Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 2345–2357, DOI 10.12694/scpe.v25i4.2811

APPLICATION OF CNN BASED REGOGNITION TECHNOLOGY IN PUBLIC ENGLISH TEACHING IN COLLEGE INTELLIGENT CLASSROOM

SHUHUI LI*

Abstract. With the advancement of intelligent technology, CNN-based recognition technology has been integrated into public English teaching in universities. This implementation contributes to enhancing teaching quality, nurturing students' English proficiency, and holds significant educational and practical value. To address the issue of low traditional attendance efficiency in intelligent classrooms for public English teaching, a face recognition model based on CNN recognition technology has been developed. R-CNN is utilized for object detection, along with pyramid pooling and non-maximum suppression to acquire the optimal candidate region for face detection. Furthermore, K-Means clustering is combined to enhance Fast R-CNN, thereby improving detection accuracy. Experimental results demonstrated that among the three networks—Fast R-CNN, Faster R-CNN, and CNN—Faster R-CNN maintained a high recognition rate and exhibited faster convergence speed, showcasing superior overall performance. Specifically, at 500 iterations, the three networks require 23.7 seconds, 26.8 seconds, and 34.2 seconds, respectively. For facial expression recognition, Faster R-CNN achieved the highest recognition rate, indicating its exceptional detection efficiency and potential for aiding teaching management. This study offers novel technical support for public English teaching in intelligent university classrooms, effectively enhancing teaching efficacy and learning experiences. Its practical significance extends to promoting educational reform and improvement.

Key words: CNN; Faster R-CNN; Face recognition; Public English teaching; Class attendance

1. Introduction.

With the rapid development of technology, artificial intelligence technology has gradually penetrated into the field of education. In recent years, with the implementation of the enrollment expansion policy, the scale of universities has become increasingly large, and there are more and more college students. Teaching management has become a major challenge [1]. As an important basic education course, public English teaching in universities has always been the focus of attention for education workers in terms of teaching effectiveness and learning experience. Intelligent classroom public English teaching in universities is an important aspect of teaching. In order to strengthen its teaching management level and improve the teaching environment, attention should be paid to classroom attendance to facilitate the smooth implementation of teaching. Classroom attendance is an important link in teaching and is generally used as one of the criteria for teachers to evaluate students' grades [2]. Intelligent technology offers a solution to the limitations of traditional teaching methods by enabling personalized and targeted approaches tailored to diverse student needs. In the context of public English teaching in universities, existing models face challenges including low student participation, inadequate assessment of learning outcomes, and a lack of personalized teaching processes. To enhance the objectivity and efficiency of classroom management in this setting, facial recognition technology is being implemented for attendance tracking in intelligent classrooms. Convolutional Neural Networks (CNN) have strong feature learning capabilities and have been widely used in fields such as computer vision and pattern recognition [3]. Therefore, leveraging CNN-based recognition technology in intelligent classrooms allows for accurate identification of students' emotions, expressions, and interactive behavior, enabling educators to make timely teaching adjustments and create a more personalized learning environment. This will help improve students' learning motivation and interest, thereby improving teaching effectiveness. CNN-based recognition technology can monitor and analyze students' expressions, postures, etc. in real time. By identifying students' emotions and interactive behaviors, it can help teachers better understand their learning status. For example, when the system detects that students have low emotions or difficult learning situations, teachers can provide timely attention and assistance. In addition, an intelligent classroom monitoring system based on

*School of Foreign Studies, South China Agricultural University, Guangzhou, 510642, China (shuhuili1980@outlook.com)

this technology can also provide teachers with data such as student participation and interest levels, helping with personalized teaching. Therefore, the application of CNN-based recognition technology in public English teaching in intelligent classrooms in universities has important practical significance and potential application prospects, which involves important ethical, social, and educational considerations. The facial data of students belongs to personal sensitive information and must be appropriately protected. To ensure the responsible use of data, strict data protection policies should be established, with clear communication and consent from students regarding the usage, storage, and processing of their personal information. Robust encryption and security measures must be implemented to prevent unauthorized access or misuse of stored data. Facial recognition technology should be implemented in a fair and unbiased manner, ensuring equal treatment for all students. By utilizing facial recognition for attendance tracking, teachers can effectively monitor student engagement. This data can also provide insights into students' learning habits and needs, facilitating personalized teaching support. Leveraging CNN for facial recognition and addressing its limitations will enhance the efficiency of attendance management in public English teaching in intelligent university classrooms, promoting scientific and standardized approaches to education.

2. Related works. Intelligent recognition is an important content in many fields at present, and recognition efficiency and accuracy make a big difference in the final result. Face recognition technology based on CNN can improve the accuracy of face recognition, it has been applied to many fields by many scholars and achieved many research results.

For the purpose of improving the teaching efficiency of multimedia English, Hao K put forward an intelligent network teaching system model, which used deep learning speech enhancement and facial expression recognition technology to judge students' understanding of emotions. The result showed a good detection effect [4]. Duan R and other scholars established an AI speech recognition correction model based on AI speech recognition technology to solve the problem of limited oral English pronunciation correction. The test results showed that the model effectively assisted teachers in correcting students' spoken English pronunciation [5]. In view of the shortcomings of intelligent speech recognition, Dong S established an intelligent English recognition and prediction system based on support vector machine and combined with wavelet packet analysis to extract features of EEG signals. The result showed that the system built by the research improved the speech recognition rate [6]. For the purpose of improving the efficiency of cross-cultural English teaching, Zhang M and other scholars established a cultural O2O English teaching system that supported cross-intelligent recognition and management to detect and recognize students' emotions. Combined with the neural network algorithm optimization system, the result showed that the system built by the research had a nice performance [7]. Li A and other researchers built an English intelligent online teaching model to solve the problem that the traditional English online teaching mode was limited by the location and used the improved deep belief network to identify the students' status and locate the location. The result showed that the model effectively enhanced the monitoring and recognition of students' online learning status, and had a good performance [8].

