Hindawi
Advances in Mathematical Physics
Volume 2021, Article ID 2209526, 12 pages
https://doi.org/10.1155/2021/2209526

Research Article

A Three-Dimensional Anisotropic Diffusion Equation-Based Video Recognition Model for Classroom Concentration Evaluation in English Language Teaching

Yanghong Liu[1] **and Jintao Liu**[2]

[1]*School of Foreign Languages, Xinyang College, Xinyang 464000, China*
[2]*Academic Affairs Office, Xinyang Normal University, 464000, China*

Correspondence should be addressed to Jintao Liu; jtliu@xynu.edu.cn

Received 26 October 2021; Revised 29 November 2021; Accepted 3 December 2021; Published 20 December 2021

Academic Editor: Miaochao Chen

In this paper, a three-dimensional anisotropic diffusion equation is used to conduct an in-depth study and analysis of students' concentration in video recognition in English teaching classrooms. A multifeature fusion face live detection method based on diffusion model extracts Diffusion Kernel (DK) features and depth features from diffusion-processed face images, respectively. DK features provide a nonlinear description of the correlation between successive face images and express face image sequences in the temporal dimension; depth features are extracted by a pretrained depth neural network model that can express the complex nonlinear mapping relationships of images and reflect the more abstract implicit information inside face images. To improve the effectiveness of the face image features, the extracted DK features and depth features are fused using a multicore learning method to obtain the best combination and the corresponding weights. The two features complement each other, and the fused features are more discriminative, which provides a strong basis for the live determination of face images. Experiments show that the method has excellent performance and can effectively discriminate the live nature of faces in images and resist forged face attacks. Based on the above face detection and expression recognition algorithms, the classroom concentration analysis system based on expression recognition is designed to achieve real-time acquisition and processing of classroom images, complete student classroom attendance records using face detection and face recognition methods, and analyze students' concentration from the face integrity and facial expression of students facing the blackboard by combining face detection and expression recognition to visualize and display students' classroom data for teachers, students, and parents with more data support and help.

1. Introduction

Current face recognition systems are not strong enough in effectively distinguishing between live and nonlive face data, i.e., they encounter challenges in determining whether the face image acquired by the system is from a real legitimate user or a forged face, and face live detection comes into being [1]. The task of face live detection is to determine whether a face is live or not by analyzing the acquired face image and to guarantee the safe and reliable operation of the face recognition system. Face live detection is an interdisciplinary research area that has received the attention and research of many researchers in related disciplines. The research content of related disciplines has also been applied to the research of live face detection technology, which has greatly misled the development of the technology. At the same time, he will also make some serious observations of spontaneous behavior and may also be accompanied by some actions related to learning during concentration. Under the research of the research group, through the observation and evaluation of the concentration of students in a short time frame, combined with related algorithms, a concentration evaluation curve of a classroom cycle can be given. Even with some changes in the human face due to

human growth and development or some external factors, it still has a lot of information to be maintained [2]. On the other hand, the fierce development of information technology and the widespread use of the internet have made artificial intelligence gradually change our lives in every aspect such as home, healthcare, travel, manufacturing, and education. Artificial intelligence technology also provides the technical basis for student concentration evaluation. Through machine vision and artificial intelligence algorithms, an automated learning status evaluation method can be implemented to assess the whole student and the whole process, assisting teachers to recognize and master the learning status of students, adopt targeted teaching methods, and improve personalized training of students.

The traditional evaluation of student learning using grades as the main indicator in teaching and learning both hinders the need for the individual development of students and is extremely one sided. The evaluation of student learning in teaching should be more process oriented. Through the changes of students' expressions and head postures in the classroom learning process, we can judge the changes in student's concentration in the learning process and set up real-time prompting means or postclass feedback so that students can understand the problems in classroom learning and improve them in time; at the same time, we can use the concentration evaluation system based on facial expressions and head postures to provide reference basis for teachers' teaching evaluation and teaching methods [3]. It also makes it possible to rationalize the teaching improvement plan according to students' characteristics. In the traditional classroom, teachers can only interpret the information conveyed by students' facial expressions and postures through close observation, and due to limited energy, they cannot take care of all students' emotions at the same time, and the quality of the classroom is also limited by teachers' ability to interpret students' facial expressions and postures, so they cannot make timely adjustments to the classroom teaching [4]. By using a system based on expression and posture recognition to determine students' concentration, we can record students' emotional and postural changes in real time and determine students' learning status and emotional difficulties to a greater extent, thus helping teachers to adjust the teaching progress according to most students' emotional and postural changes promptly and improve the efficiency of teaching. Students can also analyze their learning problems at any time according to their own emotions and posture changes, making classroom evaluation more accurate and intuitive [5].

The classroom pictures collected in real time can be recorded by face detection and face recognition methods to record the number of students attending the class and the corresponding time, so as not to affect the classroom order, without contact and interaction, and complete the class attendance of students. Through face detection and expression recognition, we can analyze the different states of students in the classroom, count the students who look up and listen to the lecture with serious expressions, and detect the students who look down, laugh, and play. By capturing the complete degree of students' faces and expression changes, statistics and analysis of students' classroom concentration are conducted. The final visualization of attendance data, concentration data, and classroom interaction data allows teachers and parents to understand the real performance of students promptly and helps to jointly improve the efficiency of students in class. To integrate the above issues, it is necessary to develop a system to assist teachers in classroom attendance and classroom concentration analysis with the characteristics of classroom teaching management and provide timely feedback to teachers, students, and parents so that the three parties can jointly supervise and improve the quality of teaching, forming a positive and virtuous cycle.

2. Current Status of Research

Face detection is a mature biometric technology that is nonintrusive, friendly, and concurrent by determining face feature information for identification. Face detection has been researched since around the 1970s and has been developed with some practicality over the long years. Traditional face detection methods are mainly classified as knowledge-based and statistics-based, where feature-based face detection methods have been successfully applied [6]. Knowledge-based face detection methods mainly include template matching-based face detection methods, skin color-based face detection methods, and shape-based face detection methods. Using a priori knowledge and rules, the face represented as the result of a combination of local organs [7]. For example, the face is roughly axisymmetric with eyebrows and eyes in the upper part and lips in the lower part and close to the nose; this a priori knowledge helps to quickly sift through nonface regions. It is possible to quickly find regions in the image that match the face contour and then further identify and judge the detected face regions, saving the time of face detection [8]. The images for fake video attacks are then obtained by secretly recording images and videos of the target of the attack without their knowledge or by finding videos on websites. This approach is more disorienting than 2D image attacks. Because videos contain biometric features such as movement information and changes in facial expressions that real faces have, faked video attacks are more difficult to discern than images [9].

