Research Article

Classroom Learning Status Assessment Based on Deep Learning

Jie Zhou,1 Feng Ran,2 Guang Li,3 Jun Peng,1 Kun Li,4 and Zheng Wang5

1School of Education, City University of Macau, Macao, China
2Beijing Dongcheng Academy of Educational Sciences, Beijing, China
3Faculty of Data Science, City University of Macau, Macao, China
4Beijing No. 171 High School, Beijing, China
5Shandong Youth University of Political Science, School of Information Engineering, Jinan, China

Correspondence should be addressed to Jun Peng; d19092105030@cityu.mo

Received 16 February 2022; Revised 11 March 2022; Accepted 17 March 2022; Published 16 April 2022

Copyright © 2022 Jie Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Student classroom behavior performance is an important part of classroom teaching evaluation, and conducting student classroom behavior recognition is important for classroom teaching evaluation. The article proposes a deep learning-based student classroom behavior recognition method, which extracts the key information of the human skeleton from student behavior images and combines a 10-layer deep convolutional neural network (CNN-10) to recognize students’ classroom behavior. To verify the effectiveness of this method, the paper conducts a comparison experiment on the student classroom behavior dataset using CNN-10 and the student classroom behavior recognition method. The experimental results show that the student classroom behavior recognition method can effectively exclude the interference of irrelevant information such as students’ physique, dress, and classroom background, highlight the key effective information, and have higher recognition accuracy and generalization ability. Using the human skeleton and a deep learning-based student classroom behavior detection approach to identify students’ typical classroom behaviors might improve intelligent classroom teaching by reflecting students’ learning status in a timely and effective manner.

1. Introduction

Classroom teaching is the main site of school education and teaching, and process evaluation of classroom teaching is of great significance to improve teaching quality, while students’ classroom behavior performance is an important part of classroom teaching evaluation. In traditional classroom teaching, teachers only observe students’ behavior in class or watch classroom videos after class to conduct a process evaluation, which takes a lot of time and energy and cannot be observed and measured continuously on a large scale. Generally, teachers mostly use Flanders interaction analysis system (FIAS) [1], S-T (student-teacher) classroom pedagogical analysis method, information technology-based interaction analysis system (ITIAS) [2], improved Flanders interaction analysis system (iFIAS) [3–5] and other methods, as well as with the help of classroom video analysis software, observation scales, and other tools to analyze students’ classroom behavior. However, in essence, these methods still belong to the category of manual analysis, which is labor-intensive and inefficient.

With the expansion of college enrollment in recent years, the number of students in classes is increasing, and university instructors are increasingly burdened with managing classrooms and teaching lectures. With the popularization of smart campuses, detecting abnormal behaviors of students in the classroom through deep learning is one of the important measures to improve classroom efficiency. With the increasing development of computer vision, abnormal behavior detection based on computer vision can be implemented in two ways. The traditional method is based on the grayscale value of image pixels and achieves the effect of detection and recognition by sliding the human-designed feature
template over the original image and matching whether the pixel points are consistent. Traditional feature extraction algorithms are SIFT (scale-invariant feature transform) [6] and ORB (oriented fast and rotated brief) [7, 8], and traditional classification is implemented by support vector machine (SVM) [9] or AdaBoost [10]. When the pixel distribution of the detected objects is constant, fast and accurate detection and recognition can be obtained, but when the same class of objects presents different states in the image, it is difficult to obtain good detection and recognition results by only relying on artificially set feature templates. In recent years, with the significant increase of computer GPU computing power and the research of deep learning by researchers around the world, convolutional neural networks have become the mainstream of detection and recognition technology with the advantage of extracting a large number of features. Such techniques are divided into two-step detection and single-step detection, the former uses a strategy of generating feature candidates through a convolutional network and then feeding the generated features into the network for detection and classification to achieve the recognition and detection function, and the related algorithms are R-CNN series [11–15]; the latter simplifies the target detection problem into a single-step regression problem by directly obtaining the anchor frame from the image and thus eliminating the need for candidate generation network, common algorithms such as YOLO series [16–19] and SSD (single shot multibox detector) [20, 21], etc.

