the loss function does not decrease or decreases below a certain threshold several times, the deep training network for the classroom state recognition method is considered to be trained and the whole exercise process is completed. Using a YOLOX-based deep learning network to analyze the classroom can

quickly and accurately assess students' learning, help teachers adjust and improve teaching methods in time, and improve teaching quality and learning efficiency.

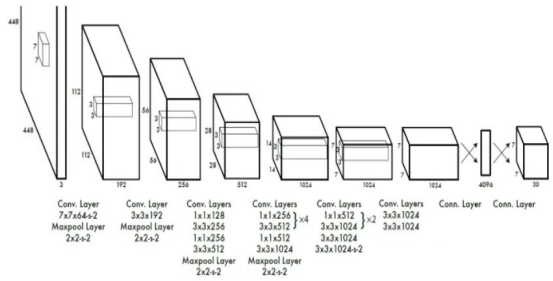

Figure 3. Yolo network structure

Ross Girshick et al. presented the object detection technique known as R-CNN, or region-based convolutional neural network, in 2014. It uses deep learning techniques and is the first algorithm to apply deep learning to object detection. On a number of commonly used datasets, it performs at the cutting edge. In Table 2, the R-CNN framework is displayed. It operates by segmenting the input image into regions of interest (RoIs) and utilizing a convolutional neural network that has already been trained to extract features from each RoI. The presence or absence of items in each zone of interest is then classified using a collection of class-specific linear SVMs using the attributes of each region of interest. Finally, the bounding box of each RoI is predicted using a regression model. The main advantage of the R-CNN algorithm over traditional object detection methods, such as HOG and SIFT, is that it learns better features from the data and achieves higher accuracy. However, the R-CNN algorithm also has some limitations, including its slow processing speed and high memory usage. Since the introduction of R-CNN, several other object detection algorithms built on its framework have been proposed, such as fast R-CNN, faster R-CNN, and masked R-CNN. these algorithms improve the speed and accuracy of the original R-CNN algorithm by optimizing different aspects of the system [14]. Likun et al. introduced an artificial neural network model based on R-CNN for hybrid English teaching assistant application. An accuracy of 97.89% and a sensitivity of 98.34% were achieved[15].

Table 2. Typical student classroom behaviors and representational actions

Region proposal（Selective Search）	
Feature extraction（CNN）	
Classification（SVM）	Bounding-box regression（regression）

4. Feature extraction and processing

4.1. Attitude data extraction-based processing

A technology called human gesture recognition uses human gestures, movements, and postures to represent the condition or purpose of the human body. This method entails modeling, fixation, and segmentation of the human body's structural components as well as additional analysis of the emotional inclinations expressed through gestures.

Figure 4 left shows the extraction of human skeletal joint point information regarding the automatic identification of student behaviour in the classroom, Xu J Z et al [16] conducted experiments and utilized the Boosting algorithm and convolutional neural network algorithm to assess the accuracy of five different models for the automatic detection of student classroom actions. The findings demonstrated that automatic student classroom behavior identification employing the human skeleton information extraction technique can correctly identify five common student actions in a challenging recognition environment like a school classroom. In a study by He X Let al. [17], the authors employed a method for extracting human skeletal information to convert the classroom behaviors of the students into skeletal data, which was then fed into a model created by a ten-layer convolutional neural network (CNN-10) for training and testing. The testing findings revealed that their self-developed student classroom behavior dataset had an average recognition accuracy of as high as 97.92%, correctly detecting seven prevalent categories of student behaviors. Peng L et al [18] employed transfer learning techniques and used a VGG-based pre-trained network model for feature extraction, and successfully achieved the detection and analysis of students' abnormal behaviors such as playing with cell phones and sleeping in the classroom. Jiang Q Y et al [19] established a classroom action recognition dataset for students and trained them on their actions through a deep residual system. With this technical tool, they succeeded in accurately identifying six typical student behaviors in the classroom. Liu X Y et al [20] used a multi-objective data regression method to build a student classroom behavior monitoring dataset and establish a classroom behavior monitoring model to effectively conduct classroom behavior monitoring. Yang B et al [21] developed a gesture recognition algorithm to achieve real-time recognition of human gestures based on the features of gestures on spatial distribution, which can be accurately recognized even in the case of complex background, As shown in figure 4, right.