The image texture-based detection method is mainly based on the analysis of the image [10]. The real face captured by the same device is compared with the forged face captured with that device, there is a loss of detail information in the image of the acquired forged face, and the difference in detail produces a difference in the texture of the face image, i.e., the image of the forged face captured under the same shooting environment and conditions has a lower image quality [11]. As an example of the attack method of printing a photo of a face, the attacker prints the image on the photo and then uses the photo to attack the face recognition system [12]. The movement of the central region of the human face produces a greater distance than the peripheral region, i.e., the real face has a three-dimensional structure, and the movement information produced by the parts at different distances from the camera is different. And the

motion information produced by different parts of the face photo is the same [13]. The method uses the motion information of multiple regions of the face to decide. The real face structure is three dimensional, and during motion, the image records the motion pattern of each region, and the motion pattern of different regions can be estimated using the optical flow method, which in turn discriminates the real face image from the faked face image [14]. Set up real-time prompts or after-class feedback so that students can self-understand the problems in classroom learning and improve in time; at the same time, use the focus evaluation system based on facial expressions and head posture to provide reference for teachers' teaching evaluation and teaching methods.

Classroom expression recognition, on the other hand, is a recognition technology that judges students' expressions by recognizing their microexpressions in the classroom learning process. With the rapid development of artificial intelligence technology and facial recognition technology, it gradually begins to be applied in education, but the application area mainly revolves around online learning and intelligent classroom, and the use for the traditional classroom is still relatively small. Looking at the current stage for the application of facial recognition in education evaluation, we can divide the current stage of classroom expression recognition and evaluation into two ways. The first one is the recognition of classroom expressions by key facial parts, through the combination of key features or characteristics of the face, to judge the expression situation of the classroom. The students' classroom expressions were judged by determining the eye openness by selecting the aspect ratio of the eye area and classifying them into two categories: focused and unfocused. However, the assessment of expression by judging only the percentage of opening and closing is less credible.

3. Analysis of a Three-Dimensional Anisotropic Diffusion Equation for Video Recognition Model in ELT Classroom Concentration Evaluation

3.1. Three-Dimensional Anisotropic Diffusion Equation Video Recognition Model Design. Diffusion is a concept in physics that refers to the transfer of molecules of a substance from a region of high concentration to a region of low concentration until the overall region is uniformly distributed. The rate of diffusive movement is proportional to the concentration gradient of the substance. By slowing down the image diffusion rate in the region with a large gradient value and speeding up the image diffusion rate in the region with a small gradient value, it is possible to retain the details and boundary information based on the removal of image noise, so they proposed the P-M equation, which is the anisotropic diffusion equation.

$$I_t = \operatorname{div}\left(a(x, y, \mathrm{t})\Delta I^2\right). \tag{1}$$

Suppose that if the location of the boundary of each region of this image is known when time is then it is desired to make an unbounded region smoother so that a bounded region can retain more information. This can be achieved by setting the conduction coefficient to 1 in the interior of each region and 0 at the boundaries. Each region will then become smoothly blurred separately within each region, with no interaction between the different regions, and the final image region boundaries will remain clear, thereby reducing the amount of calculation and improving the accuracy rate, but this way of directly merging information may easily lead to the failure of key information to be recognized, especially in facial expression recognition; there may not be a particularly big difference between similar microexpressions, if the gradient magnitude of the image pixel change is used as a conduction coefficient in the diffusion equation.

$$a(x, y, t) = g\left(\left\|\nabla I\left(x^2, y^2, t^2\right)\right\|\right). \tag{2}$$

Boundary enhancement and reconstruction of fuzzy images can be achieved by either a high-pass filter or by running the diffusion equation in the inverse direction. But by choosing a suitable conduction coefficient for the image gradient, it is possible to use anisotropic diffusion to achieve forward enhanced image boundaries. In this paper, an objective analysis method is used for the image results. Since the results are judged subjectively mainly by observing the image detail information and image quality and its criteria are influenced by the observer's state, they are not used [15]. The objective criteria are compared concerning image contrast, image standard deviation, and image mean gradient, the processing time of the i5 processor with a main frequency of 2.6 GHz is added for reference, and this reference time also includes the time of reading the image cache at the beginning of the project. The image contrast is calculated as shown in

$$A = \sum \delta(x, y)^2 \cdot P_\delta\left(x^2, y^2\right). \tag{3}$$

This indicator can reflect the layer difference between the darkest and brightest areas in the image, and for darker images, increasing this indicator can reflect the effect of image enhancement well. The image grayscale means the value is calculated as follows.

$$G = \lim_{M,N\longrightarrow\infty} \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} I\left(x^2, y^2\right). \tag{4}$$

The distance of faces in real scenes often appears to be large near and far from the characters, making the face size in the image inconsistent, and the changing multiscale faces will affect the accuracy of face detection. To enhance face detection, it can be broadly divided into two ways: image pyramid model and feature pyramid model. Image pyramid can get images of different resolutions by downsampling the images, forming a pyramid image structure from coarse to fine, and generating feature maps on images of different sizes separately, which improves the accuracy of the model to some extent. The original image is scaled, and the detected candidate frames are then reduced to equal scale to calculate

