Journal of Engineering Science and Technology
Special Issue on ICIT2022 Vol. 18, No. 6 (2023) 39 - 47

FACE RECOGNITION TECHNOLOGY BASED ON DEEP LEARNING ALGORITHM FOR SMART CLASSROOM USAGE

YONGHONG WANG, WOU ONN CHOO*, XIAOFENG WANG

Faculty of Data Science and Information Technology,
INTI International University, Nilai, Malaysia
*Corresponding Author: wouonn.choo@newinti.edu.my

Abstract

As enrolment in higher education increases year by year, there is a serious imbalance between the number of faculty and the number of students, and teachers are unable to attend to the learning of all students while ensuring the quality of instruction. In this paper, we study the learning of students aged 18 to 23 and propose an intelligent education system based on face recognition. The system is based on improved MTCNN and FaceNet models, and the experimental results show that the system achieves 98% accuracy in face recognition and 92% accuracy in student emotion recognition. The proposed method can effectively improve the efficiency of classroom check-in, monitor the teaching process, and manage the teaching effect.

Keywords: Deep learning, Education quality, Educational environment, Face recognition, Feature recognition, Smart classroom.

1. Introduction

With the development of artificial intelligence (AI), the application of AI in various fields is getting deeper day by day. Face recognition, as an important research direction of AI, has started to be widely used in various fields, such as phone unlocking, security, robotics, education, and so on [1]. In recent years, China has seen a large growth in both economy and education, with the number of students enrolled in college entrance exams increasing year by year (as shown in Fig. 1) and the number of students in school classrooms increasing. Therefore, teachers are unable to attend to the learning of all students while ensuring the quality of instruction [2].

Translated with www.DeepL.com/Translator (free version)

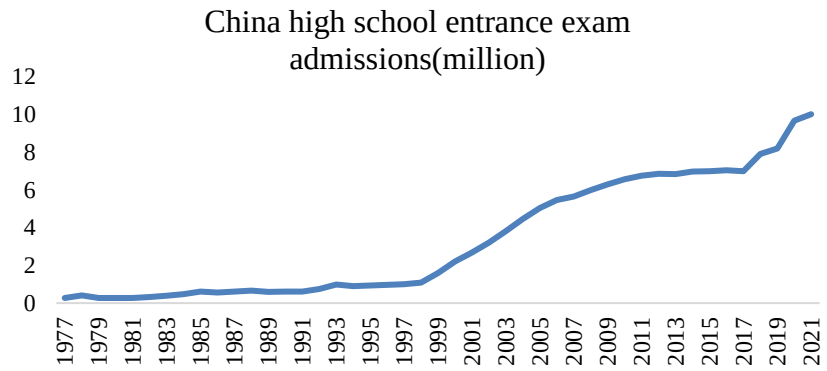

Fig. 1. The number of China high school entrance exam admissions.

AI is a machine-based technology. It can make predictions, diagnoses, suggestions, and decisions. AI has gained importance in education in recent years because of its potential to support learning in different contexts [3]. Artificial intelligence in education (AIEd) has shown technological advances, theoretical innovations, and successful pedagogical effects [4]. AI can provide specialized support to improve knowledge gap awareness and enable instructors to teach effectively through personalized and adaptive instruction [5]. AI also provides algorithm-based decision-making that enables effective real-time assessment of complex skills and knowledge [6]. In addition, AIEd can analyse classroom status and student attention, which helps to identify at-risk students in a real-time mode, allowing for timely intervention [7].

In recent years, face check-in technology has become more and more popular in various industries because of its convenience and not require manual check-in. Therefore, applying face check-in technology to the education field to assist teachers in educating students and combining education management and artificial intelligence in both directions can help improve classroom quality, assist parents and schools in monitoring the quality of students' education in real-time, and escort students to class [8].

Therefore, applying AI technology to teaching and learning activities in higher education can help teachers improve attendance in the classroom and check the

learning status of students. By analysing this collected data, teachers can get a more comprehensive picture of their teaching effectiveness.