Yu J, for the purpose of achieving task recognition in English classroom teaching, used improved CNN to recognize images and combined GPU to solve the problem that data training took a long time. The result showed that the way proposed in the study validly perceived the results according to the target location, which was conducive to task recognition in English teaching [9]. Chen X, for the purpose of enhancing the recognition of English speech emotion, established a training strategy based on transfer learning and CNN. The result showed that the proposed method had a good performance in English speech emotion recognition [10]. For the purpose of improving the efficiency of face recognition, Karkal G and other researchers built a non-invasive attendance system, which used CNN to detect faces and a neural network to recognize faces. The result showed that the system realized automatic detection of face recognition and improved the efficiency of attendance [11]. For the purpose of improving the efficiency of classroom attendance, Seelam V and others proposed an intelligent classroom attendance management system based on face recognition by using the principles of deep learning and computer vision. The result showed that compared with traditional manual attendance, the former improved the efficiency of attendance [12]. Shah V and other scholars established a recognition model based on machine learning and depth learning algorithms to improve the accuracy of emotional state recognition. The result showed that the model effectively processed text data and had a high accuracy in emotional state recognition [13].

Shen Y and other researchers built an intelligent assessment system for autistic children in view of the lack of professional evaluators for autistic children's painting. They developed an assessment model for autistic children by utilizing LSTM and CNN to segment and identify the components of portrait sketches. The result showed that this model validly judged children's autistic tendencies [14]. Lan C and other scholars, in order to enhance the recognition rate of speaker recognition system, constructed a model combining the additive Margin Softmax loss function, which combined CNN with gated recursive units. The speaker model was trained using the data enhancement method. The result showed that compared with other models, the equal error rate of this model was reduced, the recognition rate was greatly enhanced, and it had good robustness [15]. Chen Z and other scholars proposed a lightweight real-time facial recognition algorithm based on CNN to improve the accuracy of facial recognition, therefore reducing the parameters and computational complexity of facial feature extraction networks. The research results indicated that this method improved the accuracy of facial recognition and the recognition time was shorter than the current recognition methods [16]. Rao T et al. designed a multi-scale graph CNN-based facial expression recognition method to improve the accuracy of facial expression recognition. The experimental results showed that this method had a more stable performance and improved the accuracy of facial expression recognition [17]. Xianzhang P proposed a video facial expression recognition method based on CNNs and gradient direction histograms to address the low efficiency of facial expression recognition in videos. This method comprehensively extracted facial features from videos and recognized facial expressions. The research results indicated that this method effectively extracted facial expressions from videos, thereby improving the performance of facial expression recognition [18].

The above content is the intelligent recognition research conducted by scholars from different fields combined with CNN. Compared to traditional facial recognition methods, CNN technology has high accuracy and robustness in image recognition and can better recognize faces under different lighting, angles, and facial expressions. However, using CNN for facial recognition requires a large amount of labeled data and computational resources for training. In order to improve the efficiency of teacher attendance in public English teaching in intelligent classrooms in universities, this study will introduce Faster R-CNN, which can perform object detection faster, improve recognition speed and accuracy, and facilitate the smooth implementation of teacher teaching work.

3. Face recognition in public English teaching attendance for college intelligent classroom.

3.1. Research on face recognition and CNN optimization. Since the 1970s, face recognition technology has risen and developed rapidly. There are many excellent methods used in face recognition, including geometric feature matching, template matching, neural network, and other methods Among them, the matching method based on facial geometric features, as the most effective face matching method, has the advantage of simple recognition methods, but it mainly depends on the accuracy of facial organ feature localization algorithm, and the localization algorithm has a high complexity. At the same time, the accuracy of face recognition will be affected by occlusion and expression The detection method based on the feature face uses principal component analysis to project the face image into the feature face subspace, calculate the position and length of the projection point, and then recognize the face. This method has a simple calculation method and fast recognition speed, but it requires high image quality, and other external environmental factors also have a greater impact on the recognition results. The way based on template matching mainly uses coded matching to complete face recognition. Template matching can be divided into points, regions, and a mixture of the two according to the level. The calculation process of the first way is relatively complex, the calculation amount of the second method is reduced, and the matching method based on the combination of the two methods has more recognition efficiency. The recognition method based on the hidden Markov model has a high recognition rate, and can better realize the recognition of face images under different conditions. The method based on the neural network has stronger expression ability and adaptability, and the recognition efficiency is better. CNN is the representative network structure of the neural network and has made many breakthroughs in face recognition.

Face detection is a key step in face recognition. Its results have a great impact on the extraction of face features and the accuracy of face recognition. Conventional face detection mainly extracts features manually and then trains classifiers, but it has high requirements for the detection environment CNN can better extract facial features. CNN structures generally include an input layer, convolution layer, down-sampling layer, and output layer. The convolution layer and down-sampling layer appear alternately between the input layer and the output layer and are important modules for feature extraction [19]. At the same time, activation functions are also an important part of CNN, which can increase the expression ability and learning ability of models. Common activation functions include the Sigmaid function, Tanh function, and ReLU function. Among them, Sigmoid function and Tanh function are continuously differentiable functions, monotonically increasing, which is convenient for forward propagation of the network and unfavorable for back propagation. The expressions of the two functions are shown in Formula 3.1.

$$\begin{cases} f(x) = \frac{1}{1+e^{-x}} \\ tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{cases} \tag{3.1}$$

The ReLU function generally outputs hidden neurons, and the expression is shown in Formula 3.2.

$$f(x) = max(0, x) \tag{3.2}$$

The expressions (2.1) and (2.2) show that the ReLU function requires less computation than the three functions. When $X < 0$, the occurrence rate of over fitting is greatly reduced, which is not conducive to back propagation. Therefore, it is improved to obtain the Leaky ReLU function. The expression is shown in Formula 3.3.