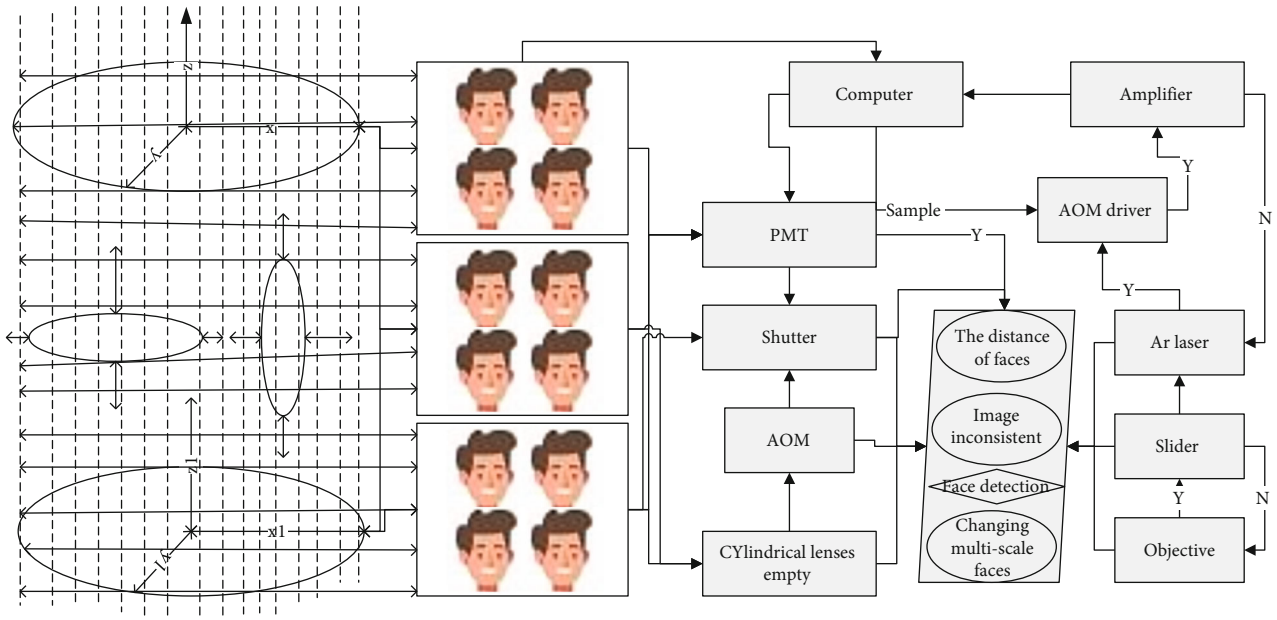

Figure 1: Three-dimensional anisotropic diffusion video recognition model.

the corresponding spatial position existing distortion. This sampling method does not get new semantic information and can only extract a single level of different resolution information, the learned features do not have a real scale change, and the sampling operation is relatively time-consuming; there are many repeated calculations, which affects the detection speed of the model, as shown in Figure 1. Up to 1080 registers can be set up, and the method of using shift registers at the same time will lead to a lot of waste of register resources. In particular, in the process of multiple iterations, the line buffer is used more and the resource occupancy is large. The image format used in this article is 1280×720; the pixel data of one row exceeds the maximum amount of the shift register.

The distance of faces in real scenes often appears to be large near and far from the character, making the face size in the image inconsistent, and the changing multiscale faces affect the accuracy of face detection. Common convolutional neural network-based face detection methods often use the highest layer feature maps for classification and candidate frame regression; although the high-level features contain rich semantic information, after several layers of convolution and pooling operations, the input to the output feature maps is constantly transformed between, the fine-grained information is gradually lost and the resolution becomes blurred, and the fixed-size feature maps are poor for small-size face detection.

$$I^2(x,y,t) = I_0\left(x,y^2\right) - t\frac{\partial I(x,y,t)}{\partial t^2}_{t=1},$$
$$\begin{cases} \dfrac{\partial I(x,y,t)}{\partial t^2} = -\nabla I^2, \\ I(x,y,0) = I^2\left(x,y^2\right). \end{cases} \tag{5}$$

To enhance the face detection capability, it can be broadly divided into two ways: the image pyramid model and the feature pyramid model. Image pyramid can get images of different resolutions by downsampling the image, which consists of a pyramid image structure from coarse to fine and generates feature maps on images of different sizes, respectively, which improves the accuracy of the model to a certain extent. The original image is scaled, and the detected candidate frames are then reduced to equal scale to calculate the corresponding spatial position existing distortion [16]. This sampling method does not get new semantic information and can only extract the single level of different resolution information, the learned features do not have real scale change, and the sampling operation is time-consuming; there are many repeated calculations, which affects the detection speed of the model.

$$h(i) = w \cdot \left[1 + g\left(i^2\right)\right],$$
$$\phi(x,y,t) = \left[g(\nabla I) + h\left(\nabla I^2\right) - \nabla I^2(x,y,t)\right]. \tag{6}$$

During our research, we found that in general, it is not possible to compute the attentional deflection angle of the gaze of an observer directly from a planar image, because its observation target cannot be determined. So, we decided to transform the gaze deflection problem into a problem of calculating the deflection angle of a triangular mapping by using some of the head key points and corresponding them to a triangular plane. After that, the attentional range of the observed person is calculated in conjunction with its relative position. In this paper, we use 70 key coordinate points in head pose feature point extraction in face key point recognition. Here, we choose four coordinate points as the target coordinate points for

head pose measurement, where point 27 is the brow coordinate point, point 30 is the nose tip point, point 36 is the external eye angle point of the left eye, and point 45 is the external eye angle coordinate point of the right eye.

$$\begin{aligned} l &= a \ln (1 - i), \\ a_{ij} &= \frac{1}{\sqrt{n+1}} \left(f_i, f_j \right). \end{aligned} \tag{7}$$

The eyes and mouth, two of the most important features of the face, are the most visible, so it is only logical that the expression of the eyes and mouth should be judged by fusing their features. In the case of the eyes, the state of the eyes is recorded by judging their opening and closing, and the same is true for the mouth. By combining multiple situations of the mouth and the eyes, we can analyze and judge different expression situations. Therefore, in classroom expression recognition, we can judge three kinds of classroom expressions: concentration, fatigue, and normal, by the combination of both. The video collection time is about 15 minutes. By decomposing the video to obtain about 27,000 pictures of each student's head posture and facial expressions, the facial expressions and head postures can be analyzed. The concentration analysis is performed in units of 30 seconds. Calculation of degree scores: through feature detection, the position of human eyes and nose is located and marked by coordinates, and the change of the coordinate points of human eyes and thus the change of vision is determined, and thus the expression situation in the classroom is estimated. By defining classroom expressions as focused and unfocused and thus determining the images in the classroom through the trained model, the classroom expressions are finally determined.

$$\begin{aligned} M_s\left(A'\right) &= \sigma\left(\text{avgpool}\left(A'\right), \text{minpool}\left(A'\right)\right), \\ e_2^u &= \sup\left\{ \lim_{K \longrightarrow \infty} \sum_{K \epsilon J} (u(K) + u_K) v\left(K^2\right) \right\}. \end{aligned} \tag{8}$$

An active state of learning also means that the student is focused on the content now, and the focus of this paper is on some external emotional expressions that change less intensely in response to the learning content within a short period. For example, students' expressions change when they are concentrating, their state of pleasure when they are seeking knowledge, and their expressions of understanding and disagreement when they are inquiring. When students are in a focused state of learning, their range of vision combined with changes in head posture will change with the point of focus, as well as some spontaneous behaviors of careful observation, which may be accompanied by some movements related to learning when focused. Under the research of the group, the observation and evaluation of students' concentration in a short time range combined with relevant algorithms can give a concentration evaluation curve of a classroom cycle, as shown in Figure 2.

There are many ways to evaluate students' concentration through classroom expressions, some evaluate concentration based on eye opening and closing or sight tracking, but their evaluation reliability is low due to single feature points; some studies evaluate concentration through the combination of multiple features of the face, but the expression categories are small and not comprehensive; some studies evaluate concentration through deep learning methods; although the recognition effect is good, the judgment for expressions is easy to be wrong. But the judgment for expressions is easy to make mistakes [17]. And through the definition of classroom expressions, we can find that their classroom expressions have nonconscious and unconscious characteristics, which is very similar to the definition of microexpressions. Therefore, in classroom expression recognition, we will have better results by using the recognition of microexpressions to identify and classify classroom expressions.

