Research Article

An Online English Writing Evaluation System Using Deep Learning Algorithm

Pei Wang1 and Xinyi Huang2

1School of Foreign Languages, Chuzhou University, Chuzhou 239099, China
2College of Information and Network Engineering, Anhui Science and Technology University, Chuzhou 233100, China

Correspondence should be addressed to Pei Wang; wangpei2828@chzu.edu.cn

Received 14 April 2022; Revised 15 May 2022; Accepted 6 July 2022; Published 12 August 2022

Abstract

In English writing, automatic evaluation systems are becoming more and more common and are receiving more and more attention. As a teaching application, an automatic evaluation system uses information technologies such as corpus and artificial intelligence to quickly score and modify students’ writing. To expedite the realization of teaching information, this system records students’ writing processes, which are closely related to the instruction requirements of the college’s English program and meets the personalized needs by combining the current situation of English writing and teaching. This study investigates the impact of field cognitive style and automatic evaluation system on college writing training. From the perspective of cognitive style differences, this paper summarizes the application strategies of an automatic evaluation system in College English writing and teaching, to better realize the integration of information technology and subject teaching and improve students’ English writing ability and level. The deep learning-based system for grading students in the classroom is completed by this paper. The network data transmission module allows several ends to all experience stable data interaction.

1. Introduction

The teaching process includes analyzing the learning environments of the pupils, formulating teaching objectives, designing teaching activities, and providing feedback and suggestions [1]. Learning evaluation includes formative evaluation and summative evaluation. In this process, teachers collect evidence and data about students’ academic progress to provide learners with further learning suggestions. Appropriate evaluation methods are an effective feedback mechanism. According to some experts, learning evaluation is a teaching activity, which aims to improve students’ participation and learning effect. Feedback is a core element in teaching activities. It can reflect students’ learning achievements and progress, support students’ learning, and monitor teachers’ teaching [2]. The purpose of teaching evaluation is not to judge the IQ (intelligent questions) of students, nor to select “good students” and “bad students” to treat them differently, but to provide evidence for students’ learning efforts at a certain stage, so that teachers can give more optimized teaching support.

The Chinese composition in the college entrance examination is 60 points, accounting for 40% of the full score of 150 points. The full score of Chinese in the middle school entrance examination is usually 120 points, and the composition accounts for about 40–50 points [3]. Based on these statistics, it is obvious that the score of Chinese composition is very important for students participating in the college entrance examination. Of course, meeting exam-oriented education is the most basic writing requirement. The purpose of writing is to improve students’ written expression ability, which is the ability and quality that students should develop continuously even if they go out of campus and finish school education [4]. Teachers give students effective writing feedback and corresponding modification suggestions to help them gradually understand and master the teaching skills of writing. The ultimate goal is to help pupils and enhance their writing skills [5]. Composition assessment
is not a product of composition instruction, but an important means to adjust composition teaching. Teachers are crucial in helping kids learn to write; teachers’ teaching behavior of giving high-quality written feedback to writing can support the teaching practice of formative evaluation. The progress of students’ writing performance in a school year is closely related to writing feedback [6]. The logical underpinning of composition evaluation is the composition evaluation instrument, an important guarantee for scientific composition evaluation, and a necessary tool for carrying out composition evaluation activities.

There are some flaws in the existing Chinese composition evaluation standard. For example, some scholars believe that the criterion for evaluation is a lack of refinement [7]. Only by making the evaluation standard more standardized and detailed, can the composition evaluation play its due value. The researchers pointed out that many composition evaluations only pay attention to results, not process evaluation and so on. The current evaluation tools have certain pertinence and completeness, but they only evaluate and analyze the overall structure of the composition, lack the evaluation of students’ composition thinking, and give corresponding guidance and suggestions to improve the writing level [8]. The following are the study’s primary contributions.

(i) The automatic evaluation system, its needs, and problems have all been discussed in detail.

