International Journal of Gaming and Computer-Mediated Simulations
Volume 15 • Issue 1

Blended English Teaching Model in Higher Education School Environment AR Constructive Technology

Jing Yang, Hulunbuir University, China
Lei Bu, Ningxia Medical University, China*

ABSTRACT

China's education modernization requirements clearly suggest exploring new teaching methods to improve teaching effectiveness. Based on this, this article investigates AR construction technology as a blended English teaching model in the university school environment. Mobile terminals are used to build the blended English teaching model, modeling tools are used to achieve split modeling, and the mobile terminals themselves are used to achieve the management and network synchronization work for the cloud classroom. For the gesture recognition model construction, 3D convolutional neural networks are used to separate and optimize the parameters. Finally, experiments are designed to simulate and analyze the AR construct of the blended English teaching model to determine accuracy. The simulation results show that the model can realize the simultaneous display of teacher explanation, and the improved algorithm of gesture recognition is improved in recognition rate, which can improve the effectiveness of blended English teaching.

KEYWORDS

3D technology, AR construction, Blended teaching, gesture recognition

1. INTRODUCTION

As the concept of education is constantly updated and progressing, AR constructive technology fits and meets the requirements of English learning in higher education (Wang, 2022). From the initial network classroom to MOOC, and then to the occurrence and development of blended teaching, all of these tools are the embodiment of the in-depth development of Internet + education. According to a research report on Online Education issued by the U.S. Department of Education in 2010, through experimental research and quasi-experimental research, it is found that blended teaching is better than pure face-to-face teaching or online teaching. At this stage, there are many studies on AR constructive technology in the field of education, mostly in the teaching of spatial three-dimensional related fields, such as medical teaching, industrial design teaching, etc. These professions are quite different

DOI: 10.4018/IJGCMS.334121

*Corresponding Author

from English teaching itself (Zhou, 2020). Teaching English has a profound impact on the learning and progress of each profession (Li, et al., 2020). Existing blended English teaching can recognize the importance of contextualized teaching and learning, and as such, online and offline teaching is constantly being combined and VR constructs have begun to be built (Ma, 2021). However, there are great problems with existing blended English teaching. Many times the concepts will be modeled well and then taught, but if teachers and students are not present at the same time and in the same space, they cannot achieve temporal synchronization and role play is very limited (Leal Filho, et al., 2019). In addition, some teachers did not master the interactive skills of online teaching at the initial stage of online and offline mixed teaching and lacked good and active communication with students, resulting in the inability to solve students' confusion and emotional problems in a timely manner and not achieving effective learning. More than 90% of the students think that the teacher's pre-class preparation is sufficient, 6.3% of the students think it is sufficient. Therefore, it is very difficult for teachers to control the teaching difficulty and class progress under such circumstances(Zhang et al., 2020).

At present, AR technology has entered a relatively mature level and has been widely applied in multiple fields. In English teaching, to fully utilize AR technology, teachers need to further improve their ability to apply information technology, conduct in-depth analysis and reflection on the teaching content, understand the objective laws of student growth, and propose practical and feasible optimization plans. Teachers should also actively participate in case analysis work, deeply consider the application effects of different methods, and provide references for the application of AR technology. Based on this background, this article studies the AR construction technology of the blended English teaching mode in a university school environment, which is divided into 5 main sections.

The first section briefly introduces blended English teaching, the application of AR technology, and the section arrangement of this study; Section 2 introduces the research progress on blended learning, AR application fields, and related technologies both domestically and internationally, and summarizes the shortcomings of current research. Section 3 constructs a hybrid English teaching model based on AR technology and mobile devices. This involves synchronous teaching using 3D Max and other technologies to achieve modeling and supporting online interaction and offline learning. We use interpolation algorithms to achieve synchronous teaching and propose a three-dimensional separable convolutional neural network to improve the algorithm for gesture recognition in model construction. Section 4 conducts a simulation analysis on the AR structure of the hybrid English model constructed in this article to measure the effectiveness of the model application. Section 5 summarizes the entire text. The simulation results show that compared with traditional algorithms, the AR model has better recall prediction and lower error rate, and this algorithm can perform fast computation and convergence, improving gesture recognition rate. The innovation of this paper lies in the design of the hybrid teaching AR architecture, using modeling tools, such as 3D Max, for AR architecture development and taking advantage of the distributed computing of mobile terminals for the overall system design. The 3D analyzable neural network architecture is proposed. In the computation, deep convolution and point-by-point convolution are used to speed up the computation, while the skip connection learning rate is introduced to control the gradient dispersion problem, compressing the AR model to reduce the computation based on ensuring the performance, and the collected dynamic model is simulated and analyzed.

From the perspective of student experience, the application of AR technology is conducive to enriching classroom contents and teaching forms, stimulating students' enthusiasm for learning, identifying and judging students' pronunciation and grammar through the cooperation of various equipment and software, understanding the gaps in students' English learning through their participation in teaching activities, guiding students to actively improve their English ability, and stimulating their subjective initiative in learning.