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ ax, & \text{if } x < 0 \end{cases} \tag{3.3}$$

The Leaky ReLU function is used as the activation function of CNN. Meanwhile, the loss function can improve the accuracy and stability of the model. Its expression is shown in Formula 3.4.

$$f(x) = -\log\left(\frac{e^{f_{yi}}}{\sum_j e^{f_j}}\right) \tag{3.4}$$

In formula 3.5, f_j represents the jth element in the classification score vector f [20]. The normalized exponential function also has a significant impact on most CNNs, and its expression is shown in formula 3.5.

$$f_j(z) = \frac{e^z j}{\sum ke^z k)} \tag{3.5}$$

The random gradient algorithm can decrease the difference between the real value and the predicted value when the parameters are updated, which is conducive to obtaining the optimal weight parameters. At the same time, CNN uses iterative training to improve its ability to extract features and combines gradient descent way to reverse adjust the weight parameters of the network layer by layer. CNN also has the function of multi-layer perception, which enables local receptive field, convolutional kernel weight sharing, and spatial subsampling. In turn, CNN has the advantages of lower network complexity, easier network tuning, and less risk of fitting Regions with convolutional neural network feature (R-CNN) have a better effect in target detection. The general steps of R-CNN are shown in Figure 3.1.

In Figure 3.1, R-CNN generates candidate regions for each image to be detected through the selective search (SS) algorithm. Among them, SS extracts the features of candidate regions by merging sub regions. First, the input image is divided into many small regions, and a hierarchical grouping algorithm is used in this process. After that, the regions are merged through the similarity between small regions, and iterative merging is performed. When all images are merged into a whole region, the iterative process is ended. A circumscribed rectangle is then made for the sub-regions merged in the iteration process, and the candidate regions are the circumscribed rectangles of the sub-regions. Finally, the candidate regions are output. Before feature extraction, it is necessary to normalize and unify the size of candidate regions, and then extract features using CNN to obtain feature vectors [21]. Then the classification processing is carried out. When it belongs to the corresponding category, the position coordinates of the candidate box to be corrected can be output after its position. It is worth noting that the detection effect of R-CNN is better than that of conventional detection

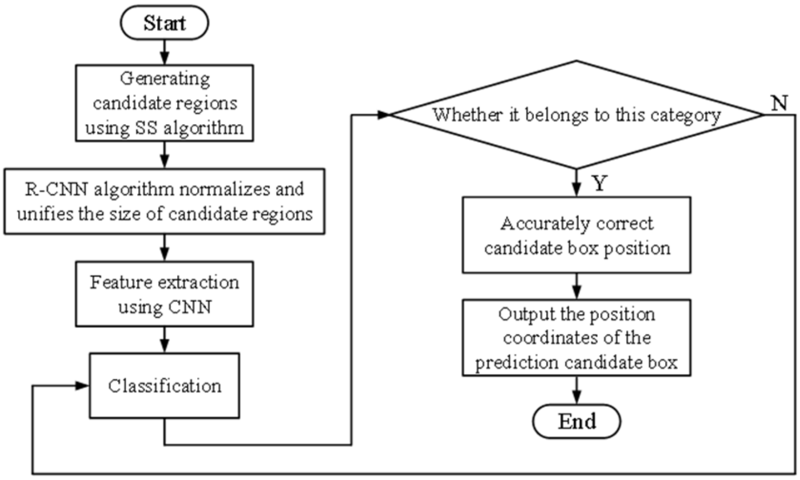

Start
Generating candidate regions using SS algorithm
R-CNN algorithm normalizes and unifies the size of candidate regions
Feature extraction using CNN
Classification
Whether it belongs to this category
N
Y
Accurately correct candidate box position
Output the position coordinates of the prediction candidate box
End

Fig. 3.1: R-CNN flow chart

methods, but there are also issues such as long training and testing time and large memory consumption in the process.

To address the issue of information loss and deformation caused by normalizing and unifying candidate regions in size, and to improve the accuracy of recognition results, Spatial Pyramid Pooling (SPP) can be employed as an intervention technique. As a multi-scale Pooling technology, image features can be extracted from more than one angle, and feature maps of different sizes can be converted into feature vectors of uniform size, thus improving the accuracy of image detection.

In addition, in the process of face target detection, since more than one target candidate box region will contain the same face, the candidate boxes will contain or cross each other At this time, the candidate region box can be filtered by using non-maximum suppression (NMS). NMS can suppress non-maximum elements, search for local maximum, eliminate overlapping windows, and then get the best candidate area for face detection

3.2. Face recognition based on improved Faster R-CNN algorithm. Fast R-CNN adopts the method of sharing the convolution layer. After a complete graph is input into the network, the operation of extracting the features of candidate regions is implemented on the final convolution layer, which can avoid repeated calculation of convolution. The overall framework of the Fast R-CNN algorithm is shown in Figure 3.2.

Due to some defects in the speed of Fast R-CNN, it takes more time to find all candidate regions using the SS algorithm, so Fast R-CNN uses the Region Proposal Network (RPN) instead of SS to improve the detection speed. Fast R-CNN represents the combination of Fast R-CNN and RPN. It mainly uses the CNN network to detect targets and uses RPN to generate candidate regions from the extracted feature map. In addition, since the area generation network and classification network use the same CNN to complete the weight sharing, the detection speed and accuracy can be improved. In view of the large number of RPNs generated by the Faster R-CNN algorithm, and in order to give consideration to the RPN size and the proportion of width and height and lose the pertinence, it is necessary to improve the face target search algorithm in Faster R-CNN, using RPN as the area search strategy. The RPN structure is shown in Figure 4.1.