(ii) The fundamentals of cognitive style, deep learning, and how to use automatic classroom attendance to its full potential.

(iii) Face detection and recognition rate (FD and FR), head posture recognition rate (HPR), head recognition rate (HR), expression recognition rate (ER), and human posture recognition rate (HPR).

The remainder of the research article is organized as follows, Section 2 will describe the relevant works that are required to comprehend the proposed strategy. The entire system design approach will be discussed in Section 3. Similarly, Section 4 discusses the model construction employed in this study as well as the system test environment. Finally, Sections 5 and 6 detail the analysis of the results as well as concluding observations.

2. Literature Review

Geoffrey Hinton and his team won the 2012 ImageNet image identification competition using the deep learning model AlexNet. Researchers led by Stanford University Professor Wu Enda and world-renowned computer expert Jeff Dean developed the deep neural network (DNN) in the same year and saw incredible improvements in image recognition accuracy, dropping it to 15% in a test using ImageNet Applications for image recognition utilizing deep learning are likely the most advanced. An image processing system based on a convolutional neural network can effectively reduce overfitting. When the neural network was trained using GPU acceleration, it was able to identify a wide range of images faster and more accurately [9].

Embedded deep learning frameworks can run many networks with relatively low computing requirements, thanks to the development of related technologies. The Tengine deep learning framework and the open source wine development tool set are used in this article. Use MobileNet-V2 face detection and image interception, VGG-16 head pose recognition, head up detection, and expression detection for classroom behavior recognition functions like head pose recognition, expression identification and human pose identification for the depth learning algorithm [10].

Because face recognition has relatively higher computing needs, they are not able to run smoothly on the embedded system, and more of the role of the algorithm can play on the GPU of the server [11]. Currently, the face pixels in an image must be at least 64 pixels wide, and preferably 128 pixels wide, in order to be recognized. Generally, during face recognition, the face may be detected for a specific angle. In conclusion, deep learning’s advanced application in the image sector enables the creation of a classroom evaluation system integrating multiple deep learning algorithms [12].

3. Mathematical System Design

The following designs are consists of three paragraph which show us the mathematically design of the system requirement.

3.1. Multitask Convolutional Neural Network (MTCNN)

For rapid and efficient face detection [13], the proposal network, refining network, and output network are the three cascaded networks used by the model. Image pyramid, boundary regression, nonmaximum suppression, and other technologies are also used in the model [14–16]. The proposed $p$-net approach first modifies all the training samples to $L_2 \times L_2 \times L_2$ using three convolution layers $\times 32 \times 32$ feature map. Finally, three different $1 \times 1$ convolution kernel are used to obtain three multidimensional outputs. The structural diagram is inspired from [18] and is illustrated in Figure 1.

For the face, the cross-entropy loss function described in (1) is utilized.

$$L_i^{\text{det}} = -(y_i^{\text{det}} \log(p_i) + (1 - y_i^{\text{det}})(1 - \log(p_i))),$$

where $p_i$ represents the probability of the sample, $x_i$ is the face, $y_i^{\text{det}}$ represents the label, and the value is 0 or 1. The regression loss through Euclidean distance is calculated for each candidate box’s regression using the following sum of the squares loss function, as given in the following formula:

$$L_i^{\text{box}} = \|y_i^{\text{box}} - y_i^{\text{box}}\|_2^2, \quad y_i^{\text{box}} \in \mathbb{R}^4,$$

where $y_i^{\text{box}}$ is the coordinate predicted by the network and $y_i^{\text{box}}$ is the true background coordinate, and they are both quads. For face marker location, we have
\[ L_i^{\text{landmark}} = \left\| y_i^{\text{landmark}} - y_i^{\text{landmark}_i} \right\|^2_2, \quad y_i^{\text{landmark}} \in \mathbb{R}^{10}, \tag{3} \]

where \( y_i^{\text{landmark}} \) is a marker coordinate predicted by the network and \( y_i^{\text{landmark}_i} \) is actual real marker coordinate. Equation (4) displays the training formula for numerous input sources.