Traditionally, educational research has concentrated on classroom instructional practises. Students are the primary subjects of learning activities in the classroom, and their conduct is a direct reflection of teaching activities, which not only reflects students’ present learning level but also, to some extent, indicates classroom efficiency. By analyzing students’ classroom behaviors, teachers can accurately grasp students’ listening status and dynamically adjust teaching progress according to students’ classroom status to improve classroom efficiency. As a result, it is critical to recognize and assess kids’ classroom behavior. Due to the complexity of students’ classroom behavior, the recognition effect is easily affected by some irrelevant factors, and the training of deep neural networks is not conducive without the support of the large-scale students’ classroom behavior data set. Pose estimation technology refers to the process of recovering human joint points from a given image or video to obtain the human skeleton, whose main task is to depict the shape of the human body in the image or video - pose estimation technology refers to the process of recovering human joint points from a given image or video to obtain the human skeleton, whose main task is to depict the shape of the human body in the image or video. This is because the human skeleton information is more suitable for characterizing behavioral actions compared to the raw image data. In this study, we propose a deep learning-based student classroom behavior recognition method (hereinafter referred to as “student classroom behavior recognition method”), which extracts the key information of the human skeleton from student behavior images and combines a 10-layer deep convolutional neural network (CNN-10) to recognize students’ classroom behaviors in order to eliminate the complex factors interference to further improve the accuracy of student classroom behavior recognition.

The following is a breakdown of the research: Section 2 discusses detecting student behavior in the classroom using the human skeleton and deep learning. Section 3 delves into the planning and execution of the experiment. The analysis and discussion of experimental data are covered in Section 4. Finally, the research job is completed in Section 5.

2. Identifying Student Classroom Behavior Based on Human Skeleton and Deep Learning

2.1. Human Skeleton Information Extraction Based on Open Pose

In this paper, the student abnormal behavior detection system is mainly composed of three modules: video acquisition, image recognition, and information response, firstly, through the classroom surveillance camera, video data is supplied to the computer for preprocessing, and then the preprocessed key-frames are input to the improved YOLO convolutional network for recognition and detection, and the location, category and time information output from the encapsulated statistical network is used as API interface for other systems to call in real time. Information as API interface for other systems to call in real time. The flow chart of the student abnormal behavior detection system is shown in Figure 1.

The video acquisition module captures video from the camera in the classroom scene, key-frames the video, and then sends the key-frame image data to the image recognition module; the function of the image recognition module is to preprocess the image for data enhancement and then detect the abnormal behavior through the deep neural network; the aberrant information is sent to the picture recognition module by the information response module. The information response module’s job is to calculate the number of aberrant behaviors in the image, the category, and the proportion of abnormal behaviors using numerical statistics on the abnormal data. The function of the information response module is to calculate the number of abnormal behaviors in the image, the category, the percentage of abnormal behaviors, etc., and encapsulate this information into a class. The information is encapsulated into classes and called by other modules.

OpenPose is an open-source library based on convolutional neural network and supervised learning with Caffe as the framework, which can obtain the skeletal key points of people in images in a timely and accurate manner, and is suitable for single or multiperson behavior recognition. After getting the detection results, the key points are then connected using the affinity of human skeletal key points to obtain the human skeleton map, as shown in Figure 2. In student classroom behavior recognition, using human skeleton information can effectively exclude much redundant information unrelated to behavior, such as students’ physique, dress, classroom background, etc., while highlighting human behavior key information, thus reducing the complexity of behavior recognition.
2.2. Convolutional Neural Network for Student Classroom Behavior Recognition. In image identification, speech recognition, and other domains, convolutional neural networks (CNN) are now frequently employed. Convolutional neural networks, such as AlexNet, VGGNet, GoogLeNet, and others, are used in practically all deep learning-based methods, particularly in the field of image identification. An input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer are all common components of convolutional neural networks. Convolutional neural networks have the powerful fitting ability and can effectively extract image features. Building a suitable convolutional neural network for student classroom behavior recognition should consider both the complexity and expressiveness of the network model and the ability to fit the established student classroom behavior dataset. Based on the previous research results, the convolutional neural network was built by repeating “overlapping convolutional layers twice, then pooling layers” and finally two fully connected layers. A 10-layer deep convolutional neural network (CNN-10) with 8 convolutional layers, 4 pooling layers, and 2 fully connected layers was built for student classroom behavior.
recognition in order to adapt the constructed convolutional neural network to the student classroom behavior dataset, as shown in Figure 3.

2.3. Behavior Recognition Based on Human Skeleton Information and Deep Learning. Figure 4 shows the construction of a behavior recognition procedure based on human skeleton knowledge and deep learning in this work. In the first step, the classroom behavior data is collected, which contains images and labels of seven kinds of student behaviors. The second step is to extract the human skeleton information, i.e., using OpenPose to obtain the human skeleton information of the student classroom behavior images-first, effectively eliminating the distracting factors such as students’ physique, dress, and classroom background; then, separating the background information from the human skeleton image while keeping the size constant; finally, only Finally, only the human skeleton information is retained. The third step, training, and testing are to input the obtained human skeleton images into the built-the third step, training, and testing are to input the obtained human skeleton images into the constructed deep convolutional neural network CNN-10 for training and testing.

