International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 5, No. 2, 2022

Video-based Student Classroom Classroom Behavior State Analysis

Yinggan Cheng

Department of Computer Science and Technology, Qingdao University, Qingdao, Shandong Province, China

* **Corresponding author**: Cheng Yinggan (Email: 792563254@qq.com)

Abstract: Looking back on the development of education, education has evolved along with the development of human society, and the development of science and technology has been the fundamental force driving the change of education. And with the increasing demand for personalized education for students, the focus of education and teaching has gradually shifted to the direction of personalization and specialization of students. In the teaching process, students' classroom behavior is an important reference indicator for evaluating students' classroom learning, but teachers cannot pay attention to each student's classroom behavior in time, and can only derive students' classroom learning status from their homework or exams. Many subjective factors can easily affect teachers' judgment of students' learning status, and the evaluation obtained is not always comprehensive and objective enough. In traditional teaching activities, teachers mainly identify and control students' classroom videos manually, but this method is not only time-consuming and labor-intensive, but also affects teachers' classroom quality. Therefore, it is urgent to study an intelligent and effective system for analyzing students' classroom behavior status. It is of great significance to improve the quality of students' classroom learning and to assist teachers in obtaining teaching feedback information. At present, the rapid development of artificial intelligence is reshaping the world's industrial ecology and pushing human society into the era of intelligence. Artificial intelligence is deeply integrating with education and promoting innovation in educational philosophy, teaching methods and management modes, and is expected to lead systematic changes in education. In fact, the development of artificial intelligence technology cannot be separated from the continuous research of computer vision technology. How to make the computer obtain external information through video acquisition device to simulate human for analysis and recognition is the main research content of computer vision technology, and its research direction includes target tracking, object recognition, human behavior recognition, etc. In this paper, through the combination of artificial intelligence technology and computer vision technology, we analyze the video images of students' classes in classroom scenes, identify students' classroom behavior status, build a judgment system on students' classroom learning status, and assist teachers to obtain teaching feedback information.

Keywords: Education, Classroom behavioral states, Video images, Computer vision technology, Instructional feedback.

1. Research Preparation

1.1. Background of the study

Artificial intelligence is an important driving force leading the new round of technological revolution and industrial change, which is profoundly changing the way people produce, live and learn, and pushing human society to usher in an intelligent era of human-machine collaboration, cross-border integration and co-creation and sharing. It is an important mission of education to grasp the global AI development trend, find the breakthrough and main direction, and cultivate a large number of AI high-end talents with innovation ability and cooperation spirit. Artificial intelligence technology is not widely used in the field of education, but it does not mean that artificial intelligence and computer vision technology have a lack of applicability in education. On the contrary, with the rapid development of information digitization nowadays, the combined technology of the two has a broad development prospect in the field of education. The traditional classroom is a time-consuming and labor-intensive way to get feedback on teaching and learning, with teachers learning about students' learning through after-class assignments on the one hand, and teachers getting feedback from students' listening in class on the other. However, teachers have to balance the quality of the content taught in class and monitor students' attentiveness, which inevitably leads to a decline in teaching quality, and teachers are unable to keep an eye on each student's status in class. By combining artificial intelligence with computer vision technology, recording students' class videos, recognizing students' actions in classroom scenes and recording them at the same time, we can automatically record and analyze the classroom students' listening status, so that teachers can understand the overall listening situation of the class more intuitively and with less time and effort, which is an important reference value and significance to assist teachers in improving education and teaching methods. The deep integration of artificial intelligence and educational scenes can not only reverse the current problem of uneven distribution of educational resources to a certain extent, but also improve the effectiveness of classroom teaching and learning efficiency, as well as transform and reconstruct the traditional education form and education mode, and promote the early realization of "teaching according to ability" that has been pursued for more than two thousand years.