In this paper, we propose a new solution for the next stage of smart classroom creation based on face recognition technology. In the smart classroom, AI technologies such as face detection, face recognition, and emotion recognition are applied to higher education teaching to achieve sensorless check-in and continuously monitor students' learning attention in the smart classroom.

2.Methodology

In this section, we propose a three-stage approach for classroom attendance as well as classroom status analysis as the method of this study, including MTCNN face detection, FaceNet face recognition, and convolutional neural network emotion recognition. The diagram of the organization structure is shown in Fig. 2.

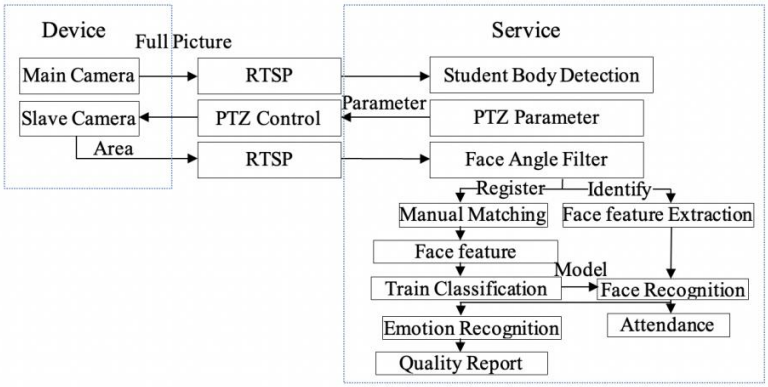

Fig. 2. The diagram of organization structure.

Phase 1: MTCNN face detection

In the face detection section, the system uses the MTCNN algorithm for face detection because of the high accuracy and speed required for multi-person face recognition [10]. The MTCNN algorithm proposes a deep cascade multitasking framework that exploits the intrinsic correlation between face detection and alignment to improve performance [11]. The model employs three deep convolutional neural networks to predict the coordinates of the face and feature points by using a coarse-to-fine approach. These three cascaded networks are the Proposal Network (P-Net) for fast candidate window generation, the Refinement Network (R-Net) for high-precision candidate window filtering selection, and the Output Network (O-Net) for generating the final wraparound box with key points of the face [12]. People widely use the MTCNN algorithm in face detection because it has high accuracy and fast detection speed.

In smart classrooms, the distance between people and cameras is usually not equal, and the size of faces on the cameras is different. If it does not change the images, it will cause some smaller faces in the images not to be detected. Therefore, before using the MTCNN algorithm for face detection, an image pyramid is generated to resize the pictures. MTCNN will set a scaling parameter factor for each scaling and perform a power operation with the factor and the number of

images. Therefore, before using the MTCNN algorithm for face detection, a picture pyramid is generated to resize the pictures. MTCNN will set a scaling parameter factor for each scaling, perform a power operation with the factor and the number of images to be scaled, and set the minimum size of the detected faces and the minimum scaling size of the images to prevent the features from overlapping and not being computed because the images are too small, as shown in Eq. (1).

$$nextSize = originsize \times \left(\frac{12}{minsize}\right) \times factor^{n}, n = \{1,2,3,4,\ldots,n\} \quad (1)$$

In which, $originsize$ is the image size before scaling, the factor is the scaling ratio, default is 0.79, n is the image to be scaled, $minsize$ is the minimum size of the detected face, default is 20, smaller $minsize$ means more comparisons are needed and longer time, considering the time requirement of the system and this paper is based on classroom scene, therefore, this system is based on The system optimizes $minsize$ and initializes $minsize$ according to the classroom scene, as shown in Fig. 3. The size of the face is calculated according to the coordinates, and the size is used as $minsize$.

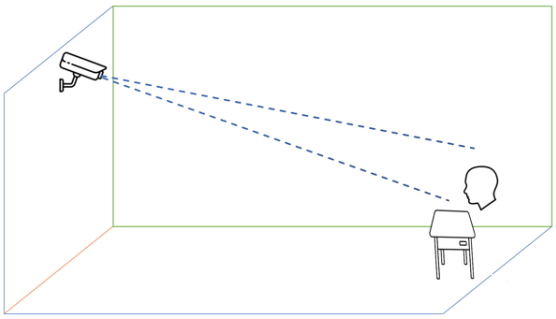

Fig. 3. The diagram of scene.