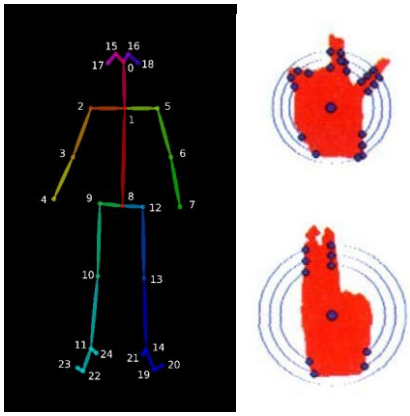

Figure 4. The left is the human skeleton joint point extraction map[14], and the right is the different region joint angle selection map[21]

4.2. Processing based on facial data extraction

Facial expression recognition refers to the analysis and recognition of expressions expressed by facial muscle movements through face images or video data to infer the emotional state of human beings. There are typically two approaches for processing facial expression recognition. To identify and categorize various facial emotions, the first method is to recognize local aspects of the face, such as by examining the locations and movements of important areas like the eyebrows and eyes [22]. The second method involves identifying the overall features of the face, or examining the features of the face as a whole and identifying the facial features that are portrayed by various facial emotions. A face image can be represented using face recognition techniques as a vector X of size n*1 (n is the image's width multiplied by its height), where N is the total number of training samples. At this point, the overall scatter matrix of the training sample set is used as the generation matrix [23]. That is:

$$C = E\left[XX^T\right] \approx \frac{1}{N}\sum_{k=1}^{N} X_k X_k^T$$

Represent the N face vector with an n*N matrix

$$X = \left[X_1, X_2, X_3, \cdots, X_n\right]$$

Then C can be expressed as: $C \approx \frac{1}{N} XX_T$

Jia L Y et al [24] developed a face model with 68 key features using the SVM method in order to successfully recognize four distinct expression styles of pupils in a teaching environment, as shown in figure 5. A bimodal emotion recognition model developed by Ma X L et al [25] enables a thorough examination of pupils' emotional states. The model determines the students' present affective states by simultaneously analyzing their body language and facial expressions. Han L et al [26] use of multi-pose face detection and facial recognition algorithms allowed them to successfully recognize five distinct expression patterns of students in the teaching classroom by extracting feature changes in three crucial student body parts—the heads, eyes, and lips.

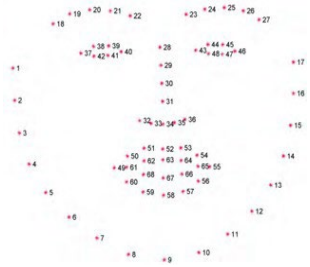

Figure 5. Facial Feature Extraction[26]

4.3. Extraction processing based on sound data

The intonation of the speech signal can be used to reflect the speaker's current emotional state while processing acoustic data. For instance, high tones frequently convey passion and joy, whereas low tones frequently convey melancholy and gloom. Iain R. Murra et al [27] have relatively effectively summarized a large amount of practical experience through the classification of prosodic parameters in terms of dialogue flow, mean fundamental frequency, fundamental frequency range, fundamental frequency variation, intelligibility, etc., which has laid the cornerstone for our speech intonation analysis and recognition. However, in a real complex environment, noise can cause serious interference to us when processing sound data, which makes it difficult to perform accurate recognition. To address these challenges, we must therefore continue to enhance the methodologies for acoustic data extraction and processing.

4.4. Extraction processing based on physiological signals

We can more precisely depict tiny variations in a character's emotions by gathering and researching several forms of physiological data (such as EEG, ECG, E&M, and body surface temperature). Students can undertake more precise analyses of classroom state behavior thanks to a dedicated smart classroom platform, which enhances instruction in the classroom. Among these, the technique for determining pupils' emotional states based on EEG signal characteristics offers the advantages of simplicity of use and superior outcomes. Scientists are becoming more and more interested in this method since it may be used to identify different emotional states and analyze EEG signal properties at four key levels: time domain, frequency domain, time-frequency region, and spatial domain [28-29].

5 Behavioral characteristics classification method

Currently, decision tree-based, Bayesian classifiers, support vector machines, artificial neural networks, and deep learning is some of the common techniques employed in behavioral categorization research. Each of these techniques can be used for various data kinds and application settings and has advantages and limitations. Choosing the best strategy for training and optimization in practical applications requires taking the data type and application environment into account.

(i)Decision tree-based classification method. Decision tree-based classification is a nonparametric method that recursively partitions the data into smaller and smaller subsets, with each partition determined by a test on one of the input features. The decision tree algorithm builds a tree-like model where each internal node represents a test on an input feature, each branch represents a possible outcome of the test, and each leaf node represents a class label or a probability distribution of class labels. The categorization error or information received at each step is minimized or maximized depending on how the decision tree is built. To classify a new instance using a decision tree, the algorithm starts at the root node and tests the input features according to the paths leading to the appropriate leaf nodes. The classification is then determined by the category labels or probability distributions associated with the corresponding leaf nodes. Decision tree-based classification methods have several advantages, including their interpretability, simplicity, and ability to handle numerical and categorical data. However, they may suffer from overfitting, instability due to small changes in the data, and poor performance when dealing with high-dimensional data. Zhou J G et al [30] used a combination of decision trees and gradient-boosting methods to successfully classify the facial expressions presented by students in class.

