International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3236-3246

https://doi.org/10.70517/ijhsa463272

Integrating Emotion Recognition and Augmented Reality in Art Education

Xiaopeng Pei[1,*]
[1] College of Humanities and Education, Hebi Polytechnic, Hebi, Henan, 458030, China
Corresponding authors: (e-mail: 15503929190@163.com).

Abstract Augmented reality technology, as an important achievement in the development of science and technology in the new era, has been widely used in the field of education. The study is based on the 3D-ResNet18 network, which is improved by adding Self-Attention layer and transformer encoder to construct an emotion recognition model based on deep learning. The model is combined with augmented reality technology and used together in art education. Through the experiments conducted on the image data collected by students in art teaching, the improved 3D-ResNet18 network model in this paper has high accuracy in recognizing the emotion of students' expressions, and the recognition accuracies of confusion, happiness, normality, and boredom are all over 90%, and the overall recognition accuracies are improved by 0.51%~13.49% compared with other methods, which reflects the high-precision emotion recognition performance of the constructed method. After being used in AR art teaching, the overall emotion score of the sample students was recognized to be about 0.65, which confirms the effectiveness and practicality of the fusion application of the emotion recognition model and AR technology, which can support the diagnosis of students and classroom situations, and is conducive to the timely adjustment of the teaching program and the promotion of the development of the quality of art education.

Index Terms deep learning, augmented reality technology, emotion recognition model, 3D-ResNet18, art education

I. Introduction

With the continuous progress of science and technology, emotion recognition and augmented reality technology gradually play an important role in various fields, especially in art education [1], [2]. Emotion recognition combined with augmented reality technology provides students with a more three-dimensional, personalized, and immersive learning experience and expands their creative space [3], [4]. Emotion recognition technology is a technology that simultaneously utilizes multiple media forms (e.g., speech, text, images, video, etc.) for emotion recognition [5]. Its basic principle is to analyze and recognize the emotions expressed in multiple media forms through techniques such as computer vision, natural language processing and machine learning [6], [7].

In art education, emotion recognition technology has a wide range of application prospects, especially in the methods of personalized teaching, educational resources development and intelligent assisted evaluation [8]-[10]. Through the analysis of students' voice, text and facial expression in the learning process, based on the recognition of the emotional state, it can determine whether students understand the teaching content, and then give the corresponding tutoring and feedback, and the teacher can adjust the teaching strategy according to the emotional state of the students, provide personalized teaching services, and improve the learning effect of the students [11]-[14]. And Augmented Reality (AR) technology is to combine the virtual world with the real world, adding virtual images to the actual scene through digital information, so that the user can get a more realistic and concrete experience [15]-[17]. In art education, AR technology can bring an immersive learning experience to students, enhance their interest and engagement, and expand their learning resources and tools [18]-[20]. The integration and application of emotion recognition and augmented reality technology in art education is an innovative attempt of technology and education, which not only enriches the teaching mode of the art classroom, but also effectively cultivates students' innovative spirit, and then continuously improves the teaching effect [21]-[24].

The article analyzes the application of augmented reality technology in art education, and selects 50 college students as the subjects to collect their data related to AR art teaching and process them. The 3D-Resnet18 network is used as the basic network to process the teaching video images in consecutive frames, the Self-Attention layer is added to 3D-Resnet18 to extract the basic feature relations in the sequences, and the transformer encoder is introduced to improve the computational efficiency, and the emotion recognition model based on the improved 3D-Resnet18 network is built . Experiments are conducted on this emotion model using the collected data to explore the optimization effect of the model and the performance of emotion recognition through the comparison of the accuracy and loss values before and after the ablation experiments and model improvement, as well as the comparison of this paper's method with other methods. Ten more students are selected to collect their AR art teaching data and apply the trained model for analysis. Finally, the application of emotion recognition and augmented reality technology in art teaching is further discussed to explore the future development of the integration of the two applications.

II. Application of augmented reality in art education

Augmented Reality (AR), that is, augmented reality, refers to the combination of virtual things with the real environment, which can enhance people's practical understanding of the environment they are in. The application of augmented reality technology in the field of education emphasizes visual, auditory and tactile multi-sensory stimulation in conveying information, which effectively makes up for the shortcomings of traditional teaching methods and general multimedia teaching. The application of augmented reality technology in art education has important practical feasibility.

First of all, augmented reality technology improves the degree of visualization of teaching content. The technology makes the virtual world of physical objects can be combined with the specific real environment, the original virtual, abstract content more image, intuitively presented to the students, so that the visual world of the students by a more intense stimulation, which in turn inspires the students to develop further imagination, creation and other higher level of art learning activities. Through the use of this technology, students in art learning can be integrated into a more realistic scene, which effectively stimulates students' emotional experience and improves the efficiency of the classroom.

Secondly, the use of augmented reality technology creates an interactive learning environment closer to reality for students' art learning. Under this technology, the use of sound, light, electricity as well as diagrams and colors interacts with the students, and the students' learning and exploratory ability is cultivated, and their thinking is effectively expanded, which improves the effectiveness of classroom teaching. Students learn under the learning scene created by augmented reality technology, which strengthens students' understanding and memory of the teaching content, while the interactivity of the technology itself increases students' interest in learning art, and students can study more deeply, greatly improving the learning efficiency.

Again, augmented reality technology enhances students' immersion in learning. The technology has the characteristics of the combination of virtual and real, which makes students in the learning in the immersive scene, through the auditory, visual feel the sound as well as the dynamic picture, not only effectively improve the students' knowledge of the art discipline, but also effectively cultivate students' creativity and imagination, change the traditional teaching under the students face the static mode of the text of the books, the students are immersed in the three-dimensional animation of the teaching scene, enhance the understanding of knowledge.