\[ \min_{i=1}^{N} \sum_{j \in \{\text{det, box, landmark}\}} \alpha_i \beta^i_j L_j, \quad \beta^i_j \in \{0, 1\}. \tag{4} \]

The whole training learning process is to reduce the previous formula, \( N \) is the quantity of practice samples, \( \alpha_i \) denotes the significance of every work, \( \beta^i_j \) denotes a model label, and \( L_j \) represents the function used to compute the loss. In the MTCNN network, \( \alpha_{\text{det}} = 1, \alpha_{\text{box}} = 0.5, \alpha_{\text{landmark}} = 0.5 \) in P or R net, otherwise, \( \alpha_{\text{det}} = 1, \alpha_{\text{box}} = 0.5, \alpha_{\text{landmark}} = 1 \). Finally, the network will output three sets of data, which are the face/nonface judgment results, the coordinates of the upper left and the five face calibration points’ coordinates, as well as the face frame’s lower right corners [18].

3.2. MobileNet-V2. In MobileNet-V1, the predecessor network of MobileNet-v2, deep separable convolution is used, which significantly increases the neural network’s operation speed while preserving accuracy. A new structure called inverted residuals with linear bottleneck is proposed in MobileNet-V2, which initially utilizes 1x1 convolution to increase the dimension of the input feature map, then uses 3x3 convolution to operate, and finally uses 1 convolution to decrease the dimension. In order to guarantee the model’s ability to express itself, the ReLU activation function is no longer employed after convolution. Instead, the linear activation function is used. Figure 2 depicts its structure.

Given input with the number of channels being Channel1. Through a series of changes, a feature is created that has Channel2 as the number of characteristic channels. The feature processing is as follows.

1. Utilizing global average pooling, we have

\[ z_c = F_{\text{sq}}(u_c) \]

\[ = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j). \tag{5} \]

2. \( W \), a parameter trained to represent the explicit correlation between feature channels, can be used to create weights for each feature channel. The \( z \) obtained by multiplying \( W_j \) by the Squeeze operation is a full connection operation, followed by a ReLU layer, then perform a full-connection operation by multiplying by \( W_2 \). Lastly, by using the Sigmoid function, \( s \) is obtained. The formula is shown as follows:

\[ s = F_{ex}(z, W) \]

\[ = \alpha(g(z, W)) \]

\[ = \alpha(W_2 \delta(W_1 z)). \tag{6} \]

Formula (7) displays the calculation formula for parameter \( W \).

\[ W_1 \in \mathbb{C}_r \times \mathbb{C}_r \]

\[ W_2 = \mathbb{C} \times \frac{\mathbb{C}}{r}. \tag{7} \]

In order to decrease the quantity of calculation and the number of channels, the scaling parameter \( r \) will be used. The final output has the dimension \( 1 \times 1 \times \mathbb{C} \), where \( \mathbb{C} \) stands for the quantity of channels.

Figure 1: Structural of p-net [18].

Figure 2: Structure of MobileNet-V2.
channels. The weight of C feature maps is denoted by the symbol $s$.

3. Utilize formula (8) for operation after obtaining $s$.

$$\bar{x}_c = F_{\text{scale}}(u_c, s_c) = s_c \cdot u_c,$$

where $S_c$ is the weight and $u_c$ is a two-dimensional matrix; multiplying each $u_c$ matrix value by $S_c$ is equivalent.

The input is $128 \times 128$ data and output is 256 floating-point value, which represents the value of the face feature vector’s abscissa and ordinate. Make sure the cosine distance threshold and the cosine distance coordinate value between the output feature vector’s cosine distance and the reference face image's cosine distance are set. If the cosine distance between the detected face and the threshold is less than one, the student is likely to be the one in question.