In the CNN-10 used in this study for student classroom behavior recognition, the input feature map size is uniformly set to $112 \times 112$ to correspond to the size of the student classroom behavior image; the convolutional kernels of the convolutional layers are all small $3 \times 3$ filters with an operation step of 1, and the number of convolutional kernels of the 8 convolutional layers is 16, 16, 32, 32, 64, 64, 128, 128 Finally, two fully-connected layers of different sizes are used for feature reduction, and the number of output neurons of the latter fully-connected layer is 7, which corresponds to the 7 categories of students’ classroom behaviors. Traditional convolutional neural networks frequently utilize the sigmoid function as the activation function, but the Sigmoid function is prone to saturation during the gradient descent phase, terminating the gradient transfer. Because the ReLU function offers the characteristics of fast convergence and a simple gradient solution, it is used as the activation function $h(x)$ in this work, whose value is shown in (1). This study uses the initial values proposed by He et al. [22] as the initial values of the network weight parameters in order to make the activation values of each layer with appropriate breadth. The parameters were updated using an Adam-based optimization method with a default value of 0.001 for the network learning rate, while the cross-entropy error was used as the loss function of the network.

$$h(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

(1)

The batch normalization layer and dropout layer are used after the first fully connected layer to increase the convolutional neural network’s robustness and training speed, while the dropout layer is used after the second fully connected layer with the dropout parameter set to 0.5. The output layer of the convolutional neural network uses the softmax function $y_k$, which is calculated as shown in (2).

$$y_k = \frac{\exp(a_k)}{\sum_{i=1}^{n} \exp(a_i)}$$

(2)

The correlation coefficient is calculated as

$$R = \frac{\sum_{m} \sum_{n} AB}{\sqrt{\sum_{m} \sum_{n} A^2} \sqrt{\sum_{m} \sum_{n} B^2}}$$

where $A$ and $B$ represent any image, $m$ and $n$ are the width and height are the width and height of the image, respectively, $N$ is the number of channels in the image. After many attempts, when the correlation coefficient $R < 0.8$ is the key-frame to be extracted, it can be determined that when the correlation coefficient is the key-frame, the video has a sudden change.

3. Experiment Preparation and Implementation

In order to verify the effectiveness of the student classroom behavior identification method, this study conducted a
3.1. Construction of Student Classroom Behavior Dataset. Due to the limited number of students’ classroom behavior data sets currently available, this study uses a SONY FDR-AX30 digital 4K camcorder to collect 400 students from an experimental primary school. The image data of students’ classroom behavior is collected by means of a single-person image, which is used to construct a data set of students’ classroom behavior. This paper collects 90 classroom videos and builds a classroom behavior library. We collected 10,000 pictures of students raising their hands, 1,000 pictures of students bending over, 10,000 pictures of students walking back and forth, 10,000 pictures of students writing on the blackboard, 10,000 pictures of students looking up, 10,000 pictures of students bowing their heads, 1,000 pictures of students standing, 1,000 pictures of students Pictures of raised hands, and 1,000 pictures of students lying on their desks. The test environment, monitoring equipment, and behavioral data are shown in Figures 5–7.

In this study, the collected images were cropped and saved according to the upper body region of the human body constructed based on the key points of the human skeleton, and then uniformly scaled to an image with a size of $112 \times 112$ (blank padding) based on the long side. Hands up, Head down, Listening to class, Standing, Low, and other pictures of normal classroom conduct with a high frequency of students are summarized in this study.

3.1.1. Experimental Procedure. First, randomly choose four-fifths of each student’s classroom conduct as a training sample and the remaining one-fifth as a test sample from the data set of students’ classroom behavior.

Secondly, based on the consideration of the balance between the performance of the experimental machine and the training efficiency, the mini-batch is set to 160 and the epoch is set to 30.

Finally, CNN-10 and student classroom behavior recognition methods were used for training and testing, and in order to reduce random errors, 10 random experiments were conducted using these two methods, and then the average recognition accuracy of these two methods was calculated and calculated.

3.1.2. Human Joint Point Settings. A total of 18 human body joints are set in this paper, as shown in Figure 8.

3.1.3. Evaluation Indicators. Model evaluation generally uses accuracy (ACC), recall (R), and precision (P), as shown in equations (4)–(6).

\[
ACC = \frac{TP + TN}{total}, \quad (4)
\]

\[
R = \frac{TP}{TP + FN}, \quad (5)
\]

\[
P = \frac{TP}{TP + FP} \quad (6)
\]

4. Analysis and Discussion of Experimental Results

In order to verify the effectiveness of the student classroom behavior identification method, this study conducted a comparison experiment on the student classroom behavior dataset using CNN-10 and the student classroom behavior identification method, respectively.