3.2. Analysis of ELT Classroom Concentration Evaluation. Head pose estimation used to predict head rotation by analyzing faces in digital images and head pose estimation based on different features can be broadly classified into two categories: methods based on face geometric features and methods based on texture features [18]. Head pose estimation based on face geometric features usually requires high image resolution, and the algorithm uses a feature model composed of many feature points about facial expressions to make inferences about head pose by computing the difference between one pose and another. The method relies on the accurate detection of face feature points and high-quality image resolution to provide accurate estimation results. How to model is one of the most important steps in face geometric feature-based methods and there are geometric methods that are more used nowadays. The geometric approach for head pose estimation uses head shape and local features to estimate the pose, where a total of five key points of the face are used, namely, the eye corner points (the outer corner points of the two eyes), the mouth corner points (left and right two), and the nose tip points; the facial symmetry axis can be found by connecting the left and right two outer eye corner points, the two mouth corner points, and connecting the midpoint of these two connecting lines by face geometric features; the distance from the nasal tip point to this midpoint connection line constitutes a certain angle to the midpoint connection line, and also, the nasal tip to each eye corner point constitutes an angle, and these angles are also of great reference value in this paper. Assuming a fixed ratio between these facial points and a fixed length of the nose, the facial orientation can be determined from the three-dimensional angle of the nose under weak perspective geometry.

$$\begin{aligned} A' &= M_s(A) \times A^2, \\ F_{K,S} &= \int_S \nabla u \cdot k^T n K^2 S ds. \end{aligned} \tag{9}$$

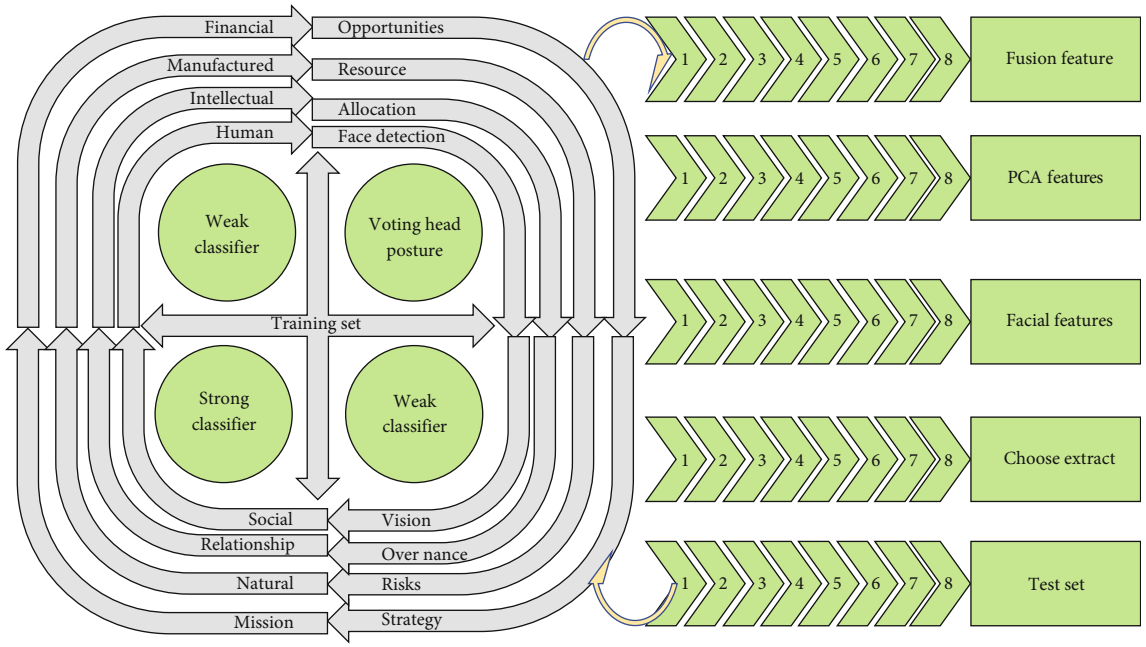

Figure 2: Integrated classifier framework.

The experimental procedure of the geometric method is simple and quick. Only a few facial features are needed to obtain a roughly complete estimate of the head pose. However, the shortcoming of this method is the lack of detection of feature points with high precision and accuracy, where high precision detection refers to the processing of edges or missing features. In addition, there is a frequent situation where some facial features may be obscured due to some unavoidable factors, such as when a person wears glasses that obscure the corners of his eyes. Face texture feature-based methods usually use the texture features of the whole face to estimate the head pose. In this method, most of the studies for the head pose problem can be replaced by classification or regression problems [19]. The algorithm first needs to extract the features in the image that are related to the head pose, and secondly, the head pose can be estimated by some classification algorithm that classifies or regresses the coordinate values of these corrected feature points. These classification algorithms in head pose estimation will build a model from a set of labeled training data, thus providing a discrete pose estimate for a new data sample. At the same time, students who bow their heads and laugh and play can also be detected. Through the captured students' facial completeness and facial expression changes, statistics and analysis of the students' classroom concentration are carried out. The advantage of this method is that it makes better use of the data information from the facial region, as shown in Figure 3.

This module is divided into two sections: grade view and class analysis. If students have doubts about the scores or details of the class, they can choose to view the classroom analysis to see their results and the specific content analysis of the teacher system to understand and correct. The data

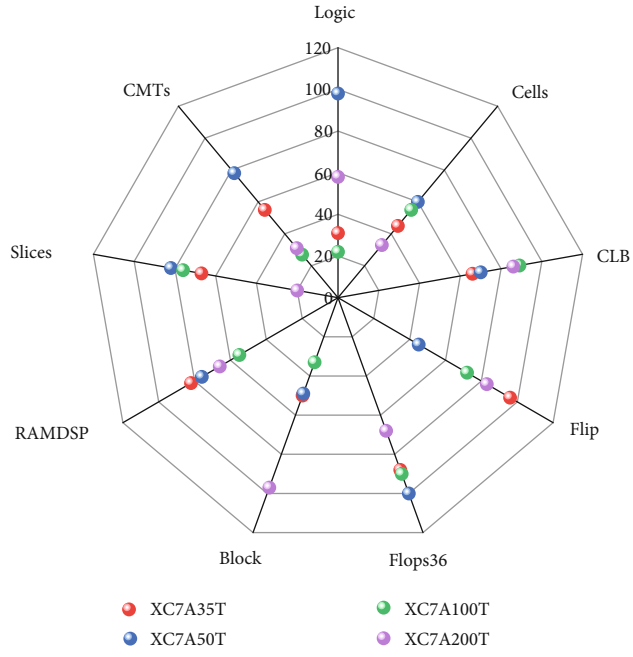

Figure 3: Comparison of internal resources of the chip.