(ii)Bayesian classifier-based approach. The Bayesian classifier-based approach is based on Bayes' theorem and uses probabilities to determine the category labels for a given data instance. The Bayesian classifier works by first estimating the prior probabilities for each category in the training data. The conditional probability of each feature in each category is then estimated for each feature of the input data. The category with the highest probability is then designated as the final classification result after combining these probabilities to get the posterior probability of each category in the input data instance. A Bayesian network classifier, which employs a directed acyclic graph to describe the dependencies between features, is more complex than a simple naïve Bayesian classifier, which assumes that all features are independent under a given class label. Its simplicity, effectiveness, and capacity for handling missing data and superfluous features are only a few of its many benefits. However, they could perform poorly if the prior probabilities do not accurately reflect actual class distributions or if the assumption of independence or conditional independence between characteristics is broken. To overcome these limitations, various extensions and variants of Bayesian classifiers have been developed, such as kernel density estimation, Bayesian belief networks, and Bayesian model averaging. Bayesian causal networks were employed by Zhou P X et al [31] to estimate classroom behavioral parameters, define several target behavior categories, and ultimately select the teaching model applied in that classroom.

(iii)Support vector machine-based approach. The Support Vector Machine (SVM) based approach's fundamental model is a linear classifier that calculates the greatest marginal hyperplane to divide two classes in the feature space, and it may be expanded to nonlinear classification using kernel functions. It works by mapping the input data into a high-dimensional feature space, and it tries to find the best hyperplane that maximizes the margin between the two classes. Support vectors are the data points closest to the margin boundary and play a key role in determining the optimal solution. SVMs have several advantages, including their ability to handle high-dimensional data, nonlinear decision boundaries, and outliers. However, they can suffer from overfitting and poor performance on unbalanced datasets. To overcome these limitations, various extensions and variations of SVMs have been developed, such as support vector regression, multicore learning, and online SVMs. Gong W [32] obtained information about the skeletal focus of learners through the OpenPose skeletal key point analysis model, which in turn led to the categorical identification of three cognitive behaviors through an SVM classifier.

(iv)Artificial neural network-based approach. In order to learn from data and make predictions or judgments, machine learning algorithms that are based on artificial neural networks (ANNs) imitate the structure and operation of biological neural networks, notably the brain. Each neuron gets information from other neurons, uses an activation function to analyze this information, and then creates an output. An artificial neural network (ANN) is composed of interconnected artificial neurons or nodes stacked in layers, each of which transmits output to the neurons in the layer below. To help the network perform better on a particular task, the connections between neurons are weighted, and the weights are modified throughout training. ANNs can be used for a wide range of applications such as image and speech recognition, natural language processing, prediction, and control systems. They can also be combined with other machine learning techniques, such as deep learning and reinforcement learning, to improve their accuracy and efficiency. One researcher collected students' behavioral data in a classroom using a Kinect deep camera and implemented a BP neural network-based classifier using MATLAB to classify and identify students' behaviors such as seating position, ear washing, and looking at the blackboard.

(v)Deep learning-based approach. Deep learning-based approaches are a subset of machine learning that has become increasingly popular in recent years due to their success in a wide range of applications such as image and speech recognition, natural language processing, and recommendation systems [33-34]. Deep learning algorithms

feed data through numerous layers of nonlinear transformations, creating high-level abstractions of the data that have a hierarchical character. This eliminates the need for manual feature engineering by enabling them to automatically learn features and representations from the original data. Convolutional neural networks, recurrent neural networks, and deep belief networks are a few examples of common deep learning architectures. These architectures are typically trained using large amounts of labeled data and powerful GPUs or specialized hardware gas pedals, such as TPUs. The success of deep learning is mainly attributed to its ability to utilize massive amounts of data, coupled with advances in computational power and optimization techniques. Zhang Yiwen et al [35] added an attention mechanism module to the YOLO-v3 algorithm framework, thus effectively enhancing the training effect on students' class behavior characteristics and effectively in the SICAU-Classroom dataset, the classroom behavioral characteristics are accurately identified. On the other hand, Hou C K [36] used the Mel inverse spectral coefficient property of deep recurrent neural networks (RNN) to obtain automatically detected classroom instructor behaviors.