The next step is to feed the data into P-Net, which is a fully convolutional neural network that contains five convolutional layers, and a pooling layer (Pool), and each stage uses the PRELU activation function. P-Net is performed in two steps. Conv-1 is used to perform face classification and get the probability value of being a face, while Conv-2 is used to perform face border localization and use border regression with non-maximal suppression for face border filtering after getting face borders. The results derived from the R-Net stage are fed into O-Net for edge correction and key point localization. O-Net adds one more convolutional layer on top of R-Net to finally input the key point information of the face and the recognized face.

Since the scene of this system is based in a classroom, to ensure the accuracy of the system under influencing factors such as low head, light, and occlusion, the scene selected for the dataset of this system in the face detection stage is a randomly shot video of college students in class, with each class shot for the 30s and video subframe, totalling 18,000 photos.

Phase 2: FaceNet face recognition

In the face recognition stage, this system uses FaceNet for face recognition, to ensure the accuracy rate and reduce the training time, the system uses the LFW

dataset to pre-train the algorithm before using its dataset and uses the pre-trained model, which is loaded into its dataset for retraining [13]. The test data is based on the set of face images obtained from the face detection phase, as shown in Fig. 3. The first convolutional layer of this network, with 3 pads and 64 features, has a 7×7 step size of 2. The output features are 112×112×64 and then ReLU, followed by pooling 3×3 kernels using maximum pooling with a step size of 2. After the pooling layer finishes sampling, the data is normalized and fed into the Inception layer, as shown in Fig. 4. Inception uses 1×1, 3×3, and 5×5 convolutional kernels built in parallel and pooled using the 3×3 pooling layer, which greatly simplifies the computational work.

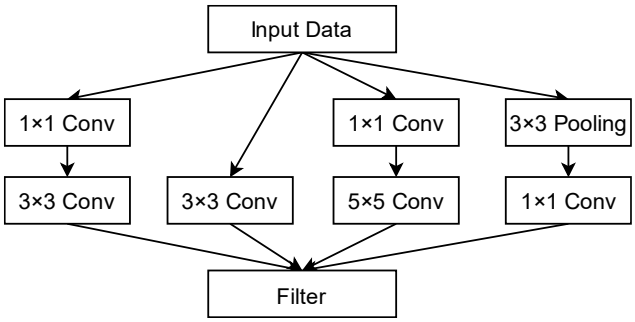

Fig. 4. The diagram of inception.

After the computation is completed, the normalization is performed with L2, which is $f(x)^2=1$, to map all image features to a hypersphere; then an embedding layer (embedding function) is accessed, and the embedding process can be expressed as a function $f(x) \in R^d$, the image x is mapped to a dimensional d Euclidean space by the function f. Finally, triple loss is used as the loss function to optimize the features, and the objective function to be optimized is given in Eq. (2), with a being the difference between the two class spacings. The function optimization is performed using mini-batch gradient descent to update the weights of FaceNet in reverse until the error converges.

$$L = \sum_i^N \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + a \right] \quad (2)$$

The dataset of the model pre-training phase used in the face recognition phase is the LFW dataset, which is used to pre-train the model to improve the accuracy; the dataset of the model training phase is about 54,000 face information obtained by cropping the faces detected in the face detection phase, of which the training set is about 43,200 images, and the validation set, and test set are about 10,800 images.

Phase 3: Convolutional neural network emotion recognition.

The dataset is the face dataset got from the first phase based on face detection. The system uses the CNN algorithm to analyse the students' emotions, and the algorithm is experimentally validated to ensure that it can be used for real-time emotion recognition in the system [14]. The convolutional neural network designed for the system consists of 2 convolutional layers, 2 pooling layers, and 2 fully connected layers, as shown in Fig. 5.