2. RELATED WORKS

Advances in computer technology, such as cloud computing and 5G networks, have brought profound changes to the field of education. Traditional teaching is no longer adapted to the needs of social development, and many scholars have conducted research on teaching models. For example, Zhang Y, et al. proposed an effective hybrid method TLNNA, based on TLBO and NNA, to optimize 30 well-known unconstrained benchmark functions and four challenging projects in hybrid teaching for industrial optimization (Zhang, et al., 2020). Yanfei Miao, et al. proposed a mobile-based information system that developed a hybrid intelligent teaching aid model for English, integrating English teaching resources, and customizing online learning courses and materials according to students' learning progress, combined with offline teaching aids (Yanfei, 2021). Pezaro, et al. (2022) constructed an offline teaching framework for online and offline teaching in their study. With the gradual maturation of AR technology, research in teaching and learning has started to emerge. In their research on interactive robot learning, Chen, et al. proposed an AR-based interactive robot teaching and programming system to interactively plan or test the path of a virtual robot and proposed to obtain the depth value of the corresponding pixel in a computer-generated image by comparing the depth image acquired by Kinect with the collision-free paths for virtual robots (Chen, et al., 2020). In their study, Cascini, et al. (2020) applied AR technology to the field of product design and found that the quality and novelty of ideas generated by projection-based AR outperformed traditional sessions or handheld display AR sessions. In their study, Xin, et al. developed an optically transparent augmented reality display based on a 3D dynamic integral imaging system that enhanced the 3D image depth range by relaying elemental images at different locations to a microlens array to generate multi-lens image planes to enhance the depth range of 3D images, which can significantly enhance the depth range of 3D reconstructed images with high image quality in micro InIm units (Shen & Javidi, 2018). Wake, et al. (2018) combined AR technology with 3D printing in their study to improve robot-assisted accuracy. (Chu, et al., 2018) used a design science research approach in their study to test and evaluate the development and performance of a mobile BIM AR system. It has been found that a slight modification to the existing information format (2D plan) to include rapid response markers can significantly improve the information retrieval process (Chu, et al., 2018). Ariansyah, et al. used a theory-driven interaction design approach in their study in terms of video versus 3D animation and interaction modes to investigate the impact of different AR modes on user performance and found that different information modes had different impacts, with 3D animation improving task completion time by 14% over video instructions (Ariansyah, et al., 2022). At present, although there is no unified teaching mode for the online and offline mixed teaching reform, there is a unified teaching pursuit. That is, we should give full awareness to the inherent shortcomings of the original teaching mode in colleges and universities and utilize advanced science and technology to change the situation that teachers output lectures unilaterally in the classroom teaching process, resulting in students' low learning initiative and poor input effect. To combine the advantages of the two, we should innovate teaching methods and improve the teaching level of colleges and universities, promoting the modernization of education.

In summary, it can be seen that there are several studies on blended English teaching, mostly focusing on educational theory and how AR teaching is achieved with the help of existing mobile terminals, but there is a lag, synchronous teaching cannot be carried out, students' learning cannot be understood in time, and there is still much room for improvement in the application of AR technology. In terms of technology research, there are many research results in the fields of model construction and technology improvement, but these research results are more often applied in industrial construction, spatial design, and tourism development, but rarely applied to specific teaching. Therefore, it is important to carry out the analysis of the AR construction of the blended English teaching model in the university school environment.

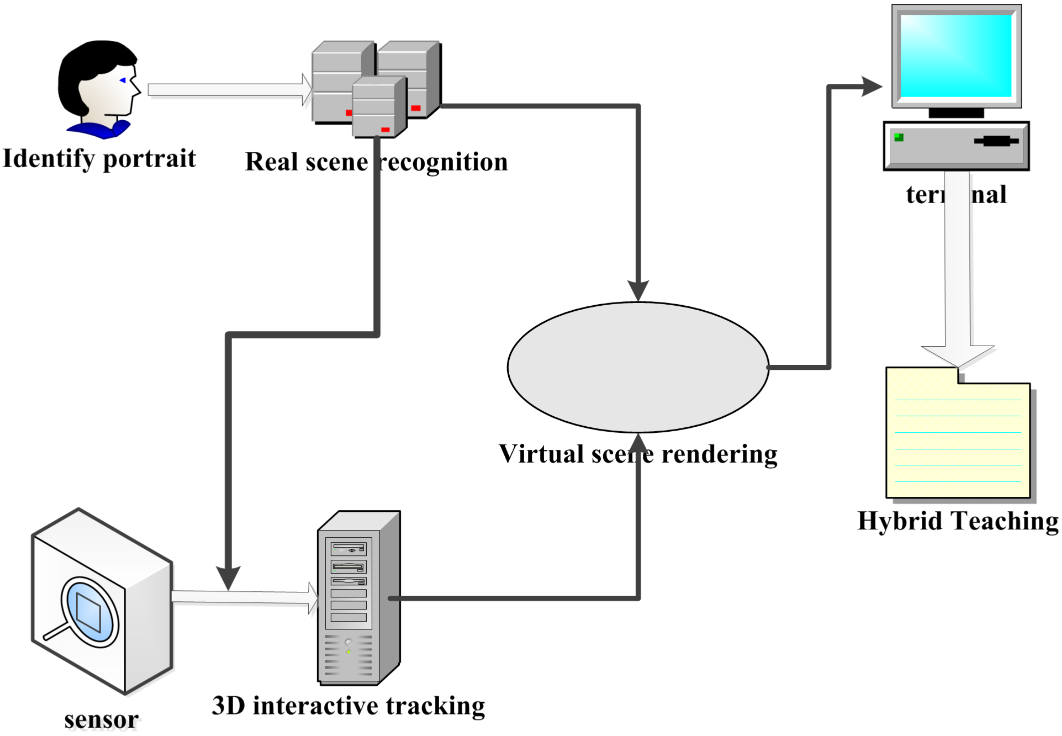
3. MATERIALS AND METHODS

Hybrid English Teaching AR System Design

With the development of computer technology, the English teaching mode also presents diversified development. Restricted by the characteristics of the English subject itself, many knowledge points need to be explained, covering complicated content, which is difficult for students to fully grasp. With the emergence of AR, students can visualize the content and use AR to simulate virtual technology into reality. Through AR technology, teachers can use computers to build various dialogue situations in advance, so that students can communicate with foreigners and classmates in English through AR equipment. In particular, some tourism, shopping, and other situations can be more fully represented by using AR technology. Teachers can also link the database in the virtual situation and use artificial intelligence technology to match the dialogue corresponding to the situation for students, to train students' language adaptability. The use of AR technology for blended English teaching is achieved mainly through computers and mobile devices to form simulated information and simulate the real English teaching environment. This mode of teaching presents a virtual display, with a camera to view the images and superimpose video, text, and pictures onto the real environment. Figure 1 shows the AR hybrid teaching system. In the application, the student side only needs to scan the picture card to be able to learn English.