storage layer is located on the server side, and its main function is the storage and management of system data. In this system, we simply divide the system database into user information database, video resource database, and resource database; the user information data is mainly the information related to the registered users; the video resource database is the storage of recorded classroom video resources,

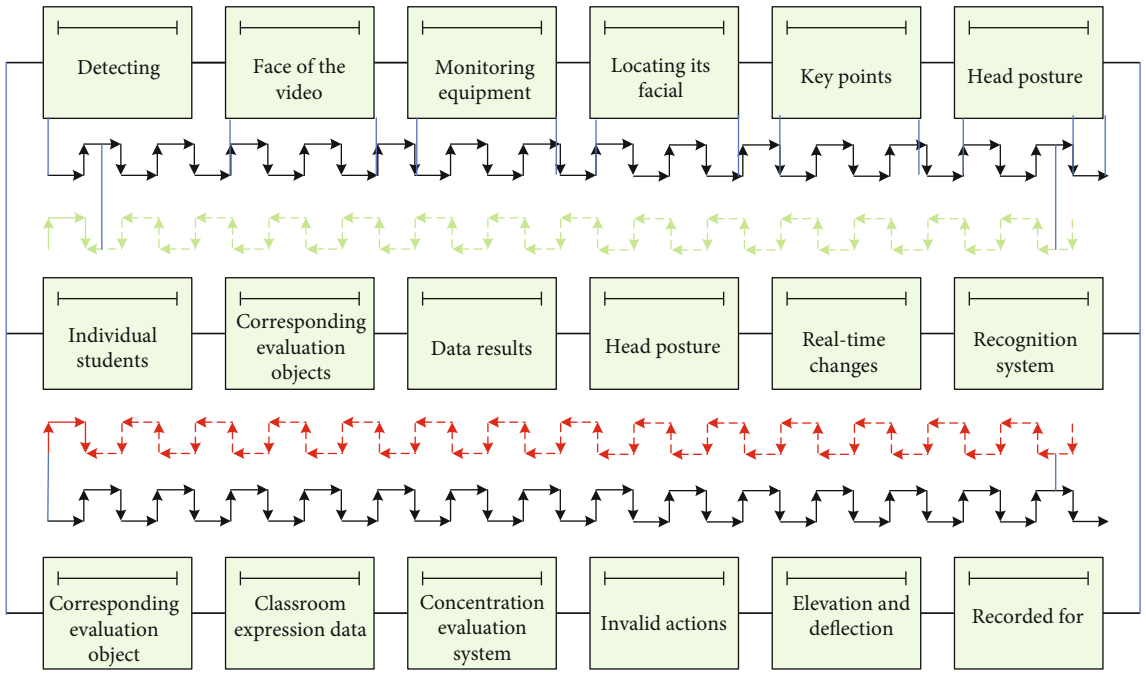

Figure 4: VGA timing.

frame-by-frame picture resources. The resource database is the recognition data of expressions and head postures and the evaluation data of concentration scores.

$$u_p = \lim_{M \longrightarrow \infty} \frac{1}{M} \sum_{i=1}^{M} u_P^i,$$
$$\alpha\left(k\frac{\partial U}{\partial n} + vun\right) + \beta \nabla u = gk. \tag{10}$$

By detecting the face of the video collected from the monitoring equipment, by locating its facial key points, and then using the head posture recognition system to judge the real-time changes of the head posture of the students in the video, the head posture angles that can be recognized are recorded for the elevation and deflection angles, while the head posture cases that cannot be recognized are recorded as invalid actions for removal. Finally, the data results are imported into the concentration evaluation system for the evaluation of concentration. By selecting the corresponding evaluation objects (individual students or all students), the results of head posture and classroom expression data of the corresponding evaluation objects are integrated and sorted and put into the designed fuzzy synthesis matrix for fuzzy synthesis operation to obtain the concentration scores of the evaluation objects. At the same time, by controlling the acquisition period, the classroom concentration of students in a period and the whole class can be obtained. Finally, the concentration scores of the assessment subjects are presented in the form of tables and line graphs.

Resource search helps teachers and students to select the required course information and view the concentration scores of individual students and all students in each subject; resource selection helps teachers and students to select relevant course resources; and resource presentation is to present the acquired data information, recorded course videos and time images in the form of data, images, or videos. User management allows three types of users in this system, i.e., teachers, students, and administrators, to manage permissions and information so that each type of user can view the relevant information resources they need to view, without interfering with each other and interrelated; data management is to organize and store user data, classroom expression data, head posture data, and classroom concentration data so that they can be easily selected and called by the server, as shown in Figure 4.

As shown in Figure 4, the model introduces an attention mechanism in each convolutional block in addition to the stacking method of convolution in VGG networks; the idea is to make the model pay more attention to the important information while ignoring the unimportant information. The attention mechanism is added because, in convolutional networks, image information is often compressed by pooling to reduce the number of operations and improve the accuracy, but such a direct way of merging information can easily lead to key information not being recognized [20]. Since the subjective judgment of the results is mainly based on the observation of image details and image quality and its standard is greatly affected by the observer's state, it is not used. In expression recognition, there may be no difference between similar microexpressions; so in this case, some key information in the expression images may lead to the model's inability to classify expressions accurately if they cannot be captured adequately. Therefore, this paper introduces an attention mechanism in the model, which is mainly achieved by calculating the importance of each channel in the convolutional layer, filtering out the unimportant