1.2. Current status of research

In recent years, the combination of artificial intelligence technology and computer vision technology has achieved many practical applications and has been put into use in the field of human behavior recognition, such as the 3D human motion simulation and video analysis system for sports training developed by Xia Shihong [1] of the Chinese Academy of Sciences. Human behavior recognition is one of the research hotspots in computer vision, which aims to analyze and understand the behavior of individuals in the video and the interaction behavior between multiple individuals. Human behavior recognition technology has a wide range of applications in security surveillance systems [2], medical diagnosis and monitoring [3], and human-computer interaction; in classroom behavior, in 2014, Hai Zhou Cai combined the valuable behavioral layer feature information and the human behavior knowledge lexicon of behavioral features through spatio-temporal interest points and visual bag-of-words model for feature training learning to generate human behavior patterns. The algorithm was applied to classroom behavior recognition and got better results [4].In 2015 Yang Yuanbo proposed an AdaBoost face detection algorithm incorporating skin color information to detect and recognize students' classroom behavior by using long-term video observation recordings for analysis [5].In 2017 Dongli Dang extracted three features of students' actions, including Zernike moment feature, optical flow features, and global motion direction features, and then combined with a plain Bayesian classifier [6] to recognize students' hand raising, standing, and sitting behaviors [7]. 2018 Pengnian Zhou et al. obtained data from three aspects: face data, contour features, and subject action magnitude, and used a Bayesian causal net model to infer subject behavior features, and then recognized student behaviors in classroom teaching videos [8]. In 2018, Peng Liao et al. made the dataset required for the experiment by classifying student classroom behavior patterns into three categories: listening, sleeping, and playing with cell phones, and the image material was preprocessed, and the system used Matcovnet third-party tools to fine-tune the parameters based on the ImageNet [9] pre-trained network model, and migration learning was performed by modifying the structure of the VGG network [10] based on the pre-trained In 2018, Bin Tan and Shuyi Yang used the target detection network Faster R-CNN [12] based on the ZFNet pre-trained network model for migratory learning to extract the characteristics of students' classroom behaviors and achieve the detection and recognition analysis of classroom behaviors such as studying, sleeping, and playing with cell phones, and the experiments achieved good detection results [13]. good detection results [13].2019 Qin Daoying detects student positions in images by deep learning model Yolov3 [14], and trains and recognizes seven classroom behaviors such as reading, sleeping, raising hands, writing, listening, standing, and looking left and right using Resnet50 [15] network [16].2019 Gong Wei uses OpenPose skeletal In 2019, Gong Wei used the OpenPose skeletal keypoint detection model to detect the keypoints of five classroom behaviors, such as raising hands, stretching, lying down, playing with cell phone, and writing, and used the direct skeletal keypoint coordinate method for raising hands and stretching, while the other three behaviors used the skeletal keypoint relationship feature method for vector extraction, and then the feature vector was trained by support vector machine for classification [17]. 2019, Ji Chongxiao used the image processing method for In 2019, Qinyi Jiang et al. proposed a deep residual network based on residual structure to build a dataset of student classroom behavior recognition including class, sleeping, playing with cell phone, taking notes, reading and In 2020, Canran Lin et al. fused human key points with RGB image information and extracted features to achieve the recognition of student behaviors [20]. In terms of classroom face detection and attention research, in 2018, J. Zhang designed three convolutional neural networks for face detection based on the Caffe framework based on the idea of Adaboost cascade algorithm as Det-A, Det-B and Det-C respectively, and achieved 92.9% recall on FDDB dataset, and then designed a prototype convolutional neural network called HeadNet head-up recognition network was then designed as a prototype of convolutional neural network, while the classroom headcount statistic was designed based on the fusion of multi-frame face location information, and the whole system achieved 92.3% recall and 94.6% accuracy on the ClassHead dataset [21]. 2019 Zuo Guocai et al. used the StackedDenoising Auto-encoders (SDA) model Auto-encoders (SDAE) to learn generalized image features from large-scale image databases aidedly, construct a deep feature extractor to extract face features, add a classifier layer on top of the general image feature extractor to construct a supervised deep learning model, and then train face images online to extract fine-tune the previously learned generalized features to complete the face recognition task. In turn, the eyes and nose of the detected target are recognized, and the eye and nose parts of the face are identified with rectangular boxes of different colors to obtain the rectangular coordinates of the eyes as well as the length and width of the rectangle, and through the localization and recognition of the face and the eyes, the frontal and lateral faces can be recognized to make a comprehensive evaluation of the students' concentration in the classroom [3].In 2019, Jia Oriyu et al. used the Yolo algorithm [22] for initial target detection of students, then use the C++ library with DLIB library for feature extraction of key points of faces, and finally classify and recognize facial expressions by SVM (Support Vector Machine) algorithm [23] [24].In 2019, Shuangxi Zhang used a dense face detection method fusing multiple deep neural networks to detect dense face detection and improved the accuracy, added SVM to FaceNet [25] face recognition method to extract facial features, and finally used VGG19 [10] and Resnet18 [15] to construct face fatigue classifier to discriminate facial concentration [26].2019 Xiaoxu Guo used an integrated deep learning framework based on facial expression recognition FATAUVA- Net [27] for classroom micro-expression recognition, combining 3D learning state space and emotional dimensional theory to achieve an effective classroom state classification method [28]. 2020 Furong Li improved MedianFlow [30] based face tracking algorithm using MTCNN [29] [31], and proposed a new CNN-based face key point detection model that The information of eye closure time, blink frequency and head position were integrated to achieve the detection of student fatigue, and 92% detection accuracy was achieved.