The face image detected by MTCNN is greyed out and cropped into 48×48×1 and input into Conv1, which has 3×3 convolutional kernels with 16 kernels and a step size of 1. The padding is 0, and the activation function is ReLU, the next step is input into the maximization pooling layer, which has 2×2 convolutional kernels with a step size of 2. The next step is input into the maximization pooling layer, where the size of the convolutional kernel is 2×2, the step size is 2, and the padding is processed in the same way as the convolutional layer.

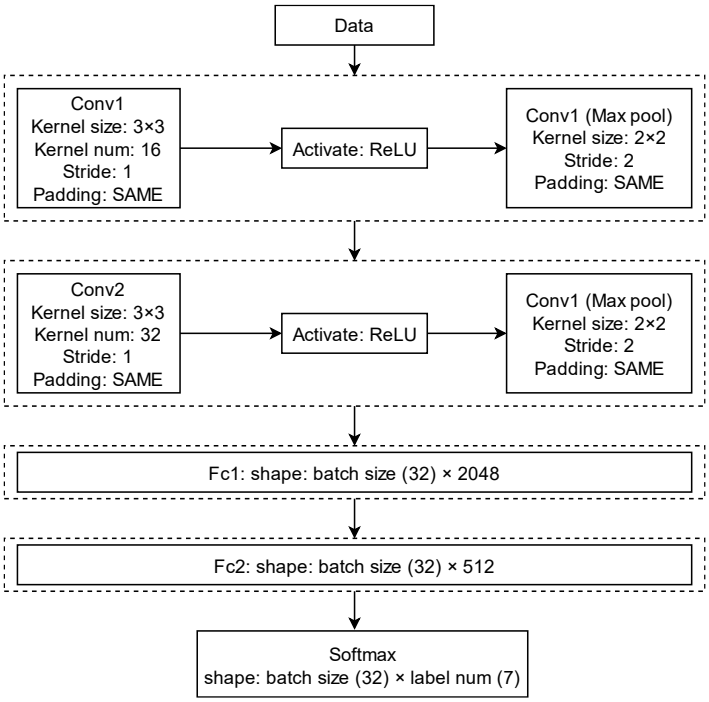

Fig. 5. The diagram of CNN.

After the first layer of feature extraction, it is sent to the second step for further feature extraction. The size of the convolutional kernel in the convolutional layer is 3×3, and the number of convolutional kernels becomes 32 to facilitate the extraction of more detailed features, and padding is added to 0. The size of the convolutional kernel in the pooling layer is 2×2, and the step size is set to 2.

After two layers of feature extraction, there are two full-connected layers. The first full-connected layer contains 2048 nodes and is partitioned according to batch_size; the second full-connected layer contains 512 nodes.

Finally, the regression classification is performed using softmax and the nodes are changed to 7 in this layer, i.e., the category of classification. softmax is specifically calculated in Eq. (3).

$$L = \frac{1}{B}\sum_i - log\left(\frac{f_{x_j - max(fx)}}{\sum_N f_{x_j - max(fx)}}\right) + \lambda W \quad (3)$$

In loss optimization, the batch gradient descent method was used, the learning rate was set to 0.0001, and each batch sample was set to 32.

3. Results and Discussion

In the face detection stage, the system uses the MTCNN deep learning algorithm, and the algorithm is optimized and adjusted for the classroom scene with insufficient light, long distances, and filling-in occlusion. 18,000 images were input into the algorithm model for deep learning, and after testing, the number of faces recognized in each video was recognizable except for students with their heads down. As shown in Fig. 6, the number of faces recognized in this picture is 14, the recognition time is 15ms, and the faces can be recognized more completely even when the students are wearing masks.

In the second stage, the FaceNet face recognition algorithm was used to recognize the faces collected in the previous stage, and the system used 43,200 images as the training set for training to get the training model. In addition, 10,800 images were used as the test set for testing. By randomly selecting 6 samples from the test set for viewing (as shown in Table 1), the accuracy rate of face recognition is over 98%.

In the test phase of the emotion recognition result, the experiment will input the test set divided in advance according to name into the model for testing, and the result will be labeled in the picture for storage, and the accuracy of the test will be output the total number of the test set is 2,000, the number of wrong pictures is 164, and the accuracy is 92%, the test result schematic is shown in Fig. 7.