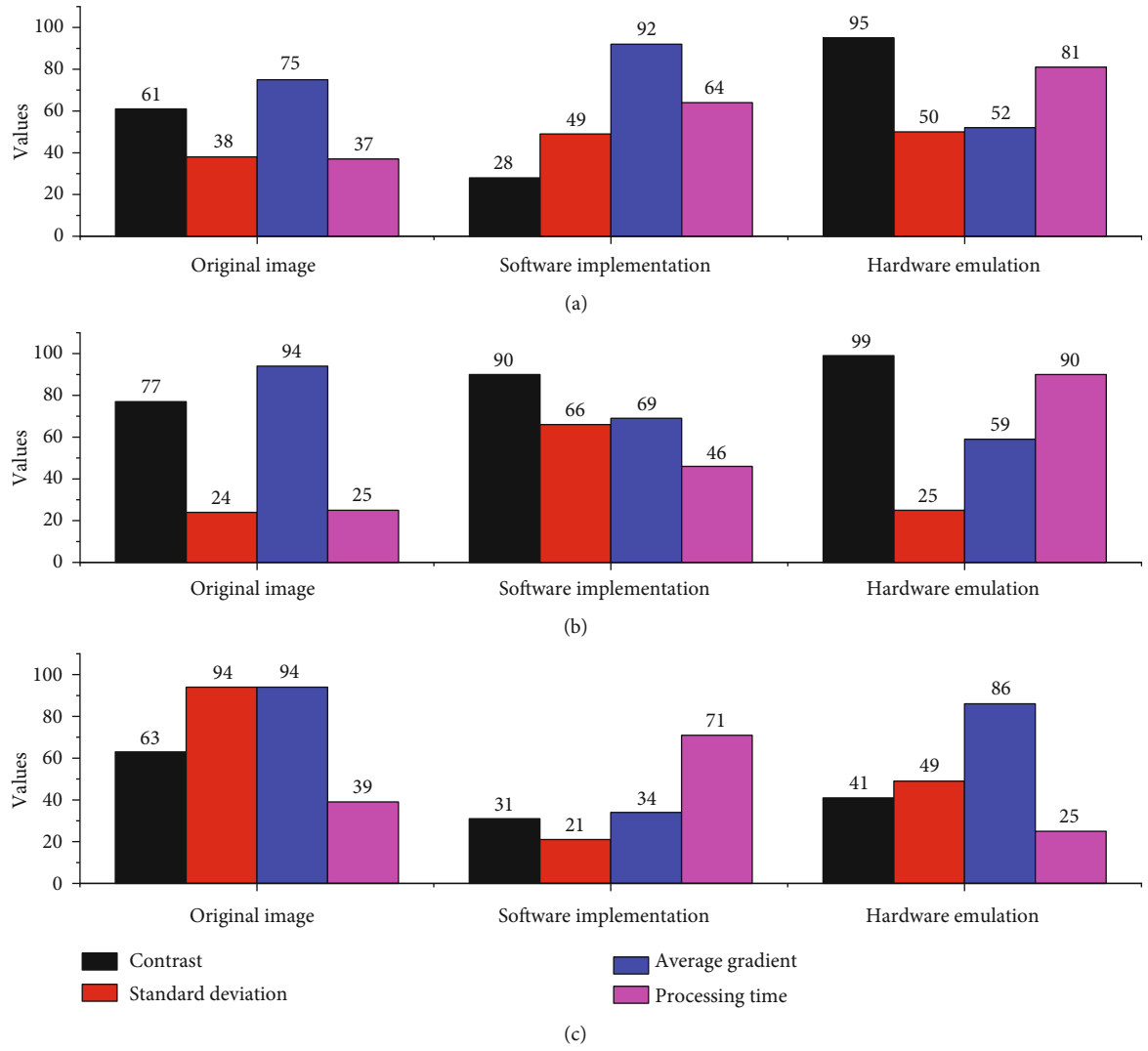

Figure 5: Parameter comparison results.

channels, and then leaving the important ones. The training of neural networks often requires a large amount of data, especially in the training of large networks; if the amount of data is too small and the number of model parameters is large, it can easily lead to overfitting problems and the generalization ability of the model is poor. The total number of samples in the dataset used in this paper is 35,887, which is not a large amount of data; this may make model overfitting and poor generalization problems, and the cost of manual data expansion is large.

4. Analysis of Results

4.1. Anisotropic Diffusion Equation Video Recognition Model Results. Since the structure, texture, and three dimensionality of each region of the face are different and each region is affected by factors such as lighting and environment to a different extent during secondary photography, it is important to focus on regions that are greatly affected by factors such as lighting and environment while reducing the consideration of less affected regions when performing in vivo determination of face images. Real face images and forged face images can show greater differences in the regions that are strongly influenced by factors such as lighting and environment. Again, the face images used in this section of the method still cropped standard face images to exclude as much as possible the effects caused by distracting factors. Six regions in the standard face image are detected and cropped: left eye, right eye, nose, left cheek, right cheek, and mouth.

There are two ways to implement line caching; one is to use shift registers to save data into registers; this method is easy to design, but there is a limit on the number of registers, up to 1080 registers can be formed, while using shift registers will lead to a large waste of register resources, especially in the process of multiple iterations; the line cache is used more, and the resource consumption is large. The image format used in this paper is 1280×720, and the pixel data in one line exceeds the maximum amount of shift registers, making it necessary to have multiple shift registers to achieve this, which also makes this implementation more inefficient. Another way is to implement the line cache

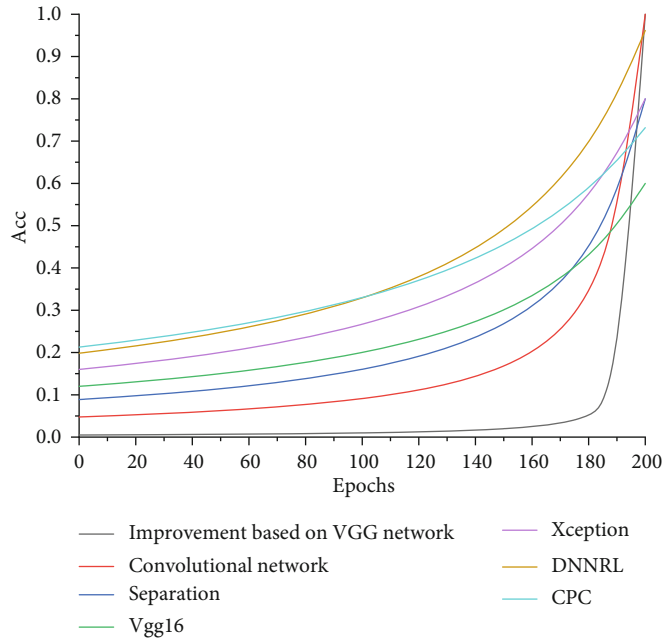

Figure 6: Accuracy curves of the two improved networks and their comparative models on the dataset.

through a combination of FIFO and counter, where the counter is used to count the amount of line cache data, and when there is a line of data in the FIFO, the data from the FIFO is output to the next line cache, and the valid signal of the data is output at the same time. This method can use fewer register resources, and the use of resources transferred to the less used block RAM so that after several iterations of processing, the use of resources is also within the acceptable range, so this paper implements the line cache by this method, as shown in Figure 5.

The hardware implementation in this section will perform FPGA on-chip experiments for the entire system, and the pinning constraints and timing constraints during the comprehensive implementation will be explained accordingly. The final experimental phase of the system design will be conducted in this subsection to demonstrate the feasibility of the hardware implementation of the algorithms and the modular design of the components. The image pyramid can obtain images of different resolutions by downsampling the image, forming a pyramid image structure ranging from coarse to fine, and generating feature maps on images of different sizes, which improves the accuracy of the model to a certain extent. After the synthesis, the entire project implementation and layout wiring are required; at this point, the external pins of the system need to be constrained, and the reference method of the constraint needs to refer to the schematic of the development board and the chip manual. After the system, external pins include pins for communication with the OV5640 camera module, pins for communication with the DDR3 chip, and pins for communication with the HDMI interface.