In the construction of a blended teaching AR for English in higher education, teaching and learning need to be brought into the same environment through the Internet, and teachers need to explain by manipulating physical objects to form models. Therefore, in the design, both teacher and student roles need to be considered, and the corresponding functional modules need to be designed with the teacher side focusing on explaining and the student side on learning. For later iterative expansion of

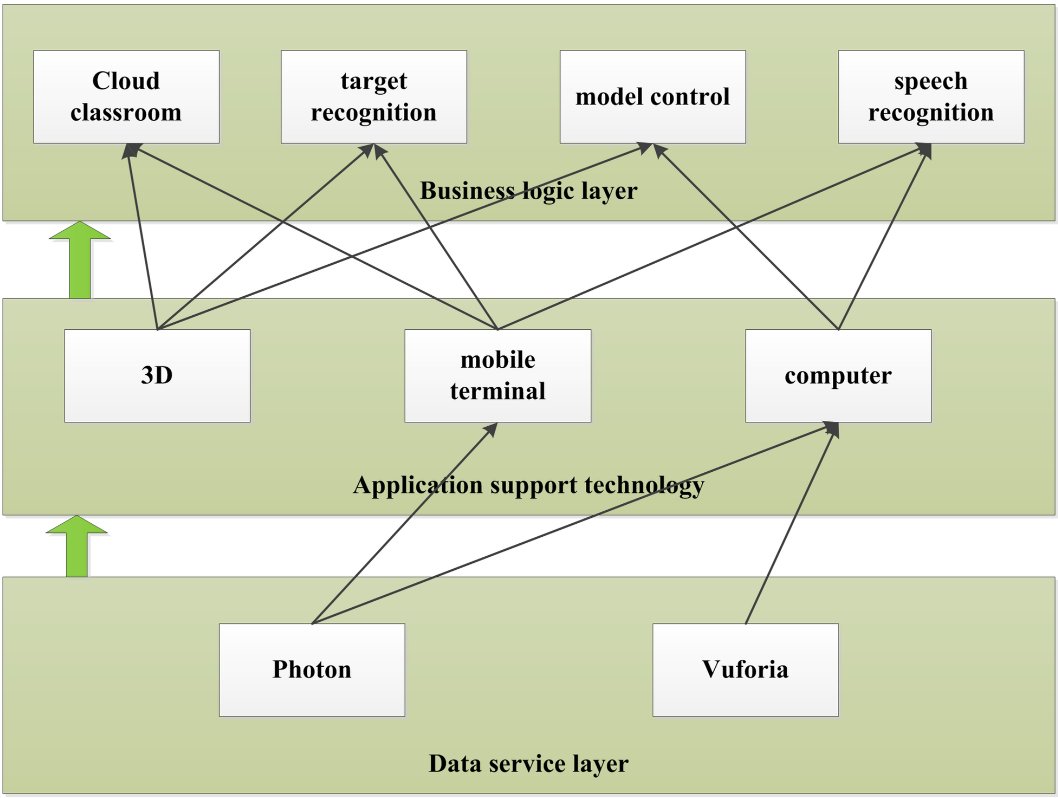
Figure 1. AR hybrid teaching system



the design, a tiered design approach was adopted, using a cloud storage platform to store data and save space (Yi, et al., 2022). Layered design divides the software into several layers, each layer only solves a part of the problems and completes the overall goal through the cooperation of all layers. A complex problem is decomposed into system subproblems, which effectively reduces the scale and complexity of each subproblem. The system architecture is shown in Figure 2, which covers the data layer, the support layer, and the logic layer, where the data service layer needs to be able to recognize images, store teacher information, and manage related operations, and is implemented through Vuforia technology. The support layer is used for the construction of images and models, using Unit3D technology and for the synchronized transmission and display of data through the use of mobile terminals. The logic layer is used for the creation of cloud classrooms, target recognition, and voice interpretation.

3D Max, short for 3D Studio Max, is a 3D animation rendering and production software based on a PC system. It is widely used in advertising, film and television, industrial design, architectural design, 3D animation, multimedia production, games, and engineering visualization. In the construction of the 3D virtual teaching model, the main use of 3D Max technology is to achieve modelling, and draw and map materials, and save the file as FBX when saved to the 3D AR project. In the construction of the cloud classroom, the management is realized through the mobile terminal, using MasterServer to create the cloud classroom, designing the number of students, secret keys, and other information. Students can access the cloud classroom through the terminal. 3D model construction requires the computer to generate 3D coordinates, angles and sizes, generate scanned images through the camera, generate a 3D model, and design 3D coordinates and generate a coordinate system through rigid body transformation (Almqvist, et al., 2021). The model is projected to reach

Figure 2. System architecture



the screen coordinate system to achieve the 3D model construction. It is assumed that the points to be identified are denoted by P and have different coordinates in space and in the camera, X_w, Y_w, Z_w and X_c, Y_c, Z_c, respectively. The focus center of the camera is denoted by O. This point is also the camera coordinate origin, and the coordinate system is established using the camera optical axis as the Z-axis, which satisfies the following relationship according to the coordinate system.

$$\begin{bmatrix} x_b \\ y_b \\ z_b \\ 1 \end{bmatrix} = \begin{bmatrix} R, t \\ o^T, 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = L_w \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

where R is the orthogonal rotation matrix, t is the translation matrix and L_w is the 4*4 matrix. The pixel plane coordinates and the camera coordinates satisfy the condition that

$$Z_b \begin{bmatrix} u \\ w \\ 1 \end{bmatrix} = \begin{bmatrix} g, 0, 0, 0 \\ 0, g, 0, 0 \\ 0, 0, 1, 0 \end{bmatrix} \begin{bmatrix} x_b \\ y_b \\ z_b \\ 1 \end{bmatrix} \tag{2}$$

where g denotes the focal length of the camera. The relationship between the pixel coordinate system and the physical coordinates satisfies the condition that

$$\begin{bmatrix} u \\ w \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx}, 0, u_0 \\ 0, \frac{1}{dy}, w_0 \\ 0, 0, 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3}$$

where dx denotes the most sensitive chip pixel size, u and w denote the central coordinates of the plane. The final coordinate transformation relationship of the image is obtained and displayed on the screen as the following equation.