2. Research Process

2.1. Objectives and content of the study

2.1.1. Research objectives.

The research goal of this paper is to objectively detect and identify the learning status of students in the classroom by means of computer vision and machine learning classification to provide objective assessment tools for teachers to improve their teaching and enhance student learning. The specific research objectives of this paper are as follows.

(1) The production of the original dataset, standard action data is needed in action recognition to train the classification model, and the action video dataset in the classroom scene lacks relevant open source data. Therefore, the primary goal of this paper is the production of standard datasets.

(2) Action recognition method combining skeletal

keypoints based on machine learning algorithm and keypoints grayscale heat map based on deep learning migration, by selecting suitable human skeletal keypoints detection methods, comparing and analyzing various machine learning classification algorithms to get the most accurate method for classification results, and also studying the keypoints grayscale heat map based on deep learning migration for the case of incomplete skeletal keypoints detection recognition method, and fusing two classification methods for action recognition.

2.1.2. Study content

This paper proposes a method of classroom action recognition based on a combination of machine learning-based skeletal keypoints and deep learning migration-based grayscale heat map of keypoints, and combines software interface development, database and other technologies to realize the algorithm for software systematization implementation. The research of this paper is as follows.

(1) Based on the research of the related technology of skeletal key point detection, we study and compare different methods of skeletal key point detection, and select a more efficient framework as the method of extracting the coordinates of key skeletal points in this paper.

(2) Based on the study of machine learning classification methods, we study various types of machine learning algorithms and compare the performance of various types of machine learning algorithms through the feature values of action data constructed in this paper.

(3) Research comparison of action classification methods based on migration learning, research comparison of deep learning image classification recognition methods, application of migration learning, application of the current mainstream deep learning classification methods to the dataset constructed in this paper, comparison of random forest and the support vector machine based skeletal key point action classification method proposed in this paper.

(4) Study of action classification method based on deep learning migration of key point heat map, and comparison with machine learning based skeletal key point action classification method, and finally combining the two for joint recognition of actions.

2.2. Characteristics and Difficulties of Research on Classroom Student Behavior States

2.2.1. Characteristics of the study of classroom student behavioral states.

The classroom student behavior state refers to the learning state of students in the classroom, mainly through the behavior state of students to discriminate, before discriminating the behavior state, first need to detect the location of the head of students in the classroom, and then to classify it to discriminate, the detection of the head is not simply equivalent to face detection, face detection is only the detection of the obvious features of the face area, while the head detection does not focus on the face area The head detection does not focus on the size of the face, but only needs to detect whether the part is the head, which is the whole head as a target.

2.2.2. Research Difficulties in Classroom Student Behavior States.

At present, the main difficulties in the study of classroom behavioral states are due to the different entry points of the study and the different characteristics and factors that exist in each.

(1) Establishment of classroom student learning state datasets. The basic work for the study of student learning states, considering the scenario in the classroom classroom, is to establish a learning state dataset, and to clarify what kind of dataset needs to be constructed in order to make the subsequent experiments more feasible. By reviewing the literature and data, we found that there is little information on the direction of classroom student learning state research and no publicly available data sets, so we need to collect a large amount of data and build the corresponding experimental classroom student learning state database accordingly.

(2) The classroom student learning status database is divided into classroom student head data set and classroom student learning status data set, both the annotation of the student head data and the subsequent cutting of the learning status data need to be done manually, which is a great workload. All of these have brought a great workload to the study of classroom students' learning status.

3. Summary and Outlook

3.1. Summary of research content

This paper investigates the problems related to classroom behavior analysis system, and proposes a method of classroom action recognition by combining machine learning based skeletal keypoints with deep learning migration based grayscale heat map of keypoints, which then constitutes a video-based classroom behavior analysis system. The whole research includes the construction of the original dataset, the study of human posture recognition based on skeletal keypoints, the study of classification methods based on machine learning, and to combine different classification methods, this paper uses the convolutional neural network based on migration learning to train the network model on the constructed dataset in this paper. The proposed deep learning migration-based key point heat map recognition method is integrated with the skeletal key point classification method. Finally, the video-based classroom behavior analysis system will be finalized through the interface development framework and overall system design. The main work planned is as follows.