Without strong light exposure, the test object is affected by natural light and reflects relatively strongly. The overall brightness relatively reduced after processing because of the high contrast of the image due to the strong reflected light. In the processed image, the detail of the QR code area increases, and all text detail information is well preserved, which slightly improves the degree of image information extraction compared to the unprocessed image. Also, there is no tearing of the image during the system implementation, and the results show that the algorithm is well adapted to the scene, as shown in Figure 6. Through a variety of combinations of mouth and eyes, we can analyze and judge different expressions. Therefore, in classroom expression recognition, we can judge the three classroom expressions of concentration, fatigue, and normal through the combination of the two.

Experimental results show that both improved models proposed in this paper outperform the mainstream methods in terms of recognition accuracy. And the improved method based on the VGG model performs better on the dataset compared to the model with deep separable convolution. The accuracy reaches 73.252% on the public test set and 73.846% on the private test set, while the artificial accuracy of this dataset is 65% earth 5%, so the accuracy of the model proposed in this paper has reached a good level. In Figure 6, by comparing the accuracy curves of the two improved methods, it is found that the improved model based on depth-separable convolution converges more easily compared to the improved model of VGG.

4.2. Evaluation Findings. Therefore, to prevent the overfitting phenomenon caused by this situation, the improved model based on the VGG network eliminates the two parametric maximum connected layers, reduces the number of convolutional layers accordingly, and introduces the attention mechanism so that the model can capture the key information to improve the performance of the model. In addition, this section also draws on the idea of Xception network construction to build a separable convolutional model

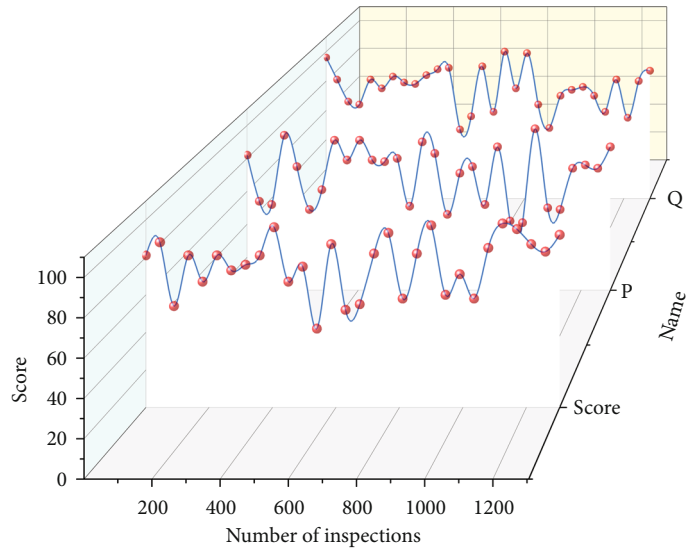

Figure 7: Overall class focus line graph.

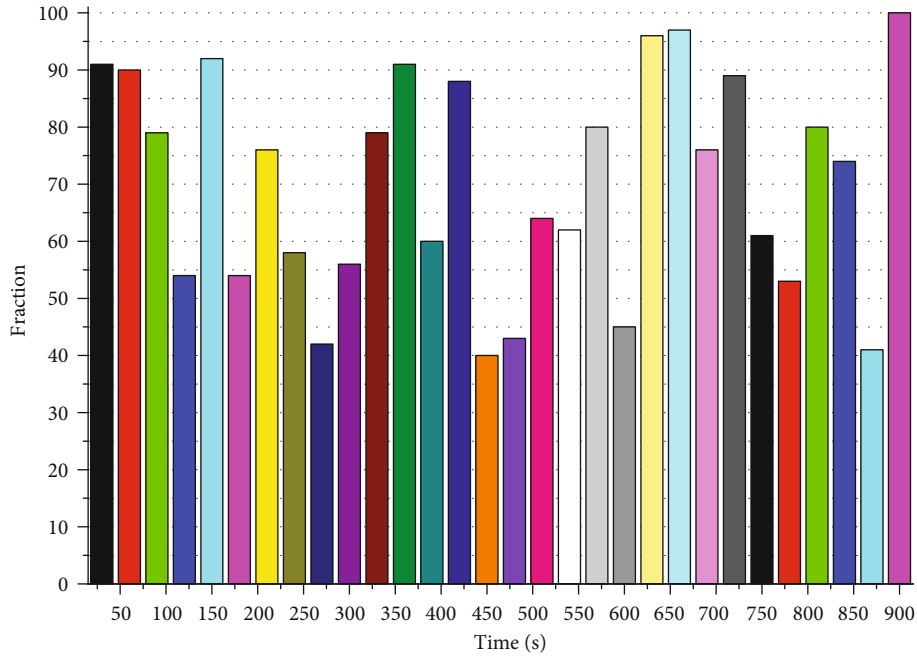

Figure 8: Concentration scores for all students.

with attention mechanism using deep separation convolution and attention mechanism, which can guarantee the accuracy and has good convergence. Finally, the training and optimization of the expression recognition algorithm and the comparative analysis of the experimental results show that the two models proposed in this paper outperform the baseline model on both the public and private test sets, and the improved method based on the VGG model performs better on the dataset compared to the model with deep separable convolution. The accuracy reaches 73.252% on the public test set and 73.846% on the private test set, and the accuracy of the models proposed in this paper has reached a good level from the results of the experiments, as shown in Figure 7.

From the overall class concentration line graph, it can be seen that in the first and second class around 200 tests and after 160 tests, the overall concentration situation of the class was poor in this period because it was tested every second, which means that the overall concentration of the class was low in the first seven minutes of class and the first five minutes of class, but compared to the first class, the overall concentration situation of the class in the second class was better. Then, for the third class, the concentration situation was low in the first 300 times of class and 300 times of the

first 300 times of the next class, which means that the class overall concentration was low in the first ten minutes of class and the first ten minutes of the next class in the third class.

Comparing the overall concentration situation of the class in the three lessons, the concentration situation in the first and second lessons is better than the third lesson. In the third class, the best concentration situation reached 0.7506333, while the best concentration situation in the remaining two classes could reach 0.908922 and 0.9186163, respectively, which may be related to the content and manner of the teacher's class. However, in general, the overall concentration of the class in the first ten minutes of the class is low, although the concentration gradually increases. In the period from 15 to 30 minutes of the class, the concentration situation is the best, the overall concentration of the class is high; especially in the first and second class, the overall concentration of the class in this period is high; in the first ten minutes of the class, the class as a whole is in a low concentration state, and the overall concentration is on a downward trend, as shown in Figure 8.