6 Conclusion

Although the use of computer vision recognition in the classroom helps professors monitor students' learning progress in real-time and enhance the caliber of lectures, there are still a number of challenges and issues that require immediate attention.

(i)Target behavior recognition in a real setting requires increased accuracy. Due to the limitations of the camera's field of vision and the actual classroom setting, which includes things like desks that obscure student targets, the subsequent research is challenging. The picture information, optical flow information, and audio information acquired in the video may be fully merged using the technique of fusing target information from different perspectives, and deep neural networks are formed for feature extraction, improving the detection effect.

(ii)Improving the individualized feedback for instructor behavior. While few people pay attention to teachers' behaviors in the classroom, classroom teaching evaluation using artificial intelligence technology currently places a greater emphasis on students' classroom learning behaviors and uses intelligent means to help teachers complete monitoring of students' learning status in order to achieve the goal of timely adjusting teaching content and improving teaching methods. In order to increase individualized feedback on instructors' classroom actions, we should strive to apply artificial intelligence technologies to recognize teachers' teaching habits.

(iii)Pay attention to data protection and information security. In terms of privacy invasion, the application of AI for teaching assessment necessitates accurate recording of instructor and student movements, language, and actions throughout the classroom. To fully safeguard privacy, regulations should be put in place for the responsible use and safe preservation of data.

As AI and education become more and more deeply integrated, classroom teaching and research can achieve a full range of classroom trajectory monitoring by tracking multimodal data (e.g., voice, image, video, space, and posture), analyzing and extracting multi-dimensional classroom data, and achieving intelligent diagnosis and feedback, and intelligent acquisition and analysis technologies based on multimodal data will also serve classroom teaching evaluation practices more extensively, contributing to a shift in classroom teaching evaluation toward efficiency, accuracy, and intelligence.

Computer vision introduces artificial intelligence into the teaching classroom by providing a new perspective on multidimensional recognition and analysis of classroom teaching scenarios. This will enable teachers' teaching behaviors to be analyzed and quantified supported by real data, thus providing a basis for developing a scientific and reasonable teaching approach. It is foreseeable that future AI technology will be further applied to teaching and provide strong support for teaching reform.

References

[1] Jabbar M A, Kantipudi M V V P and Peng S L, "Machine Learning Methods for Signal," Image and Speech Processing, CRC Press, 2022.

[2] Tian Y, "Artificial intelligence image recognition method based on convolutional neural network algorithm," IEEE Access, vol. 8, pp. 125731-125744, 2020.

[3] Yang M, Kumar P and Bhola J, "Development of image recognition software based on artificial intelligence algorithm for the efficient sorting of apple fruit," International Journal of System Assurance Engineering and Management, pp. 1-9, 2021.

[4] Chen H, Geng L and Zhao H, "Image recognition algorithm based on artificial intelligence," Neural Computing and Applications, pp. 1-12, 2021.

[5] Feng H H and Zou B, "Exploring and Reflecting on Student Motivation Enhancement in Traditional Classrooms," Journal of Higher Education, vol. 7, no. 25, pp. 56-59, 2021.

[6] Zhang J X, Zhu L and Wu Z F, "Classroom Immediate Assessment Based on Information Technology," Journal of Yangzhou University (Higher Education Research Edition), vol. 22, no. 2, pp. 80-85, 2019.

[7] Nash B L, "Constructing meaning online: Teaching critical reading in a post-truth era," The Reading Teacher, vol. 74, no. 6, pp. 713-722, 2021.

[8] Malin H, "Teaching for purpose: Preparing students for lives of meaning," Harvard Education Press, 2021.

[9] Munna A S and Kalam M A. "Teaching and learning process to enhance teaching effectiveness: a literature review," International Journal of Humanities and Innovation (IJHI), vol. 4,no. 1, pp. 1-4, 2021.