4. Overall System Scheme Design

The technology is primarily used to track kids’ attendance and behavior in the classroom, form information-based index parameters, and finally get the classroom evaluation results. Firstly, this paper evaluates the system functional requirements in light of the research goals and use case scenarios, and then presents the design specifications. Then, according to the system function and design requirements, the corresponding hardware platform construction scheme and software design scheme are given. Finally, the key problems and solutions of the system are put forward.

4.1. System Functional Requirements Analysis and Design Requirements. The system’s functional needs are examined in accordance with the study’s goals and potential application scenarios.

(i) Automatic attendance. Attendance is not only the basic requirement of teachers for students’ classroom management but also the basis of classroom evaluation. The traditional roll call attendance method is inefficient, and there will be loopholes in the way of random inspection. Realizing “automatic no inductive attendance” is the most basic functional requirement of the system.

(ii) Full-coverage face scanning in the classroom. In the process of classroom attendance, it is vital to look at every student’s face in the room. So that the camera can effectively gather face images of pupils dispersed across the classroom, the system must have an effective full-coverage face scanning capability.

(iii) Face detection and recognition. In the process of class attendance, students’ faces need to be scanned and their identities confirmed. Therefore, the system needs to have reliable face detection and recognition function, be able to cut out the student face image from the student area image collected by the camera, and correctly identify the student’s identity to complete the attendance function.

(iv) Classroom behavior recognition. The recognition of students’ classroom behavior is the premise of classroom evaluation. In the process of classroom behavior identification, students’ classroom behavior needs to be identified, recorded, and analyzed. Therefore, the system needs to have the ability of identifying the classroom’s behavior recognition. So that the system can correctly obtain the students’ classroom behavior recognition results and analyze the results.

(v) According to the system function requirements, the paper plans the following evaluation indicators for the system design and the requirements are as follows.

4.2. Student Attendance. Students’ attendance in class is a fundamental prerequisite for the successful growth of classroom instruction and is a crucial assurance for developing students’ good discipline, learning atmosphere, and good learning habits. Attendance generally includes normal attendance, absence from the class, late arrivals, and early departure. Student attendance is determined by face detection and face recognition. Considering that, the face recognition rate in the classroom can reach more than 90%. The system sets the face detection and recognition rate of 90%. The attendance error is not more than 10%. At this accuracy, because there may be misidentification in each class, students are allowed to submit reports in time to correct the misidentification data within the specified time in combination with the system planning absence notice.

(i) Class attendance rate. Students’ classroom attendance rate is one of the most important indicators of classroom evaluation, which is the direct embodiment of students’ learning attitude and behavior. It represents the design of the study method on the one hand and the caliber of the course’s instruction on the other. According to the attendance rate along with the phased growth characteristics of students, analyzing the key points and difficulties in designing style of study in each stage, formulating and implementing corresponding measures, is a major challenge [2]. In addition, it provides a crucial foundation for assessing the course’s quality.

(ii) Classroom concentration. Whether students listen carefully in class is the classroom concentration, which not only reflects students’ learning attitude but also reflects the attraction of teachers’ class and their teaching effect to a certain extent, which is an important indicator for classroom evaluation. Students’ concentration can be judged by the students’ head posture and facial expressions. For example, if the student faces the platform and looks neutral or happy, the student is attending the class carefully; if the student is not sitting upright and shows a
disgusted or angry expression, the student is not listening carefully.

4.3. System Design. The system adopts the architecture of an edge computing system, which is composed of an edge end, server end, and client. Figure 3 displays the system’s overall structure diagram.

4.4. Workflow of the System. Divide the student seats in the classroom into multiple areas and record the PTZ coordinates (pan-tilt-zoom camera). When the camera is facing each area, the following possibilities exist:

(i) At the beginning and near the end of the classroom, the edge computing and processing center controls the camera to collect images of students in the classroom according to the set coordinate regions, intercepts the students’ faces collected twice, saves them, respectively, to the storage directory of the edge computing and processing center, and sends the intercepted face images to the server through TCPP protocol for face recognition algorithm processing.