RPN is a full convolutional network, and shares all convolutional features with all CNNs used for detection. It can extract high-quality detection areas and save area recommendation time. In Figure 3.3 , the output characteristic diagram of the last convolution layer of the shared convolution network uses a × 3. The window is fully connected. The feature is mapped to a low-dimensional vector and then sent to the parallel box classification layer (cls) and box regression layer (reg). The positions of the points of the convolved feature graph and the original graph are mapped, and every point on the former graph is mapped to the latter at the same time, the center of each sliding window has anchor points at the position corresponding to the original

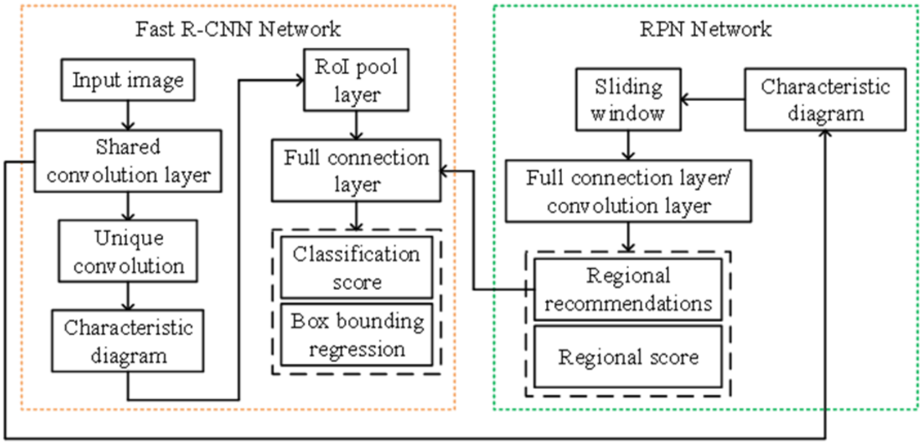


Fig. 3.2: Overall framework of Fast R-CNN algorithm

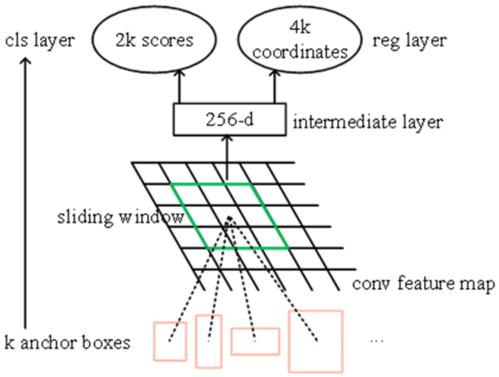


Fig. 3.3: RPN structure

drawing, and each anchor point corresponds to a size and aspect ratio [22]. RPN has three dimensions, length to width ratio, so each feature map element generates a total of 9 regional recommendations. When the RPN network is trained, the anchor points generate different anchor frames according to the corresponding rules due to the particularity of their positions. The Faster R-CNN algorithm takes each anchor point as the center and generates three anchor frames with different length-width ratios and pixel areas. Then all generated anchor frames are classified, including positive sample sets and negative sample sets, according to the Image Interaction Over Union (IoU) standard. The calculation way of training RPN. IoU is shown in Formula 3.6.

$$\mathrm{IoU} = \frac{\mathrm{area}(C) \cap \mathrm{area}(G)}{\mathrm{area}(C) \cup \mathrm{area}(G)} \tag{3.6}$$

At Formula 3.6, $area(C)$ expresses the generated Candidate bound,$area(G)$ expresses the original Ground truth bound, and IoU represents the overlap rate and the ratio of their intersection and union, which can indicate the accuracy of object detection in a specific data set. When $IoU = 1$, it means that they are all overlapped, indicating that the accuracy is the best.$IoU < 0.7$ divides the anchor frame into negative samples, and $IoU > 0.7$ or the value of the anchor frame and target into positive samples at most.

Given that the face target image typically contains a small number of pixels and the aspect ratio of the

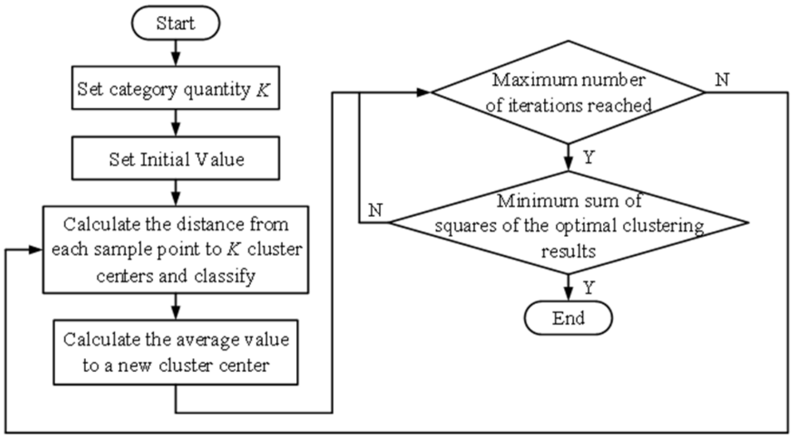

Start
Set category quantity K
Set Initial Value
Calculate the distance from each sample point to K cluster centers and classify
Calculate the average value to a new cluster center
Maximum number of iterations reached
N
Y
Minimum sum of squares of the optimal clustering results
N
Y
End

Fig. 3.4: Basic process of K-Means

face's bounding box is relatively small, utilizing the features of the bounding box as the face candidate area for the RPN network can enhance feature extraction and overall speed in face detection. In view of this, an unsupervised K-Means clustering way is used to learn the characteristics of face candidate regions from many labeled samples. First, the input sample set is $D = X_1, X_2, ...X_m$, the amount of clusters is K, n represents the maximum amount of iterations, $C = C_1, C_2, ...C_k$ describes that the output is cluster division, and the basic flow of K-Means clustering is shown in Figure 3.4.