4.1. Construction of Student Classroom Behavior Dataset. Under the same experimental conditions, after 10 random experiments, this study compared the accuracy of CNN-10 and the student classroom behavior recognition method in identifying multiple classroom behaviors on the student classroom behavior dataset. Through calculation, this study concluded that the average recognition accuracy of CNN-10 was 93.65%, while the average recognition accuracy of students’ classroom behavior recognition method was 97.92%. It can be seen that both methods have high recognition accuracy. And the result of target recognition is shown in Figure 9.

Accuracy (ACC), Recall (R) and Precision (P) are all commonly used evaluation metrics in target detection.
The results of the evaluation indicators of students’ classroom behavior recognition are shown in Table 1. As can be seen from Table 1, the deep learning-based classroom learning status assessment proposed in this paper has very good accuracy and recall in identifying students’ classroom behaviors, and the behavior recognition framework has achieved good results in identifying students’ and teachers’ classroom behaviors.

It can be found that hand-raising behavior recognition mainly relies on local information such as hands, which is easily disturbed by factors such as students’ physique, dress, and classroom background; students’ listening behavior has a certain action amplitude, and if its amplitude is too small, it is easily misjudged as other actions; some listening behaviors and head raising behaviors are too similar, and the distinction is subtle, which is also easy to cause misjudgment.
Figure 7: Examples of some student classroom behavior datasets (including human skeleton information).

Figure 8: Human joint point settings.

Figure 9: Recognition result graph.
Limited by the student classroom behavior dataset, it is difficult to satisfy the classification of all student classroom behaviors by only using convolutional neural networks to obtain behavioral features directly from the original image data. After using the human skeleton information and deep learning method proposed in this study, the gap between the recognition accuracy of each classroom behavior narrowed, and the overall recognition accuracy increased by 4.27%, among which the recognition accuracy of hand-raising behavior was as high as 98.3%. Compared with the original image data, the human skeleton information extracted when using the student classroom behavior recognition method to identify six typical classroom behaviors is more efficient in describing the characteristics of classroom behaviors and actions; combined with the deep learning method, the student classroom behavior recognition method can effectively exclude the influence of irrelevant factors such as students’ physique, dress, and classroom background, and highlight the key effective information, with stronger generalization ability and higher recognition accuracy.

5. Conclusion

A crucial problem in developing the integration of blended learning engagement research with university teaching practice is the efficient and accurate assessment and analysis of students’ classroom learning behavior engagement. The differences in classroom behaviors of different individuals determine that it is difficult for researchers to measure them scientifically with traditional methods. With the innovation and development of technology, relying on computer vision and other technologies to accurately identify individual behaviors has become a reality, which provides researchers with the possibility to measure and evaluate classroom learning behavior engagement. This study proposes a deep learning-based method for student classroom behavior recognition, and validates the effectiveness of this method through comparative experiments: this method extracts human skeleton information from images, inputs the skeleton information of student classroom behavior into a 10-layer convolutional neural network (CNN-10) for training and testing, and recognizes a variety of classroom behaviors on an autonomously constructed student classroom behavior dataset. The average recognition accuracy of this method reached 97.92%, and it can effectively exclude the interference of irrelevant factors such as students’ physique, dress, and classroom background, and highlight the effective information, thus verifying the effectiveness of this method. Students’ normal classroom actions may be recognized using a student classroom behavior recognition approach based on the human skeleton and deep learning (including raising hands, listening to lectures, etc.), which can timely and effectively reflect students’ learning status, and it can help teachers accurately grasp students’ classroom learning conditions, thus helping intelligent classroom teaching. It should be noted that the accuracy of student classroom behavior recognition is relatively low. The recognition accuracy of student classroom behavior recognition method depends on the accuracy of posture estimation results, and the recognized student classroom behavior is simple the classroom behaviors identified are simple and limited in variety. To enhance this technique, we need to test other posture estimate algorithms and gather data from bigger sample size and a wider range of classroom activities. More diversified data may be collected to improve the approach.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Beijing Municipal Educational Science “Thirteenth Five-Year Plan” General Project in 2019 “Research on Evaluation Index System in the Context of General High School History Curriculum Reform” (project no.: CDDB19182) and the 2021 General Project of Beijing Society of Education “Development and Application of WeChat Official Account for Departmental History Textbook Aided Teaching” (project no.: DCYB2021-076)

References