(1) Original dataset construction: After determining the research direction, this paper needs relevant standard datasets for the training of classification models. Since there is no real and publicly available dataset for student action videos in classroom scenes, this paper uses video recording equipment to record student behavior videos in real classroom scenes, and then uses video editing software to create and classify the standard dataset, which is divided into five types of actions: hand raising, cheek resting, lying down, playing with cell phone, and writing. The sixth category was added to the system.

(2) Research on the action classification algorithm of skeletal key points based on machine learning: In order to achieve action classification, this paper firstly researches and compares the existing key point detection methods, selects the more efficient and accurate OpenPose human posture recognition model as the skeletal key point detection framework, studies the network construction process of OpenPose in detail, conducts a detailed research on the clustering problem of the detected key points, and finally

calculates the limb vector field and uses the Hungarian algorithm to find the connection method of multiple skeletal key points by combining the key point connection weights to form a multi-person posture recognition map. Finally, we calculate the limb vector field and use the Hungarian algorithm to find the connection method of multiple skeletal key points by combining the key point connection weights to form a multi-person pose recognition map. Then, the key skeletal points are normalized by the min-max dimensionless normalization method through the screening of key skeletal points to reduce the influence caused by individual differences and the distance of the human body from the acquisition device, and the construction of action feature values is completed. Finally, by comparing various machine learning classification algorithms, we finally adopt the most efficient way to train the classification model.

(3) Research on action recognition methods based on migration learning: In order to compare the methods used in this study, research experiments related to the migration of deep learning networks are conducted, firstly, the existing deep learning infrastructure network structure is studied and described, and then the current mainstream convolutional neural networks are introduced, and the Res Net50 network is applied on top of the migration learning methods to train the classification model on the data set constructed in this paper. The study involves the selection of loss functions, optimization algorithms and parameter tuning. Finally, a comparison and analysis with the method used in this paper is performed.

(4) Research on joint recognition method based on machine learning skeletal key points and migration learning based key point grayscale heat map: In the study of machine learning based skeletal key point action recognition method, it is found that when there is incomplete recognition of skeletal key points will cause poor recognition effect or even recognition algorithm failure, based on this, a migration learning based key point grayscale heat map recognition method is proposed, and the experiment proves that the method can still perform effective recognition in the fuzzy state. It is proved that the method can still perform effective recognition in the blurred state, and finally the two methods are fused.

3.2. Research Outlook

Classroom behavior state analysis is a meaningful and quite difficult research topic. In this paper, we have achieved the recognition of some classroom actions through a skeletal key point detection framework and a machine learning classification model, but there is still much room for improvement in the classification direction and accuracy of this topic in the future. The main aspects are as follows.

(1) Application of fusion algorithm in multi-person recognition: Since the heat map recognition method in the fusion algorithm needs to intercept each single person in the video in the process of practical application and then carry out the process of heat map composition recognition, the single person data set in this paper are manually intercepted composition, for the video data in the classroom multi-person scenario, the automatic tracking interception function of each individual in the video needs to be further studied to achieve In this paper, the single-person dataset is manually intercepted.

(2) Improving accuracy: Because there are many random factors and background environment, occlusion and other problems of human action, the recognition method in this paper has an uncertain impact on the accuracy of human action classification in the presence of large area occlusion and similar action characteristics, plus in the classroom scene, there are also cases where the head is partially occluded due to more targets and smaller size, so for small targets and partially occluded head how to achieve better detection in the future can have further research enhancements in predicting the estimation of obscured human pose and the classification of similar actions.

(3) Adding action categories: This paper only classifies and identifies the hand raising, lying down, cheek resting, playing with cell phone, writing and the last added sitting action, more iconic actions can be added in the subsequent study, such as head up listening to lecture, head up fidgeting, head down sleeping, head down taking notes, left looking and right looking, etc. to improve the systematic action classification categories.

(4) Multi-feature fusion: This paper mainly focuses on the recognition and analysis of classroom behavioral actions, which are less studied at present. In the future, further multi-feature fusion methods can be combined with more features of human emotion expressions such as facial expressions and semantic analysis to analyze classroom teaching feedback and better achieve the purpose of assisting teachers to improve teaching.

References

[1] Xia Shihong. Three-dimensional human motion simulation and video analysis system for sports training [J]. High Technology and Industrialization, 2008(09): 28.

[2] Ma Yuxi, Tan Li, Dong Xu, Yu Chongzhong. Behavior recognition for intelligent surveillance [J]. Chinese Journal of Graphic Graphics, 2019, 24(02): 282-290.