[10] Shao Y C, Li C D and Cao Y, "Intelligent Analysis of Classroom Behavior in Teaching Reform," Teaching and Management, 2021(15): 29-33.
[11] Chiu Y C, Tsai C Y and Ruan M D, Mobilenet-SSDv2: "An improved object detection model for embedded systems", 2020 International Conference on system science and Engineering (ICSSE), 2020: IEEE, pp. 1-5.
[12] Zhang W, Xu Q, "Optimization of college English classroom teaching efficiency by deep learning SDD algorithm," Computational Intelligence and Neuroscience, 2022, 2022.
[13] Wang G H, Zhang X and Zheng H, "Analysis of students' classroom learning status based on deep learning," Journal of Higher Education, vol. 8, no. 31, pp. 1-5, 2022.
[14] Rosati R, Romeo L and Silvestri S, "Faster R-CNN approach for detection and quantification of DNA damage in comet assay images," Computers in Biology and Medicine, 2020, 123: 103912.
[15] L Yuan, "Research on English Hybrid Assisted Teaching System Using Contextual Support of R-CNN," Wireless Communications and Mobile Computing, 2022, 2022.
[16] Xu J Z, Deng W and Wei Y T, "Automatic Recognition of Student's Classroom Behaviors based on Human Skeleton Information Extraction," Modern Education Technology, vol. 30, no. 5, pp. 108-113, 2020.
[17] He X L, Yang F and Chen Z Z, "The Recognition of Student Classroom Behavior based on Human Skeleton and Deep Learning," Modern Education Technology, vol. 30, no. 11, pp. 105-112, 2020.
[18] Liao P, Liu C M and Su H, "A deep learning-based system for detecting and analyzing abnormal student behavior in the classroom," Electronic World, 2018(08): 97-98.
[19] Jiang X Y, Zhang Z W and Tan S Q, "Student classroom behavior identification based on residual network," Modern Computers, 2019(20): 23-27.
[20] Liu X Y, Ye S P and Zhang D H, "Improved multi-objective regression student classroom action detection method," Computer Engineering and Design, vol. 41, no. 9, pp. 2684-2689, 2020.
[21] Yang B, Song X N and Feng Z Q," Gesture Recognition in Complex Background Based on Distribution Features of Hand," Journal of Computer-Aided Design & Computer Graphics, vol. 22, no. 10, pp. 1841-1848, 2010.
[22] Pantic M., Rothkrantz and L.J.M.,"Facial Action Recognition for Facial Expression Analysis From Static Face Images ," IEEE transactions on systems, man, and cybernetics, Part B.Cybernetics, vol. 34, no. 3, pp. 1449-1461, 2004.
[23] Ding R, Song G D and Lin X G, "Comparison of Eigenface and Elastic Matching in Human Face Recognition," Computer Engineering and Applications, 2002(07): 1-2+19.
[24] Jia L Y, Zhang C H and Zhao X Y, "Analysis of Students Status in Class Based on Artificial Intelligence and Video Processing," Modern Education Technology, vol. 29, no. 12, pp. 82-88, 2019.
[25] Ma X L, Guo S N and Wu Y H, "The Recognition and Application of Educational Emotion Based on Image Analysis," Modern Education Technology, vol. 30, no.2, pp. 118-121, 2020.
[26] Han L, Li Y and Zhou Z J," Research on the Relationship of Critical Thinking and Community of Inquiry Model," Modern Distance Education Research, 2017(04): 97-103+112.
[27] Iain R.Murray and John L.Arnott,"Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion ,"The Journal of the Acoustical Society of America,vol. 93,no. 2, pp. 1097-108, 1993.
[28] Wang H, Guo H and Zhang K, "Automatic sleep staging method of EEG signal based on transfer learning and fusion network," Neurocomputing, 2022, 488: 183-193.
[29] Hemanth D J, "EEG signal based modified Kohonen neural networks for classification of human mental emotions,"Journal of Artificial Intelligence and Systems, vol. 2,no. 1,pp. 1-13, 2020.
[30] Zhou J G, Tang D M and Peng Z," Students' Expression Analysis in the Classroom based on Gradient Boosting Decision Tree and Convolution Neural Network," Journal of Chengdu University of Information Engineering, vol. 32, no. 5, pp. 508-512, 2017.
[31] Zhou P X, Deng W and Guo P Y," Research on Intelligent Recognition of S-T Behavior in Classroom Teaching Video," Modern Education Technology, vol. 28, no. 6, pp. 54-59, 2018.
[32] Gong w, "Design and Implementation of Student Learning Behavior Recognition System Based on Key Points Detection of Bones," Jilin University, 2019.
[33] Hinton G E and Salakhutdinov R R, "Reducing the dimensionality of data with neural networks," Science, 2006, 313 (5786) : 504－507.
[34] Jermsittiparsert K, Abdurrahman A and Siriattakul P, "Pattern recognition and features selection for speech emotion recognition model using deep learning," International Journal of Speech Technology, vol. 23, pp.799-806, 2020.
[35] Zhang Yiwen, Wu Zhe and Chen Xianjin, "Classroom behavior recognition based on improved yolov3", 2020 International Conference on Artificial Intelligence and Education (ICAIE) , 2020: IEEE, pp. 93－97.
[36] Hou C K, "Research on Automatic Recognition of classroom teacher behavior based on multimodal fusion," Huazhong Normal University, 2020.