Fig. 6. The diagram of face detection.

Fig. 7. The diagram of emotion detection.

Table 1. Test results of 6 randomly sampled students.

Student name	Number of images	Number of unrecognized images	Accuracy rate
Hui Li	273	4	98.5%
Jianhua Yao	284	6	97.9%
Yuan Fang	276	3	98.9%
Yida Li	271	2	99.3%
Juanhua Qi	203	1	99.5%
Ming Wang	189	2	98.9%

4. Conclusion

We propose an intelligent education system based on face recognition, which is an important application of combining artificial intelligence with education. The system uses face recognition and emotion recognition to assist teachers in managing students. On the one hand, it can assist teachers in monitoring students' classes in real-time and giving reminders, and it can analyse students' classes and propose reference solutions for teachers to make targeted classroom improvements based on students' situations. The system realizes the sensorless check-in of students using face recognition, and the established improved MTCNN and FaceNet models have an accuracy of 98% for face recognition, which can check-in students entering the classroom within 2s.

Acknowledgments

The authors would like to express our gratitude to the Faculty of Data Science and Information Technology at INTI International University in Nilai, Malaysia, for their support and assistance.

References

1. Li, T. (2021). Research on intelligent classroom attendance management based on feature recognition. *Journal of Ambient Intelligence and Humanized Computing*, 13(4), 1-8.
2. Madyatmadja, E.D.; Noverya, N.A.R.; and Surbakti, A.B. (2021). Feature and application in smart campus: a systematic literature review. *Proceedings of the 2021 International Conference on Information Management and Technology (ICIMTech)*. Jakarta, Indonesia, 358-363.
3. Hwang, G.-J.; Xie, H.; Wah, B.W.; and Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001.
4. Roll, I.; and Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599.
5. Guan, C.; Mou, J., and Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4), 134-147.
6. Chen, X.; Zou, D.; Xie, H.; Cheng, G.; and Liu, C. (2022). Two decades of artificial intelligence in education: contributors, collaborations, research topics,

challenges, and future directions. *Educational Technology and Society*, 25(1), 28-47.

7. Tsai, S.-C.; Chen, C.-H.; Shiao, Y.-T.; Ciou, J.-S.; and Wu, T.-N. (2020). Precision education with statistical learning and deep learning: a case study in Taiwan. *International Journal of Educational Technology in Higher Education*, 17(1), 12.
8. Faritha Banu, J.; Revathi, R.; Suganya, M.; and Gladiss Merlin, N.R. (2020). IoT based cloud integrated smart classroom for smart and a sustainable campus. *Procedia Computer Science*, 172, 77-81.
9. Razzaq, S.; Shah, B.; Iqbal, F.; Ilyas, M.; Maqbool, F.; and Rocha, A. (2022). DeepClassRooms: a deep learning based digital twin framework for on-campus class rooms. *Neural Computing and Applications*, 35(11), 8017-8026
10. Tong, Y.; Ma, H.; Zhang, S.; Wu, X.; and Chen, W. (2022). Research on object detection in campus scene based on faster R-CNN. *Journal of Physics: Conference Series*, 2203, 012050.
11. Gao, Z.; Huang, Y.; Zheng, L.; Li, X.; Lu, H., Zhang, J.; Zhao; Q.; Diao, W.; Fang, Q.; and Fang, J. (2021). A student attendance management method based on crowdsensing in classroom environment. *IEEE Access*, 9, 31481-31492.
12. Wu, H.; Pan, Y.; Weng, X.; and Chen, H. (2021). Design of campus health information system using face recognition and body temperature detection. *Proceedings of the 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech*). AB, Canada, 873-878.
13. Muhammad, W.; Ahmed, I.; Ahmad, J.; Nawaz, M.; Alabdulkreem, E.; and Ghadi, Y. (2022). A video summarization framework based on activity attention modeling using deep features for smart campus surveillance system. *PeerJ Computer Science*, 25(8), e911.
14. Sun, X.; Ning, Y.; and Yang, D. (2021). Research on the application of deep learning in campus security monitoring system. *Journal of Physics: Conference Series*, 1744, 042035.