$$Z_b \begin{bmatrix} u \\ w \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx}, 0, u_0 \\ 0, \frac{1}{dy}, w_0 \\ 0, 0, 1 \end{bmatrix} \begin{bmatrix} g, 0, 0, 0 \\ 0, g, 0, 0 \\ 0, 0, 1, 0 \end{bmatrix} \begin{bmatrix} S, t \\ o^{\mathrm{T}}, 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{4}$$

In the network synchronization, real-time technology selection is implemented using interpolation algorithms. In the simulation between the virtual object operation and the physical object, the angle, as well as the position changes need to be taken into account, so the displacement of the object needs to be solved in the modeling (Anderson, et al., 2021). In the study of relevant movement laws, in relation to splitting the time axis limit, a linear approximation can be used to simulate the displacement, so interpolation can be used to solve this problem. In this paper, the linear interpolation method is used in the study to achieve network synchronization, and this algorithm is faster in comparison and can obtain a more realistic model. The linear interpolation algorithm is expressed as the following equation.

$$Vis_{pi} = (1 - \delta_{pi})Vis_{p1} + \delta_{pi}Vis_{p2} \tag{5}$$

where Vis denotes the inserted filter frame, $p1$ and $p2$ denote the two adjacent over frames, and δ_{pi} takes values in the range of [0,1]. Considering that the teacher's lip variation has a large influence on pronunciation in hybrid English teaching in colleges and universities, the range of δ_{pi} parameter is set at [0.5,1] when selecting parameters. This method of controlling the parameter range variation is able to form a database of the obtained large number of teachers' lip shape over pictures, and after training, to get the model that best matches the real one.

Gesture Recognition Synchronization Synthesis

In language teaching, gesture recognition can serve as an auxiliary tool to help students better understand and remember language. Below, the authors will provide a simple example to illustrate how to construct gesture recognition to improve language teaching. Suppose we are teaching basic English vocabulary. We can create a gesture for each word so that students can better remember and understand it. For example, if we want to teach the word "run," we can create a gesture where both hands clench their fists on both sides of the body and move forward, symbolizing the action of running. To associate these gestures with words, we can use videos or animations to display the corresponding gestures for each word. In this way, students can learn gestures while learning words. Then, students can use these gestures to express corresponding words in language communication. Through this approach, gesture recognition can help students associate words with their meanings, thereby enhancing their vocabulary. In addition, gestures can help students better understand and memorize complex grammar and sentence structures.

The problem of gesture recognition needs to be considered in AR modeling, and in the field of gesture recognition, feature values are important points as a way to distinguish different gestures. The selection of feature values will directly affect the recognition results, but there are many gesture feature points that can be selected, such as palm position coordinates, movement speed, etc. (Georgakopoulos, et al., 2018). The more desirable feature value is the direction angle. Different gestures of the same hand will have different movement speeds, and while the position coordinates will change a lot, the direction angle is regular (Pham, et al., 2020). Therefore, in gesture recognition, the directional angle is used as a feature value for calculation.

Assuming that the hand position coordinates are $\left(x_t, y_t\right)$ at moment t and the coordinate position at the next moment $t+1$ is $\left(x_{t+1}, y_{t+1}\right)$, the gesture eigenvalue direction angle formula can be expressed as

$$\theta = \arctan[(y_{t+1} - y_t) \,/\, (x_{t+1} - x_t)] \tag{6}$$

This value is then encoded and quantized. If the circular angle encoding covers 12 directions, the angle is 30 degrees in each direction. The gesture coordinates are a class of sequences with directionality, and these similar values can be described by the same symbol, reducing the error. The current number of HMM observations is large, and the observation value in this paper is set to

12. It is generally believed that the larger the observation value is, the more accurate the directional description is. But in the actual measurement, the observation value is too large, which will divide the plane into more pictures. Considering the accuracy of the motion, the error is that very small changes will be recognized as different gestures, and the recognition effect will not be good. Therefore, it is set to 12. In improving the recognition rate of gesture algorithm, the convolutional neural network is applied more. The dual-stream convolutional neural network structure uses RGB images as input information and uses SVM to classify video motion, which focuses more on moving objects and can improve recognition accuracy. In this paper, a similar theory is used in the research and analysis, and the input images are selected from HSV and MHI-processed images to improve the computational speed (Funke, et al., 2019). Based on ensuring the recognition accuracy, it is also necessary to consider the reduction of parameter variations and computational effort.

$$y_{\ 1}= D_k \cdot D_k \cdot M \cdot N \cdot D_f \cdot D_f \tag{7}$$

where $D_k \cdot D_k$ denotes the convolutional kernel size, M and N represent the number of input channels and output channels, respectively, and $D_f \cdot D_f$ denotes the feature map. The expression of the convolutional calculation of 3D convolutional neural network is as follows.

$$y_2 = D_k \cdot D_k \cdot D_k \cdot M \cdot N \cdot D_f \cdot D_f \cdot D_f \tag{8}$$

where D_k denotes the number of convolutional kernels and D_f denotes the frame. It can be seen that if the total number of neighboring feature maps is fixed in the convolution calculation of the 2D convolutional neural network, the other values can also be fixed. When the convolutional layer is changed to 3D convolution, the input layer is divided into several copies, the parameter format is increased, and the amount of parameter variation can be expressed as

$$\frac{D_k D_k D_k mn}{D_k D_k MN} = \frac{D_k}{cD_f} \leq 1 \tag{9}$$

The change of convolutional neural network convolutional computation compared to the original is shown below.

$$\frac{D_k D_k D_k mn D_{\mathrm{F}} D_{\mathrm{F}} D_{\mathrm{f}}}{D_k D_k MND_{\mathrm{F}} D_{\mathrm{F}}} = \frac{D_k}{c} \tag{10}$$