(ii) The edge computing processing center identifies heads, head postures, facial expressions, and human postures for each student in the entire classroom, and forms the above behavior recognition results into statistical result files, which are sent to the server for data processing and classroom evaluation after the course.

(iii) The server performs face recognition on the received face image, saves the recognition results to the attendance table of the corresponding course in the database to complete the student attendance, and analyzes and processes the received student behavior recognition results to obtain the quantitative scores of various indicators such as classroom concentration, classroom activity, and classroom link richness. Combined with the classroom attendance rate, the classroom evaluation score is given by systematic weighted calculation, and the relevant indicators and evaluation results are saved in the curriculum evaluation table corresponding to the database to complete the classroom quality evaluation; make statistics on students’ classroom behavior, and save the statistical results to the student classroom behavior statistical table corresponding to the database for teachers’ reference in evaluating students.

(iv) On the client side, by entering the user interaction interface, obtain the attendance and classroom evaluation of the specified course from the server database, and the relevant indicators and evaluation results are displayed in the corresponding position of the interface.

4.5. System Software Scheme Design. The system software design mainly includes the functional software of the edge end, server end, client end, and data transmission software.

4.5.1. Edge End. The edge software mainly completes the collection of facial images of students in various areas of the classroom and the analysis of students’ classroom behavior during class. To realize the above functions, it is necessary to realize the functions of camera pan tilt and focal length control. Specific functional software includes the following.

(i) Pan tilt and focus control. Divide the classroom into multiple areas, adjust the camera angle with pan tilt and zoom through pan tilt and focal length control, and collect the students’ faces in each region of the classroom, so that the system may capture a reasonably clear face image of each student there.

(ii) Face image interception and face detection. After the camera is aligned with the set image acquisition area, the face detection algorithm is used to detect the face of the image, and the successfully detected face image is intercepted as a single picture and saved to the memory at the edge end.

(iii) Classroom behavior recognition. Head, head posture, expression, and human posture are all examples of classroom conduct that can be recognized. To complete the recording and analysis of students’ classroom behavior, the classification algorithm is used at the edge to recognize and classify the students’ posture, their expressions as well as the statistical findings, which are then stored in the memory.

4.5.2. Server Side. The server-side software mainly completes the face recognition of the student’s face image received from the edge end, forms the attendance record, calculates the attendance rate, carries out data analysis, index extraction, and score quantification of the student’s classroom behavior recognition results received from the edge end, and finally gives the statistical results of student’s classroom behavior, classroom evaluation results, and database management. Specific functional software includes the following.

(i) Face recognition. After receiving the student’s face image from the edge end, the server uses the face recognition algorithm to compare the face image with the student’s front photos stored in the server and saves the recognition results to the memory. Based on this, the student attendance table is formed and the classroom attendance rate is calculated.

(ii) Classroom evaluation index calculation. After receiving the statistical result file of student behavior classification and recognition sent from the edge end, the server extracts the index parameters such as student concentration, classroom activity, and richness of classroom links from the file and scores them quantitatively, and then weights and calculates
the comprehensive evaluation value according to the contribution of each index to classroom evaluation.

(iii) Statistics of students’ classroom behavior. The server analyzes the recognition results of students’ classroom behavior collected from the edge end, makes statistics on students’ classroom behavior, and obtains students’ head up rate, class attendance rate, hand raising rate, and sleep rate.

(iv) Database management. A database is needed to manage the basic information of students and courses, student’s classroom behavior, classroom attendance, and classroom evaluation results, to facilitate the storage and query of data. The server enters the results of the face recognition and classroom assessment index calculation after finishing their respective tasks into the attendance table, student classroom behavior statistics table, and classroom evaluation result table in the database, and finally obtains the attendance situation, student classroom behavior statistics and classroom evaluation score of each course.