The K-Means clustering method can be used to obtain candidate regions more suitable for the face dataset in this study for RPN, and the anchor size obtained is closer to the face bounding box of the dataset in this study, which can save training time, improve detection accuracy and speed up convergence. Among them, K-Means clustering uses the square of Euclid distance to calculate the distance, as shown in formula 3.7.

$$\text{dist}(x,y)^2 = \sum_{i=1}^{n}(x_i - y_i)^2 = \|x-y\|_2^2 \tag{3.7}$$

In formula 3.7,x and y are two different samples, and n is used to describe the sample dimension [23]. At the same time, the sum of the squared error (SSE) is used to judge the clustering results. When SSE is the minimum value, the best clustering result is obtained. The calculation way of SSE is shown in formula 3.8 .

$$\text{SSE} = \sum_{i=1}^{k}\sum_{x\in C_i}\text{dist}(x, c_i) \tag{3.8}$$

In formula 3.8,c_i represents the center of cluster C_i . Common evaluation indexes are cited to appraise the results, including precision, recall rate, false positive rate, PR curve, average accuracy (AP), and mean average precision (mAP). In addition, true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) are introduced into the calculation. The calculation method of accuracy is shown in Formula 3.9.

$$P = \frac{TP}{TP + FP} \tag{3.9}$$

In formula 3.9,P represents the precision,TP expresses the number of positive samples predicted to be positive samples, and $TP + FP$ represents the number of correct examples. The calculation way of the recall rate is shown in formula 3.10.

$$R = \frac{TP}{TP + FN} \tag{3.10}$$

Table 4.1: System parameter

Number	Project	Size	Unit
#1	Operating system	Windows 10	/
#2	Experimental platform	Tensor Flow	/
#3	GPU	GTX 1080Ti	SM
#4	Memory	1024	Mb
#5	Working voltage	220	V

In formula 3.10, R represents the recall rate and $TP+FN$ represents the number of correctly labeled instances. The calculation way of the FP rate is shown in formula 3.11.

$$F = \frac{FP}{TP + FN} \tag{3.11}$$

In formula 3.11,F is the FP rate, and is the amount of correct target instances predicted from the wrong target [24]. The calculation way of AP is shown in formula 3.12.

$$AP_C = \frac{1}{n}\sum_{i=1}^{n} P_{C,i} \tag{3.12}$$

In formula 3.12,AP_C is the AP of Category. The mAP is calculated as shown in formula 3.13.

$$mAP = \frac{1}{m}\sum_{i=1}^{m} AP_C \tag{3.13}$$

In formula 3.13,m represents the number of categories to be detected. Generally, the larger the mAP value is, the better the detection algorithm is.

4. Face recognition training experiment and result analysis.

4.1. Experimental analysis of face recognition model training. In the analysis of the application of CNN-based recognition technology [25, 26, 27] in public English teaching in intelligent classrooms in universities, a significant step was to gather a substantial amount of student face data within university classroom settings. Established facial recognition datasets like CASIA WebFace and LFW were utilized as a starting point. However, to enhance the effectiveness of the model in practical applications, data from specific classroom environments was incorporated. The data was appropriately annotated with student identity information, ensuring consistent distribution for training and testing purposes. The face recognition process was trained and a simulation model was created, facilitating the implementation of the model on the Tensor Flow platform. Leveraging the capabilities of Tensor Flow, the model could be automated for seamless operation. The platform also provided support for various algorithms such as CNN and RNN, ensuring a high degree of compatibility and assistance throughout the process. Other parameter settings of the simulation model are shown in Table 4.1.

The CNN excitation function was set as the Sigmaid function, Tanh function, ReLU function, and Leaky ReLU function respectively to learn face features and observe the performance of the CNN to obtain four different excitation function training networks. The results are shown in Figure 4.1.

In Figure 4.1, the horizontal axis expresses the number of iterations of the CNN network structure training, and the vertical axis expresses the convergence value of the loss function during training. Sigmoid function, Tanh function, ReLU function, and Leaky ReLU function are represented by red, purple, orange, and green lines respectively. In Figure 4.1, among the four excitation functions, the loss value of the last three functions was significantly lower than that of the first function, indicating that the last three functions had a high recognition rate. Among the three functions, the Leaky ReLU function had the fastest convergence speed at the same time, so the Leaky ReLU function was used as the excitation function of the model. Then the three networks were trained to get the training results, as shown in Figure 4.2.