Dynamic gesture recognition can further enhance the effectiveness of language teaching. We can create a dynamic gesture for each word or phrase, which can be modified and adjusted in different situations. For example, if we are teaching a word that expresses emotions or attitudes, we can demonstrate different gestures to students based on the context. In addition, dynamic gesture recognition can also be used to evaluate students' learning effectiveness. For example, we can judge students' understanding of language by observing the gestures they use in communication. If the gestures used by students are not consistent with what we teach, it may indicate that they need more practice and understanding. In short, by constructing gesture recognition to improve language teaching, we can help students better understand and remember language, enhance their vocabulary, and enhance their language skills. Dynamic gesture recognition can further enhance the effectiveness of language teaching and provide a new method for evaluating students' learning effectiveness. For the hybrid ELT gesture recognition problem, a 3D convolutional neural network is implemented, and this network covers the image and processing layer, the convolutional base layer, and the pooling layer composition. The historical images of teachers teaching are first segmented using HSV, and then the neural network is used to extract the feature values of the gestures and predict the gestures using the softmax function. In the activation function, the Leaky Relu function is chosen with a coefficient of 0.2. The video clips of hybrid English teaching are represented by C , and t is used to represent time. Each video clip is processed to obtain $L_0 \in (C_0, D_0, \ldots, D_{m-2})$. Multiple frames are obtained from the teacher history video images, which are merged with the initial clips to form an m-frame input image, and this image is fed into the neural network with the following cases.

$$I_{ij}^{xyz} = f[\sum_{m} \sum_{p=0}^{i=1} \sum_{q=0}^{i=1} \sum_{r=0}^{i=1} w_{ijm}^{pqr} l_{(i-1)m}^{(x+p)(y+q)(z+b)} + b_{ij}] \tag{11}$$

where f denotes the activation function, j denotes the value in the feature map, m denotes the value of the convolution kernel connected to the feature map, r denotes the time dimension, p denotes the height value of the convolution kernel, and q denotes the width of the convolution kernel. This formula is used to calculate the convolution kernel, and the final 3D pooling layer is similar to this calculation. Finally, the obtained feature values are fed as a one-dimensional vector into the concatenation layer to obtain the result. Decomposition by the standard convolution process is able to reduce the computational process. Extending this process to the 3D convolution process, the input sequence is computed separately for each frame channel, without combining the feature values. For 3D point-by-point convolution, the computational process is expressed as the following equation.

$$d_{ij}^{xyz} = f(\sum_{p=0}^{i=1} \sum_{r=0}^{i=1} w_{ijm}^{pqr} l_{(i-1)h}^{(x+p)(y+q)(z+k)} + b_{ij}) \tag{12}$$

$$k = ceil(\frac{j}{m}), h = j - km - 1 \tag{13}$$

$$l_{ij} = f(\sum_{m} w_{ijm} d_{im}^{xyz} + b_{ij}) \tag{14}$$

The post-convolution calculation procedure for the final decomposition is Eq. (15).

$$y_3 = D_k \cdot D_k \cdot D_k \cdot M \cdot D_F \cdot D_F \cdot D_f + D_k \cdot M \cdot N \cdot D_F \cdot D_F \cdot D_f \tag{15}$$

Gradient dispersion is the disappearance of a gradient, and the derivative is 0. The problems caused by gradient dispersion include: the hidden layer near the output layer has a large gradient, and the parameters are updated quickly, so it will converge quickly; The gradient of the hidden layer near the input layer is small, and the parameter update is slow. It is almost the same as the initial state, and it is randomly distributed (Jiménez, et al., 2017). Using 3D convolutional neural network, the number of network layers is increased by a factor of 1, which can improve the nonlinear characteristics of the network and better classification, but it also brings the problem of gradient dispersion. If the learning rate is large, the network will not converge, and if the learning rate is too small, the shallow network cannot be updated. Therefore, in this study, we use skip connection to solve this problem, reduce the gradient dispersion problem, and improve the accuracy.

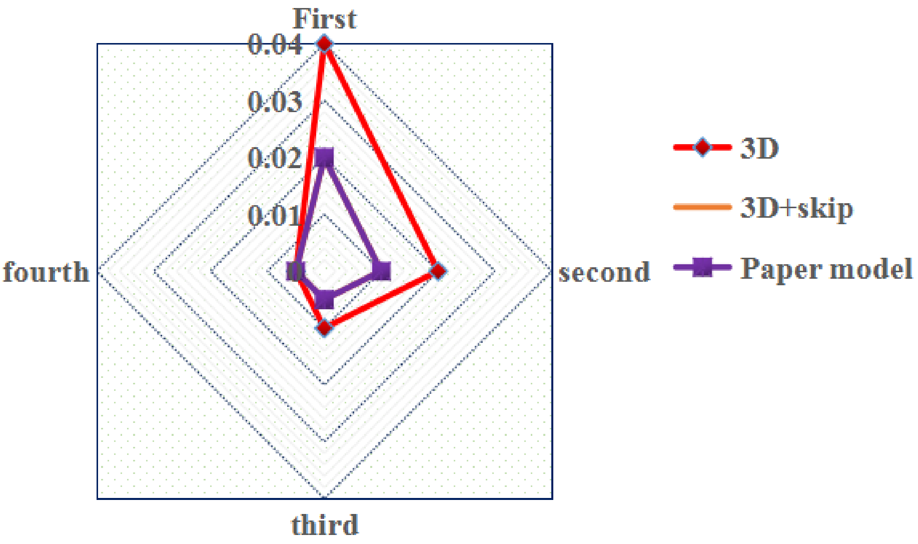
4. RESULT ANALYSIS AND DISCUSSION

Simulation Analysis of Hybrid ELT AR Architecture

Due to the pooling layer, the detection model does not require the size of the input image. But considering the batch training effect, the input image needs to be reduced to a uniform size, and a linear interpolation method is used to unify the input image to a size of 512 × 512 × 3. The pixel values are taken in the range of [0,255], and the pixels are normalized. The minimum pixel value is subtracted, and the difference between the maximum and minimum pixels is divided to obtain the input image. Too small a sample size within the dataset will cause model fitting, and considering the insufficient sample data, data augmentation is required to improve the generalization ability. The imgaug library is used to enhance the data, and image enhancement is achieved by panning, flipping, clipping, and radiative transformations.