[3] Lu, Zhujun, Li, Jile, Liang, Satellite, Chen, Pingping. Research on 3D medical image assisted diagnosis and treatment system based on Kinect[J]. Software Guide, 2015,14(01):143-145.

[4] Cai Hai Zhou. Research and application of classroom human behavior recognition based on spatio-temporal interest points [D]. Master's thesis (Nanjing Normal University), 2014.

[5] Yang Yuanbo. Key technology research on student classroom behavior video observation recording system [D]. Master's thesis (National University of Defense Technology), 2015.

[6] Rish I. An empirical study of the naive Bayes classifier [J]. Journal of UniversalComputer Science, 2001, 1(2):127.

[7] Dongli Dang. Human behavior recognition and its application in educational recording system [D]. Master's thesis (Xi'an University of Science and Technology), 2017.

[8] Zhou Pengxiao, Deng Wei, Guo Cultivation, et al. Research on intelligent recognition of S-T behaviors in classroom teaching videos[J]. Modern Educational Technology, 2018, 028(006):54-59.

[9] Jia D, Wei D, Socher R, et al. ImageNet: A large-scale hierarchical imagedatabase[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition,Miami, 20-25 June 2009, 248-255.

[10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-ScaleImage Reconginitional[J]. ar Xiv preprint ar Xiv:1409.1556, 2015:1-14.[11] Liao P, Liu C M, Su H, et al. A deep learning-based system for detecting and analyzing students' abnormal classroom behaviors[J]. E-World, 2018, 000(008):97-98.

[11] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time ObjectDetection with Region Proposal Networks [J].

IEEE Transactions on PatternAnalysis and Machine Intelligence, 2017, 39(6):1137-1149.

[12] Tan B, Yang S. Enthalpy. Research on student classroom behavior detection algorithm based on FasterR-CNN [J]. Modern Computer:Professional Edition, 2018(33):47-49.

[13] Redmon J, Farhadi A. Yolov3: An Incremental Improvement [J]. ar Xiv preprintar Xiv:1804.02767, 2018:1-6.

[14] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C].2016 IEEE Conference on Computer Vision and Pattern Recognition, LasVegas,27-30 June 2016, 770-778.

[15] Qin Daoying. Student classroom behavior recognition based on deep learning [D]. Master's thesis (Huazhong Normal University), 2019.

[16] Gong Wei. Design and implementation of a student learning behavior recognition system based on skeletal key point detection [D]. Master's thesis (Jilin University), 2019.

[17] Ji Chongxiao. Research on classroom behavior recognition method based on digital image processing [D]. Master's thesis (Taiyuan University of Technology), 2019.

[18] Qinyi Jiang, Yewen Zhang, Siqi Tan, et al. Student classroom behavior identification based on residual network [J]. Modern Computer (Professional Edition), 2019, 000(020):23-27. 2021 Master's Degree Thesis, Guizhou University for Nationalities

[19] Lin Chan-Ran, Xu Wei-Liang, Li Yi. An investigation of classroom student behavior recognition techniques based on multimodal data [J]. Modern Computing, 2020, 678(06):70-76.

[20] Zhang J. Research and implementation of a head-up rate detection system for university classrooms [D]. Master's thesis (HuaUniversity of Science and Technology of China), 2018.

[21] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-TimeObject Detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016, 779-788.

[22] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines[J]. IEEE Transactions on Netural Networks, 2002, 13(2):415-425.

[23] Jia, Oriyo, Zhang, Zhao, Zhao, Xiaoyan, et al. Classroom student state analysis based on artificial intelligence video processing[J]. Modern Educational Technology, 2019, 224(12):83-89.

[24] F Schroff, Kalenichenko D, Philbin J . FaceNet: A Unified Embedding for Face Recognition and Clustering[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, 7-12 June 2015, 815-823.

[25] Zhang Shuangxi. Research and application of face recognition and concentration discrimination method based on deep learning [D]. Master's Degree Dissertation (Nanjing Normal University), 2019.

[26] Chang W Y, Hsu S H, Chien J H. FATAUVA-Net: An integrated deep learning [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, 21-26 July 2017, 1963-1971.

[27] Guo Xiaoxu. Research on a classroom concentration analysis system based on micro-expression recognition for students [D]. Master's thesis. (Yunnan Normal University), 2019.

[28] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.

[29] Kalal Z, Mikolajczyk K, Matas J. Forward-backward error: automatic detection of tracking failures[C]. 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23-26 August 2010, 2756-2759.

[30] Li Furong. Multi-feature fusion student fatigue detection based on MTCNN [J]. Information Technology, 2020,044(006):108-113, 120.