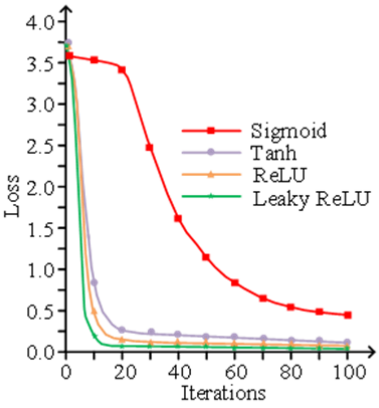


Fig. 4.1: Training of CNN by excitation function

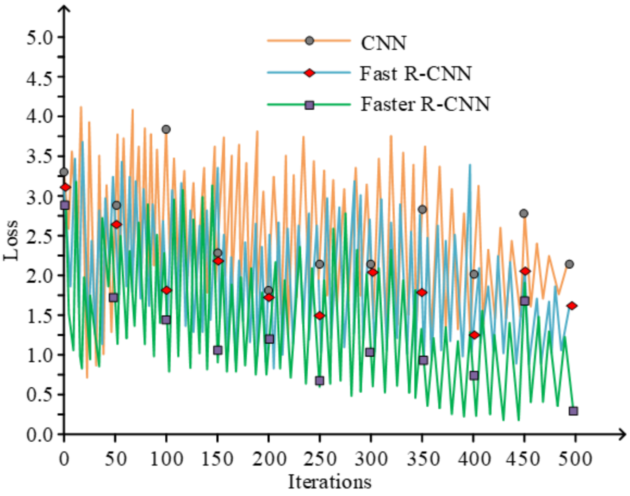


Fig. 4.2: Training of three networks

In Figure 4.2, the abscissa expresses the amount of training iterations, and the ordinate expresses the loss value of the loss function in the training process. The orange line expresses the training iteration of CNN, the blue line expresses the training iteration of Fast R-CNN, and the green line expresses the training iteration of Fast R-CNN. In Figure 6, the convergence rates of the three networks differed greatly. When the number of iterations was 250, the Loss values of CNN, Fast R-CNN, and Faster R-CNN were 2.19, 1.51, and 0.72 respectively. When the number of iterations was 500, the loss values of CNN, Fast R-CNN, and Faster R-CNN were 2.43, 1.67, and 0.27 respectively, that is, Faster R-CNN converged faster among the three networks, maintained a high level of recognition rate, and had a Faster convergence rate. The evaluation indicators mainly included precision, recall, AP, and mAP to analyze the proportion of correctly recognized faces to the total number of faces, as well as the accurate recognition of each student in public English teaching. Cross-validation methods were used to evaluate the stability and generalization ability of the method, tests were conducted in different classroom environments, and the adaptability of the method was analyzed. Three network structures were evaluated in combination with evaluation indicators, and the outcomes are shown in Table 4.2.

In Table 4.2, among various evaluation indicators, the precision, recall, AP, and mAP of the three networks

Table 4.2: Comparison of three network evaluation indicators (%)

Iterations	Index	CNN	Fast R-CNN	Faster R-CNN
100	Precision	66.3	68.2	88.1
	Recall	77.1	82.5	86.7
	AP	80.2	83.4	85.6
	mAP	79.7	82.1	83.4
200	Precision	72.5	82.3	91.4
	Recall	80.5	84.6	87.1
	AP	81.6	84.7	86.3
	mAP	80.1	82.9	83.7
300	Precision	73.6	83.8	93.7
	Recall	82.7	85.3	89.4
	AP	82.4	85.3	87.6
	mAP	83.6	83.9	84.3
400	Precision	77.1	86.3	94.5
	Recall	83.4	86.9	92.7
	AP	83.9	86.2	88.1
	mAP	84.1	85.2	86.4
500	Precision	82.5	93.4	96.2
	Recall	86.3	88.4	94.8
	AP	84.2	87.5	89.4
	mAP	82.9	85.7	87.2

Table 4.3: Recognition rate result of facial expression recognition (%)

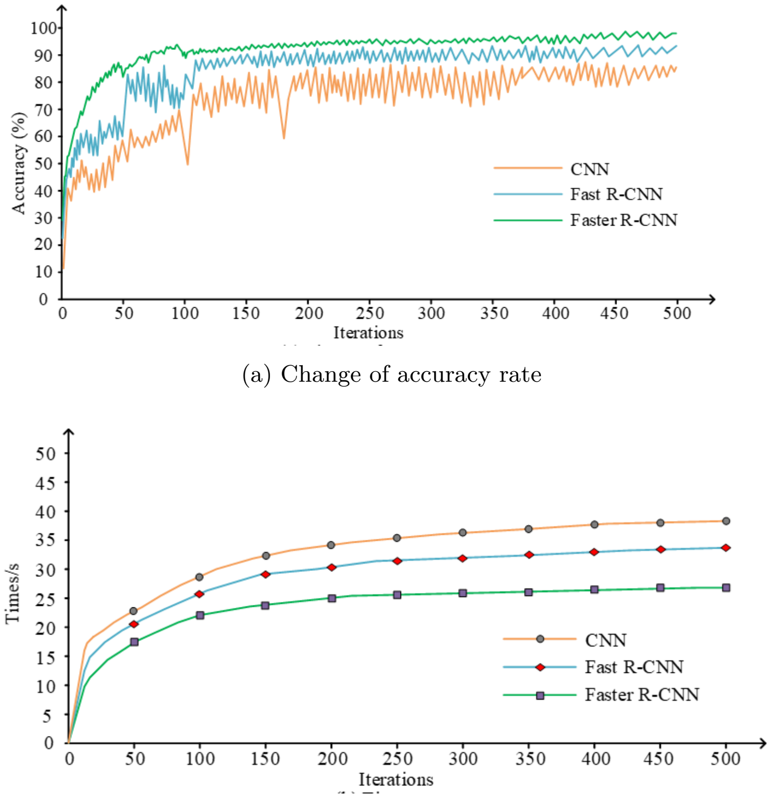
Number	Expression	CNN	Fast R-CNN	Faster R-CNN
#1	Smiling face	80.2	89.3	94.2
#2	Doubt	68.4	70.4	91.3
#3	Drowsiness	76.9	68.7	90.7

were Faster R-CNN, Fast R-CNN, and CNN in descending order. The improved Faster R-CNN had the strongest comprehensive performance and validly enhanced the detection efficiency.

4.2. Analysis of face recognition results based on optimized CNN network structure. The change in recognition accuracy and training iteration time of the three network models in the face recognition iteration process were simulated, and the results are shown in Figure 4.3.