The network model is trained, and the simulation analysis is performed using a 3D convolutional neural network, 2D convolutional neural network, separable convolutional network, and convolutional network with skip connection operation added. Multiple gesture recognition needs to be judged in the training. The softmax function is used for training in the multi-classification process with probabilities ranging from 0 to 1. To facilitate the calculation, the cost function is chosen as the NLL function, which is a commonly used classification loss function that can be extended to multiple problem analysis. The RMSProp training optimizer is used with 32,000 iterations and a learning rate size of 0.0005. To improve the convergence speed, the learning rate decay step is set to 4000, and the decay rate is set to 0.8. A stochastic gradient descent method is used for training. In the propagation, the gradient change will occur after the introduction of the skip connection method, so the results under different learning rates are measured. The details are shown in Figure 3.

Figure 3. Model parameter design



Compared with the two-dimensional convolutional neural network, the use of three-dimensional convolutional networks to optimize gesture recognition has improved performance and can increase accuracy, especially in the negative example class. In practical hybrid ELT applications, the tolerance of false positive examples is not high, so the use of 3D convolutional neural network optimization can improve the model classification accuracy. The use of skip connection can make the gradient descent smoother and the learning rate should be reduced. The computational volume, size, and accuracy under different models are measured, and the results are shown in Figure 4.

From the data in the figure, we can see that in comparison, the proposed AR model in this paper has better recall rate prediction and lower error rate. Comparatively, the other two algorithms also have their own advantages in terms of accuracy, but the model proposed in this paper reduces the amount of rate parameter calculation, improving the performance.

For convolutional neural network in the instance segmentation, the change of loss values in the training and validation sets are measured, and the measurement results are shown in Figure 5. From

Figure 4. Calculation amount, size, and accuracy under different models

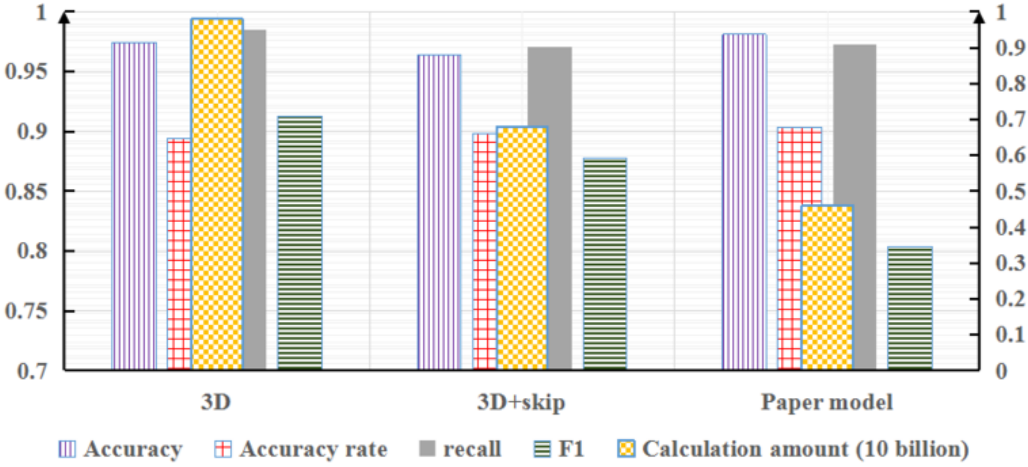
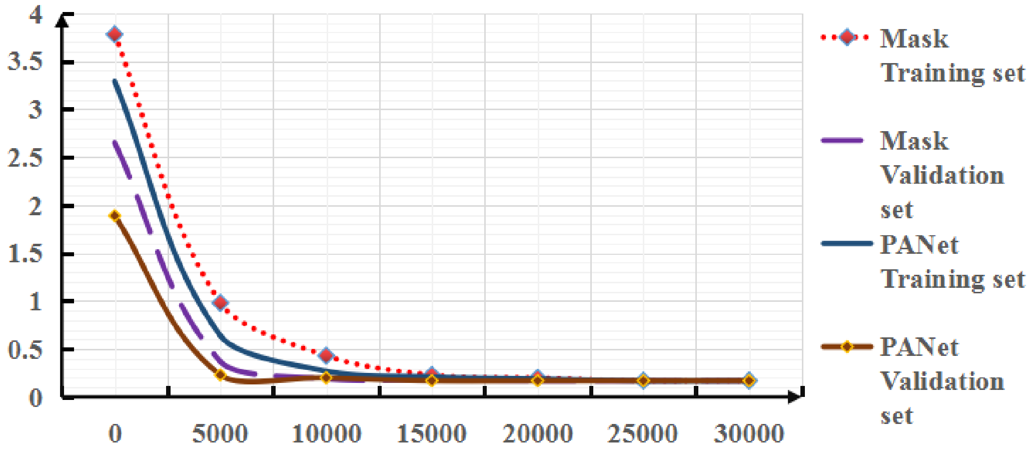


Figure 5. Loss curve change



which it can be seen that the function converges in both the training and validation sets without any overfitting phenomenon.

Analysis of Gesture Recognition Accuracy Rate

To compare and analyze the gesture recognition results, all the gestures were divided into four categories covering upward, downward, leftward and rightward. All five groups of experiments were conducted, with 10 times each group, and the recognition rate was measured by taking the average value. The measurement results are shown in Figure 6, and it can be seen that the recognition rate of all four gestures is above 92%, which indicates that the model has a high recognition rate.

Under the mainstream two-stage instance segmentation model, the AP values of this paper's algorithm and other algorithms in recognizing four gestures were measured, and the measurement results are shown in Fig. 7. From the data in the figure, we can see that the upward recognition rate is the largest compared to the AP values of the other three gestures, but it is also above 92%.