In addition to the above functional software design, the software design of data transmission needs to be completed at the edge, server, and client to realize end-to-end data transmission. To perform the operations of face detection, the system uses a deep learning algorithm. Face detection adopts MTCNN face detection algorithm and is implemented on Tengine deep learning framework mounted on EAIDK-610 development board; ace recognition adopts MobileNet-V2 algorithm and is realized on Open VINO Development Tool Suite equipped on Intel FPGA accelerated cloud platform: head recognition.

5. System Test

5.1. System Test Environment. This work completes the deep learning-based classroom evaluation system using a webcam and PC. Table 1 displays the hardware test environment specifications for the system’s edge end.

5.2. System Function Test. The deep learning-based classroom evaluation system created in this work should be able to track student behavior statistics, identifying inappropriate behavior in the classroom and evaluating the quality of the learning environment. The aforementioned operations are then each put to the test. The experiment was carried out in the author’s lab. The test should be administered to 10 and 9 participants.

5.2.1. Class Attendance. Before starting the system, first use the onvif equipment debugging tool to debug the angle and focal length of the camera, so that the camera can cover each acquisition area in the classroom from different angles and record the values corresponding to the PTZ angle coordinates and focal length multiplier of each area. Write the PTZ coordinates and focal length times into the program in the order of the acquisition area, in order for the camera to aim at each acquisition area in the predetermined order. The paper turned on the system after positioning the camera in a lab corner. The camera captures an image of an acquisition area, and in this image, uses a green box to frame the face area of all students. Within 3 to 5 seconds after the camera stays in this area, the student’s face image in the green box in this area is intercepted and the size is modified to $128 \times 128$ and saved to the storage directory of EAIDK-610 in order. The server side runs the data receiving program to make the
server side to receive the data sent from the edge side. After collecting the face images of all students in the classroom, the edge end runs the data sending program to send all face images to the storage directory of the server end. After receiving the face image, the server will recognize the face image in order and the face recognition results, including the student’s ID and recognition time. Finally, upload the attendance results to the database.

### 5.3. System Performance Test

#### 5.3.1. Face Detection Test

This paper uses the TCNN network to test face detection. Table 2 displays the test results.

<table>
<thead>
<tr>
<th>Category</th>
<th>Parameter</th>
<th>96%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tests</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Number of successful detection</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Detection rate</td>
<td>90%</td>
<td></td>
</tr>
</tbody>
</table>

#### 5.3.2. Face Recognition Test

This paper uses mobilenet-v2 as the backbone network for the face recognition test. Table 3 displays the test results.

<table>
<thead>
<tr>
<th>Category</th>
<th>Parameter</th>
<th>94%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tests</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Number of successful detection</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Detection rate</td>
<td>94%</td>
<td></td>
</tr>
</tbody>
</table>

#### 5.3.3. Head Posture Recognition Test

This paper uses the vgg-16 network for the head pose recognition test. Table 4 displays the test results.

<table>
<thead>
<tr>
<th>Category</th>
<th>Parameter</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tests</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Number of successful detection</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Detection rate</td>
<td>80%</td>
<td></td>
</tr>
</tbody>
</table>

#### 5.4. Summary of Test Results

This study’s classroom assessment system meets the design requirements in terms of FD, FR, HPR, HR, ER, and HPR. The stability of the system is up to par with the original design intent. The technical specifications are as shown in Table 5.

### 6. Conclusion

This paper details the entire development process for a classroom evaluation system based on DL. This paper begins with a summary and presentation of the problems with the current system of classroom evaluation and concludes with a presentation of the research objectives. Using the Intel FPGA accelerated cloud platform, this research implements services such as face recognition and classroom assessment index calculation. The user interface design is implemented on the client using a PC as the system client. The aforementioned procedure ensures the execution of classroom attendance and the recognition of appropriate classroom conduct for students assigned fixed seats in the classroom, as well as the acquisition of the classroom evaluation function, and experimentally demonstrates the superiority of our method.

### Data Availability

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

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References