Similarly, by analyzing the concentration situation of all students correspondingly, it is possible to determine the change of concentration of all students and thus analyze the reasons for the change of students' concentration, and at the same time, through the judgment related to the data, it is possible to evaluate the teacher's teaching situation and teaching style and other aspects. Therefore, by using the students' classroom learning video as experimental data, the concentration of all students obtained and the concentration line graph is drawn for teacher and student analysis. The length of the video is about 15 minutes, the video is decomposed to obtain 27,000 pictures of each student's head posture and facial expression so that the analysis of facial expression and head posture can be carried out, the concentration analysis is carried out in 30 seconds to calculate the concentration score, and finally, the concentration of all students in the whole class is calculated.

According to the curve, we can see that the teacher's teaching style and teaching level are good and can attract students' attention so that they are basically in a state of concentration, but because the teacher's introduction is not good, students' attention is not able to enter the class in time in the beginning; this is something that the teacher needs to improve. As the middle section of classroom teaching belongs to the area where students' concentration is highly concentrated, the more important knowledge points can be introduced during this period to let students learn and investigate better, while students can learn or review their weak knowledge points independently during this period.

5. Conclusion

Through the classification of classroom expression and head posture, a classroom evaluation system based on classroom expression and head posture was designed using a fuzzy comprehensive evaluation algorithm, and the data of both were displayed in the form of scores and interpreted using line graphs to provide a theoretical basis for the subsequent analysis of students' classroom and teachers' teaching. The line graph of classroom concentration of individual students in the experiment was displayed, and the concentration curve was combined to analyze and judge the students' concentration and teachers' teaching; at the same time, the concentration curve of the whole class was displayed for all students, and the classroom concentration of all students and teachers was analyzed and evaluated by combining with the concentration curve, to verify the rationality and relevance of the relevant system, by designing the teacher evaluation. To verify the relevance of the system, a teacher evaluation scale was designed to allow teachers to manually score the five students in the video and compare their scores with the system scores to determine the relevance of the system and finally determine its feasibility and validity. The algorithm acceleration system was modularized, and a series of modules were designed from image acquisition to storage, to processing and finally image display, and hardware simulation experiments were conducted for the important parts of each module to ensure the correctness of its integrated function and acceleration effect, after which the hardware that led the implementation of the modules was carried out through constraints, and the final experimental results were obtained, which proved the good effect of the algorithm improvement and illustrated that the hardware led algorithm.

Data Availability

All data, models, and code generated or used during the study appear in the submitted file.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Sulistyani and R. Riwayatiningsih, "Modeling online classroom interaction to support student language learning," *Ideas: Journal of Language Teaching and Learning, Linguistics and Literature*, vol. 8, no. 2, pp. 446–457, 2020.

[2] D. P. Shea, "Compelled to speak: addressing student reticence in a university EFL classroom," *The Asian Journal of Applied Linguistics*, vol. 4, no. 2, pp. 173–184, 2017.

[3] S. L. Campbell and M. Ronfeldt, "Observational evaluation of teachers: measuring more than we bargained for?," *American Educational Research Journal*, vol. 55, no. 6, pp. 1233–1267, 2018.

[4] T. Ounis, "Exploring secondary teachers' perceptions of classroom assessment in a Tunisian context," *International Journal of Language and Linguistics*, vol. 4, no. 2, pp. 116–124, 2017.

[5] M. A. Dakhiel, "Factors affecting the quality of English language teaching in preparatory year, University of Jeddah," *English Language Teaching*, vol. 10, no. 7, pp. 43–60, 2017.

[6] A. M. AlJaser, "Effectiveness of using flipped classroom strategy in academic achievement and self-efficacy among education students of Princess Nourah Bint Abdulrahman

University," *English Language Teaching*, vol. 10, no. 4, pp. 67–77, 2017.

[7] A. M. Songbatumis, "Challenges in teaching English faced by English teachers at MTsN Taliwang, Indonesia," *Journal of Foreign Language Teaching and Learning*, vol. 2, no. 2, pp. 54–67, 2017.

[8] D. Zou, "Gamified flipped EFL classroom for primary education: student and teacher perceptions," *Journal of Computers in Education*, vol. 7, no. 2, pp. 213–228, 2020.

[9] M. Ahmad, A. Shakir, and A. R. Siddique, "Teacher-student interaction and management practices in Pakistani English language classrooms," *Journal of Language and Cultural Education*, vol. 7, no. 3, pp. 115–134, 2019.

[10] N. Yürük, "Edutainment: using Kahoot! As a review activity in foreign language classrooms," *Journal of Educational Technology and Online Learning*, vol. 2, no. 2, pp. 89–101, 2019.

[11] N. Halwani, "Visual aids and multimedia in second language acquisition," *English Language Teaching*, vol. 10, no. 6, pp. 53–59, 2017.

[12] M. K. Ibrahim and Y. A. Ibrahim, "Communicative English language teaching in Egypt: classroom practice and challenges," *Issues in Educational Research*, vol. 27, no. 2, pp. 285–313, 2017.

[13] M. C. Limlingan, C. M. McWayne, E. A. Sanders, and M. L. López, "Classroom language contexts as predictors of Latinx preschool dual language learners' school readiness," *American Educational Research Journal*, vol. 57, no. 1, pp. 339–370, 2020.

[14] A. Rehman and A. Perveen, "Teachers' perceptions about the use of authentic material in Pakistani EFL classrooms," *International Journal of Language Education*, vol. 5, no. 2, pp. 63–73, 2021.

[15] T. Z. Oo and A. Habók, "The development of a reflective teaching model for reading comprehension in English language teaching," *International Electronic Journal of Elementary Education*, vol. 13, no. 1, pp. 127–138, 2020.

[16] H. Küçükler and A. Kodal, "Foreign language teaching in overcrowded classes," *English Language Teaching*, vol. 12, no. 1, pp. 169–175, 2018.

[17] T. Osman, "The obstacles against the success of 'Suggestopedia' as a method for ELT (English language teaching) in global classrooms," *American Journal of Applied Psychology*, vol. 6, no. 5, pp. 98–105, 2017.

[18] H. R. M. Salleh and H. A. Halim, "Promoting hots through thinking maps," *International Journal of Education*, vol. 4, no. 26, pp. 104–112, 2019.

[19] W. van Peer and A. Chesnokova, "Reading and rereading: insights into literary evaluation," *Advanced education*, vol. 5, pp. 39–46, 2018.

[20] S. Zuhriyah and B. W. Pratolo, "Exploring students' views in the use of Quizizz as an assessment tool in English as a foreign language (EFL) class," *Universal Journal of Educational Research*, vol. 8, no. 11, pp. 5312–5317, 2020.