5. CONCLUSION

This article studies the AR construction technology of the blended English teaching mode in a university environment. Using the existing mobile terminal technology that combines AR and Vuforia as the development framework and using commonly used modeling tools to achieve model construction. The teacher end can achieve model propagation, contour brushing, and surgical recognition. In addition to synchronous learning, the student end can also facilitate offline communication. To solve the problem of high computational complexity in dynamic gesture recognition, we suggest using 3D convolutional networks to simplify computational complexity and improve learning rate, to solve the gradient dispersion problem (Wang, et al., 2019). The research results show that this algorithm can achieve convergence in both the training and validation sets. Compared with other algorithms, the AR model has better recall prediction and lower error rate, and significantly improves gesture recognition rate (Skalic, et al., 2019). It should be noted that in 3D construction, AR architecture technology can be combined with VR technology. The structure

Figure 6. Analysis of gesture recognition rate

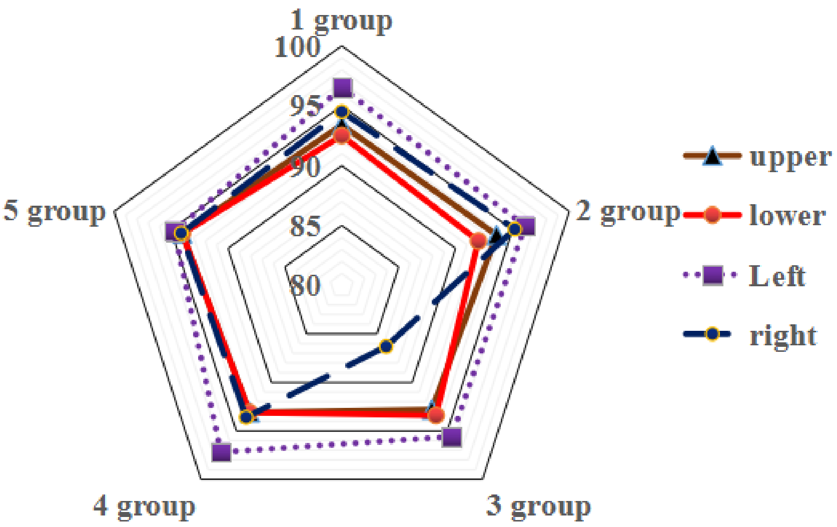
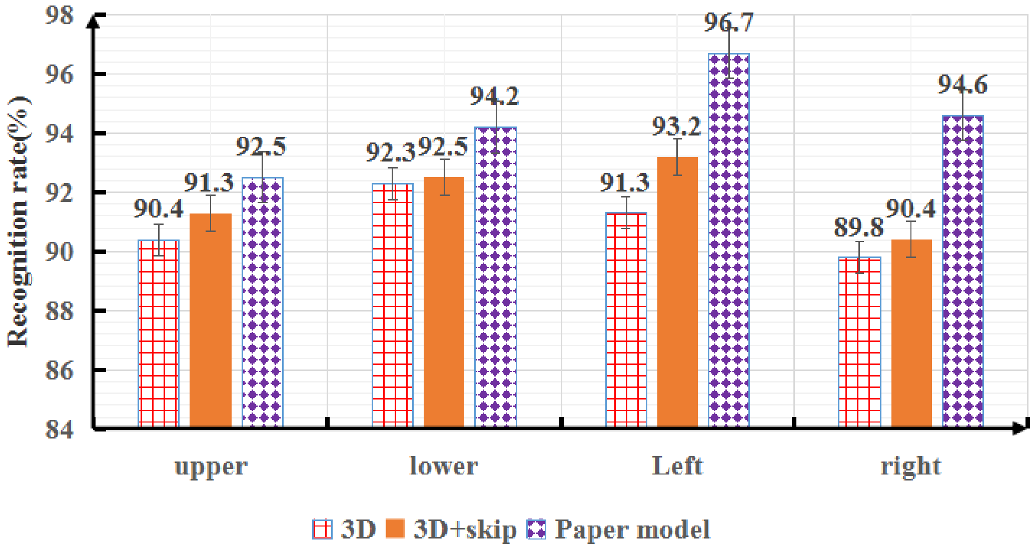


Figure 7. recognition rate under different models



and quantity of convolutional neural networks can affect the effectiveness of gesture recognition, and the depth of network layers is also an important influencing factor. The construction of gesture recognition models needs improvement.

6. DATA AVAILABILITY

The figures used to support the findings of this study are included in the article.

7. CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

8. FUNDING STATEMENT

This work was supported by the Ningxia Medical University, Medical ESP Course Online Teaching Model, YJG202007.

9. ACKNOWLEDGEMENTS

The authors would like to show sincere thanks to those techniques who have contributed to this research.

REFERENCES

Almqvist, V., Berg, C., & Hultgren, J. (2021). Reliability of remote post-mortem veterinary meat inspections in pigs using augmented-reality live-stream video software. *Food Control*, *125*, 107940. doi:10.1016/j.foodcont.2021.107940

Anderson, M., Guido-Sanz, F., Díaz, D. A., Lok, B., Stuart, J., Akinnola, I., & Welch, G. (2021). Augmented reality in nurse practitioner education: Using a triage scenario to pilot technology usability and effectiveness. *Clinical Simulation in Nursing*, *54*, 105–112. doi:10.1016/j.ecns.2021.01.006

Ariansyah, D., Erkoyuncu, J. A., Eimontaite, I., Johnson, T., Oostveen, A. M., Fletcher, S., & Sharples, S. (2022). A head mounted augmented reality design practice for maintenance assembly: Toward meeting perceptual and cognitive needs of AR users. *Applied Ergonomics*, *98*, 103597. doi:10.1016/j.apergo.2021.103597 PMID:34598078

Cascini, G., O'Hare, J., Dekoninck, E., Becattini, N., Boujut, J. F., Guefrache, F. B., & Morosi, F. (2020). Exploring the use of AR technology for co-creative product and packaging design. *Computers in Industry*, *123*, 103308. doi:10.1016/j.compind.2020.103308