In Figure 4.3, the orange line, blue line, and green line respectively represent the accuracy and time use of CNN, Fast R-CNN, and Faster R-CNN in the training iteration process. Subgraph (a) represents the accuracy rate change, subgraph (b) represents the time use, and the abscissa of the two subgraphs shows the number of iterations. The ordinate of subgraph (a) represents the accuracy, in%. The ordinate of subgraph (b) represents the time, in seconds (s). In subgraph (a), the accuracy of the three network models from low to high was CNN, Fast R-CNN, Faster R-CNN, and the accuracy of Faster R-CNN tended to be stable first. In subgraph (b), when iteration was 500 times, the time of R-CNN, Fast R-CNN, and Faster R-CNN was 34.2s, 26.8s, and 23.7s respectively. That is, in terms of the accuracy and time of face recognition, Faster R-CNN had the highest accuracy and the shortest time, after that, three models were used to recognize facial expressions, and the outcomes are shown in Table 4.3.

Table 4.3 shows that Faster R-CNN still had the highest recognition rate for specific facial expression recognition. Faster R-CNN had 94.2%, 91.3%, and 90.7% recognition rates for smiling faces, doubts, and sleepiness, 89.3%, 70.4%, and 68.7% respectively for smiling faces, doubts, and sleepiness, and 80.2%, 68.4%, and 76.9% respectively for smiling faces, doubts, and sleepiness. Among the three models, smiling faces had the highest recognition rates for different facial expressions Therefore, Faster R-CNN was used to analyze students'



(a) Change of accuracy rate

(b) Time Usage

Fig. 4.3: Changes in recognition accuracy and time consumption during iteration

mental state and classroom performance on the basis of face recognition, which was beneficial for teachers to master students' learning in teaching.

In order to further analyze the research methods, the proposed model was compared with more advanced You Only Look Once (YOLO) and Single Shot Multibox Detector (SSD). These methods made significant breakthroughs in the field of object detection and recognition and were applied to public English teaching in intelligent classrooms in universities. The results are shown in Figure 4.4.

From Figure 4.4, the accuracy of the research method was comparable to other advanced methods. However, the research method had a higher recognition rate for different expressions in college intelligent classroom English teaching, indicating that the research method had good comprehensive performance and good application effects in public English teaching in college intelligent classrooms.

5. Conclusion. With the increasing number of students in school, teachers' classroom attendance will take more time, increase the difficulty of attendance and teachers' workload, and have some adverse effects on teachers' other teaching work. Public English teaching in intelligent classrooms in colleges and universities also faces this problem. As a required course for every student in colleges and universities, teachers will spend more energy on attendance. Therefore, it is urgent to improve the efficiency of attendance by using intelligent attendance methods. By using the CNN recognition technology and improving its shortcomings, Faster R-CNN is obtained. The experimental results showed that the Leaky ReLU function had the fastest convergence speed and was used as the excitation function of the network model. Faster R-CNN, Fast R-CNN, and CNN have the lowest loss value in the training process. When the number of iterations was 500, the loss values of the three networks were 2.43, 1.67, and 0.27 respectively. Based on the evaluation index, the three network structures

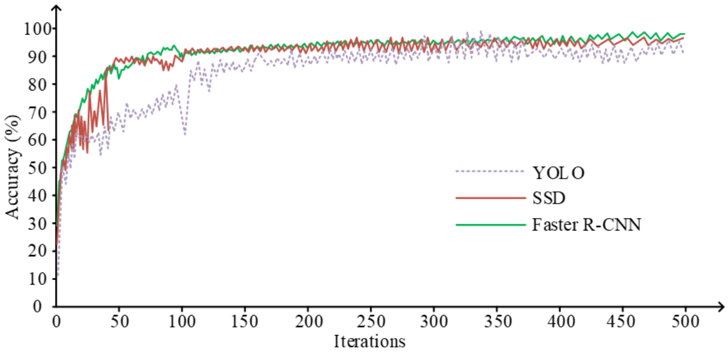


Fig. 4.4: Comparison between research methods and other advanced methods

were evaluated, and it was found that the precision, recall, AP, and mAP of Faster R-CNN were the best. In terms of accuracy and time of face recognition, Faster R-CNN had the highest accuracy, the shortest time, and tended to be stable first. In terms of specific facial expression recognition, Faster R-CNN still had the highest recognition rate. That is, Faster R-CNN maintained a high level of recognition rate and had faster convergence speed and better performance. It enhanced the efficiency of classroom attendance in teaching and promoted the development of intelligent attendance. At the same time, how to achieve intelligent attendance in the classroom on mobile devices is also the main research direction in the future. The advantages of Faster R-CNN include high accuracy, fast detection speed, and end-to-end training process, but there are complex network structures, weak target recognition ability, and complex training and parameter tuning problems. Efforts should be made to continuously enhance algorithms, minimizing false recognition rates. Facial recognition can be combined with other intelligent technologies to enrich classroom interaction. Additionally, regular evaluations of data management policies can be conducted to ensure compliance with the most recent ethical and legal standards. Students can be actively involved in the evaluation and improvement process, ensuring their needs and concerns are considered and addressed effectively. Overall, the use of this technology requires a balance between the benefits of technological progress and the accompanying ethical and social responsibilities. Through continuous technological improvement, policy formulation, and stakeholder engagement, the positive impact of these systems can be maximized while minimizing potential negative effects.

6. Fundings. The research is supported by The 2023 Guangdong Institute of Higher Education ”14th Five Year Plan” Higher Education Research Project “Theory and Practice Research on Blended Teaching Model of English Movie Audiovisual Listening and Speaking Teaching from the Perspective of POA” (Project No. 23GYB08), The 2023 South China Agricultural University- Curriculum Ideological and Political Demonstration Project ”Oral English in Practical Situations” (Project No. KCSZ2023024).