Chen, C., Pan, Y., Li, D., Zhang, S., Zhao, Z., & Hong, J. (2020). A virtual-physical collision detection interface for AR-based interactive teaching of robot. *Robotics and Computer-integrated Manufacturing*, *64*, 101948. doi:10.1016/j.rcim.2020.101948

Chu, M., Matthews, J., & Love, P. E. (2018). Integrating mobile building information modelling and augmented reality systems: An experimental study. *Automation in Construction*, *85*, 305–316. doi:10.1016/j.autcon.2017.10.032

Funke, I., Mees, S. T., Weitz, J., & Speidel, S. (2019). Video-based surgical skill assessment using 3D convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, *14*(7), 1217–1225. doi:10.1007/s11548-019-01995-1 PMID:31104257

Georgakopoulos, S. V., Kottari, K., Delibasis, K., Plagianakos, V. P., & Maglogiannis, I. (2018). Pose recognition using convolutional neural networks on omni-directional images. *Neurocomputing*, *280*, 23–31. doi:10.1016/j.neucom.2017.08.071

Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A. S., & De Fabritiis, G. (2017). DeepSite: Protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics (Oxford, England)*, *33*(19), 3036–3042. doi:10.1093/bioinformatics/btx350 PMID:28575181

Leal Filho, W., Shiel, C., Paço, A., Mifsud, M., Ávila, L. V., Brandli, L. L., Molthan-Hill, P., Pace, P., Azeiteiro, U. M., Vargas, V. R., & Caeiro, S. (2019). Sustainable Development Goals and sustainability teaching at universities: Falling behind or getting ahead of the pack? *Journal of Cleaner Production*, *232*, 285–294. doi:10.1016/j.jclepro.2019.05.309

Li, X., Xie, Y., & Liu, T. (2020, May). Research on oral English teaching system based on VR in the background of AI. []. IOP Publishing.]. *Journal of Physics: Conference Series*, *1550*(2), 022031. doi:10.1088/1742-6596/1550/2/022031

Ma, L. (2021). An immersive context teaching method for college English based on artificial intelligence and machine learning in virtual reality technology. *Mobile Information Systems*, *2021*, 1–7. doi:10.1155/2021/2637439

Mattia, G. M., Sarton, B., Villain, E., Vinour, H., Ferre, F., Buffieres, W., Le Lann, M.-V., Franceries, X., Peran, P., & Silva, S. (2022). Multimodal MRI-based whole-brain assessment in patients in anoxoischemic coma by using 3D convolutional neural networks. *Neurocritical Care*, *37*(S2, Suppl 2), 303–312. doi:10.1007/s12028-022-01525-z PMID:35876960

Pezaro, S., Jenkins, M., & Bollard, M. (2022). Defining 'research inspired teaching' and introducing a research inspired online/offline teaching (riot) framework for fostering it using a co-creation approach. *Nurse Education Today*, *108*, 105163. doi:10.1016/j.nedt.2021.105163 PMID:34741912

Pham, H. H., Salmane, H., Khoudour, L., Crouzil, A., Velastin, S. A., & Zegers, P. (2020). A unified deep framework for joint 3d pose estimation and action recognition from a single rgb camera. *Sensors (Basel)*, *20*(7), 1825. doi:10.3390/s20071825 PMID:32218350

Salama, E. S., El-Khoribi, R. A., Shoman, M. E., & Shalaby, M. A. W. (2018). EEG-based emotion recognition using 3D convolutional neural networks. *International Journal of Advanced Computer Science and Applications*, *9*(8). doi:10.14569/IJACSA.2018.090843

Shen, X., & Javidi, B. (2018). Large depth of focus dynamic micro integral imaging for optical see-through augmented reality display using a focus-tunable lens. *Applied Optics*, *57*(7), B184–B189. doi:10.1364/AO.57.00B184 PMID:29521988

Skalic, M., Varela-Rial, A., Jiménez, J., Martínez-Rosell, G., & De Fabritiis, G. (2019). LigVoxel: Inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics (Oxford, England)*, *35*(2), 243–250. doi:10.1093/bioinformatics/bty583 PMID:29982392

Wake, N., Bjurlin, M. A., Rostami, P., Chandarana, H., & Huang, W. C. (2018). Three-dimensional printing and augmented reality: Enhanced precision for robotic assisted partial nephrectomy. *Urology*, *116*, 227–228. doi:10.1016/j.urology.2017.12.038 PMID:29801927

Wang, G. (2022). Formative assessment system of VR teaching in English translation class. *OAlib*, *9*(2), 1–7. doi:10.4236/oalib.1108356

Wang, Y., Teng, Q., He, X., Feng, J., & Zhang, T. (2019). CT-image of rock samples super resolution using 3D convolutional neural network. *Computers & Geosciences*, *133*, 104314. doi:10.1016/j.cageo.2019.104314

Yanfei, M. (2021). Online and offline mixed intelligent teaching assistant mode of English based on mobile information system. *Mobile Information Systems*, *2021*, 1–6. doi:10.1155/2021/7074629

Yi, B., Sun, R., Long, L., Song, Y., & Zhang, Y. (2022). From coarse to fine: An augmented reality-based dynamic inspection method for visualized railway routing of freight cars. *Measurement Science & Technology*, *33*(5), 055013. doi:10.1088/1361-6501/ac3c1c

Zhang, D., Wang, M., & Wu, J. G. (2020). Design and implementation of augmented reality for English language education. *Augmented reality in education: A new technology for teaching and learning*, 217-234.

Zhang, Y., Jin, Z., & Chen, Y. (2020). Hybrid teaching–learning-based optimization and neural network algorithm for engineering design optimization problems. *Knowledge-Based Systems*, *187*, 104836. doi:10.1016/j.knosys.2019.07.007

Zhou, Y. (2020). *VR technology in English teaching from the perspective of knowledge visualization. IEEE Access. Jing Yang*. Foreign Language College, Hulunbuir University.