REFERENCES

[1] Yongjun, Z., Wenjie, L., Haisheng, F., Yongjie, Z., Zhongwei, C. & And, Q. and face recognition based on sample expansion. *Applied Intelligence: The International Journal Of Artificial Intelligence, Neural Networks, And Complex Problem-Solving Technologies.* **52**, 3766-3780 (2022)

[2] Liu, X. & Yang, Z. Computer-aided teaching mode of oral English intelligent learning based on speech recognition and network assistance. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **41**, 5771-5771 (2021)

[3] Shen, J. & Wu, J. Speech Recognition in Noise Performance Measured Remotely Versus In-Laboratory from Older and Younger Listeners. *Journal Of Speech, Language, And Hearing Research: JSLHR.* **65**, 2391-2397 (2022)

[4] Hao, K. Multimedia English teaching analysis based on deep learning speech enhancement algorithm and robust expression positioning. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **39**, 1779-1791 (2020)

[5] Duan, R., Wang, Y. & Qin, H. Artificial intelligence speech recognition model for correcting spoken English teaching. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **40**, 3513-3524 (2021)

[6] Dong, S. Intelligent English teaching prediction system based on SVM and heterogeneous multimodal target recognition. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **38**, 7145-7154 (2020)

[7] Zhang, M. & Zhang, L. Cross-cultural O2O English teaching based on AI emotion recognition and neural network algorithm[J]. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **40**, 7183-7194 (2021)

[8] Li, A. & Wang, H. An artificial intelligence recognition model for English online teaching. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **40**, 3547-3558 (2021)

[9] Yu, J. Analysis of task degree of English learning based on deep learning framework and image target recognition. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **39**, 1903-1914 (2020)

[10] Chen, X. Simulation of English speech emotion recognition based on transfer learning and CNN neural network. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **40**, 2349-2360 (2021)

[11] Karkal, G., Reddy, K., Singh, K., Hosangadi, N. & Patil, A. Feature Learning Approach for Facial Recognition Using Deep Metric Learning. *Journal Of Computational And Theoretical Nanoscience.* **17**, 4125-4130 (2020)

[12] Seelam, V., Penugonda, A., Kalyan, B., Priya, M. & Prakash, M. Smart attendance using deep learning and computer vision. *Materials Today: Proceedings.* **46**, 4091-4094 (2021)

[13] Shah, V. & Mehta, M. Emotional state recognition from text data usingmachine learning and deep learning algorithm. *Concurrency And Computation: Practice And Experience.* **34** pp. 17 (2022)

[14] Shen, Y., Wang, X., Chen, Z., Sun, Q., Zhang, X., Liang, H. & Pan, J. Intelligent recognition of portrait sketch components for child autism assessment. *Computer Animation And Virtual Worlds.* **33** pp. 3 (2022)

[15] Lan, C., Wang, Y., Zhang, L. & Zhao, H. Research on Additive Margin Softmax Speaker Recognition Based on Convolutional and Gated Recurrent Neural Networks. *Journal Of The Audio Engineering Society: Audio, Acoustics, Applications.* **70**, 611-620 (2022)

[16] Chen, Z., Chen, J., Ding, G. & Huang, H. lightweight CNN-based algorithm and implementation on embedded system for real-time face recognition. *Multimedia Systems.* **29**, 129-138 (2023)

[17] Rao, T., Li, J., Wang, X. & Sun, Y. Chen H. *Acial Expression Recognition With Multiscale Graph Convolutional Networks.* **28**, 11-19 (2021)

[18] Hog, X. & Neural, C. Network spatial-temporal features for video-based facial expression recognition. *IET Image Processing.* **14**, 176-182 (2020)

[19] Yiyang, T., Chenguang, S. & Bin, W. Toward jointly understanding social relationships and characters from videos. *Applied Intelligence: The International Journal Of Artificial Intelligence, Neural Networks, And Complex Problem-Solving Technologies.* **52**, 5633-5645 (2022)

[20] Sumun, K., Connor, P., Yassmin, P., Kaiti, D., Arvind, U., Navid, P., Joseph, L., Age, D. & Using, A. Facial Recognition Software in Rhinoplasty Patients: A Proof-of-Concept Study. *The Journal Of Craniofacial Surgery.* **33**, 1540-1544 (2022)

[21] Godson, A. Tetteh. Effects of Classroom Attendance and Learning Strategies on the Learning Outcome. *Journal Of International Education In Business.* **11**, 195-219 (2018)

[22] Gao, Q., Cao, B., Guan, X., Gu, T., Bao, X., Wu, J., Liu, B. & Cao, J. Emotion recognition in conversations with emotion shift detection based on multi-task learning. *Knowledge-based Systems.* **8861**, 1-10886 (2022)

[23] Gutz, S., Stipancic, K., Yunusova, Y., Berry, J. & Green, J. Validity of Off-the-Shelf Automatic Speech Recognition for Assessing Speech Intelligibility and Speech Severity in Speakers with Amyotrophic Lateral Sclerosis. *Journal Of Speech, Language, And Hearing Research: JSLHR.* **65**, 2128-2143 (2022)

[24] Chakraborty, S., Amrita, C., Sille, R., Dutta, C. & Dewangan, B. Multi-view Deep CNN for Automated Target Recognition and Classification of Synthetic Aperture Radar Image. *Journal Of Advances In Information Technology.* **13**, 413-422 (2022)

[25] Lee, S., Abdullah, A. & Jhanjhi, N. A review on honeypot-based botnet detection models for smart factory. *International Journal Of Advanced Computer Science And Applications.* **11** (2020)

[26] Azeem, M., Ullah, A., Ashraf, H., Jhanjhi, N., Humayun, M., Aljahdali, S. & Tabbakh, T. Fog-oriented secure and lightweight data aggregation in iomt. *IEEE Access.* **9** pp. 111072-111082 (2021)

[27] Gaur, L., Solanki, A., Wamba, S. & Jhanjhi, N. Advanced AI techniques and applications in bioinformatics. (CRC Press,2021)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: Nov 8, 2023
Accepted: Jan 10, 2024