Finally, augmented reality technology enhances the fun of art teaching. In the application of this technology, students can transition from the traditional art painting and coloring simple learning activities to the three-dimensional transformation of their own works, not only to convey knowledge, but also multi-sensory stimulation can be more effective in attracting the attention of students to enhance the learning of fun.

III. Data Acquisition in AR Art Teaching

III. A. Experimental organization

The subjects tested in this study were 50 college students, 25 male and 25 female, with an age range of 18 to 27 years old. The experimental equipment involved were cameras (using high-definition cameras with autofocus function), AR experience equipment, actual teaching resources (online retrieval of art course content as teaching resources, and four videos of about 5 min duration provided to the test subjects), ELAN annotation tool (the subjects labeled their process emotions through the annotation tool), and emotion labeling tool. The emotion labeling software developed independently by PyQt5 was used, which has the functions of data import, data information, data deletion, and label type.

III. B. Data collection

III. B. 1) Data collection and selection

First, data collection. When learners watch AR art teaching videos, their facial expression data can be captured with the help of a camera to mark the emotion of the testee's self-assessment and save the data. Second, screening data. There is bound to be unqualified image data in the collected data, such as incomplete image data, too much facial occlusion, unclear images, etc., and the unqualified data is deleted through machine screening and manual review.

III. B. 2) Data annotation

Data labeling included self-labeling by the test subjects and labeling by the researcher. With the help of the ELAN annotation tool, the testees labeled their facial expression emotional state at different times during the learning process in advance. Combined with the results of self-labeling by the testees, the emotion labeling tool was used to label the data information.

III. B. 3) Sentiment Classification and Selection

The four affective states of normality, doubt, happiness, and boredom are noted as the possible learning emotions in the AR art teaching process, and there are three reasons for such a division. First, this study applies affective computing to the AR art teaching learning scenario, where there is a single mode of interaction between teachers and students, and there are fewer types of learner emotions and less fluctuation. Second, after analyzing the collected data, it was found that the test subjects often showed the four states of normality, doubt, happiness, and boredom in AR art teaching practice, and other emotions rarely appeared. Thirdly, it was found through the study that the emotions that often appeared in the learners during the learning process were frustration, boredom, and doubt. Therefore, this study classified the emotions into four types: normality, doubt, happiness, and boredom, on the basis of which the emotions were labeled and analyzed.

III. B. 4) Data set partitioning

A total of 4571 pieces of facial expression data were collected, and the datasets were all divided into training set, validation set and test set in the proportion of 60%, 20%, 20%, and 2,743 training sets, 914 validation sets, and 914 test sets of facial expression data were obtained.

IV. Deep learning-based emotion recognition model

The expression video is composed of multiple consecutive frames of expression images, and it is difficult to effectively model the temporal information in the video in general 2D convolutional neural networks that do not have a module dedicated to modeling temporal information. Therefore, in this paper, based on the 3D-ResNet18 network, we improve it to construct an expression emotion recognition model for the art education classroom.

IV. A. ResNet18 network

Currently, there are many mature neural network models, the earliest neural network model is Le Net, after that many classical neural network models have been proposed, such as AlexNet, VGGNet, GoogleNet, ResNet and so on. In this section, ResNet18 is improved by optimizing it for the problem of expression emotion recognition in a real classroom environment.

The complexity of the network model is mainly affected by the depth of the network and, in real experiments, the problems of gradient vanishing and network degradation occur as multiple layers of the network are stacked up to a certain level. Although the problem of gradient vanishing can be improved by methods such as regularization and batch normalization, the problem of network performance degradation due to deeper network depth still exists.

To balance the problem of network depth and performance, the classical residual neural network (ResNet) is proposed. The network mainly proposes a residual learning module, and the core idea of the module is to superimpose the features of the shallow network features and their features learned in the deeper network by means of constant mapping, and continue learning as the input of the next layer. The superposition of the above residual module makes the network layers deepen while maintaining the same performance as the shallow network, thus solving the problem of network degradation caused by deepening layers. The structure of the residual module is very simple, mainly including a residual learning branch and a constant mapping branch to the output. The two branches are added together and passed through a nonlinear activation function to form a complete residual module. The gradient descent chain rule for computing the loss in the network is:

$$\frac{\partial Loss}{\partial X_1}=\frac{\partial F_N\left(X_{L_N},W_{L_N},b_{L_N}\right)}{\partial X_L}\cdots\frac{\partial F_2\left(X_{L_2},W_{L_2},b_{L_2}\right)}{\partial X_1} \tag{1}$$

To solve the problem of vanishing gradient, the gradient of backpropagation becomes after adding the residual unit:

$$\begin{aligned}\frac{\partial Loss}{\partial X_1}=\frac{\partial Loss}{\partial X_L}\frac{\partial X_L}{\partial X_1}=\frac{\partial Loss}{\partial X_L}\left(\frac{\partial X_1+\partial F\left(X_1,W_1,b_1\right)}{\partial X_1}\right)\\=\frac{\partial Loss}{\partial X_L}\left(1+\frac{\partial F\left(X_L,W_L,b_L\right)}{\partial X_L}\right)\end{aligned} \tag{2}$$

The network after adding the residual unit structure effectively solves the problems of gradient disappearance and deep network optimization caused by too many layers in the deep network. Nowadays, there are mainly ResNet18, ResNet34, ResNet50, ResNet101 and ResNet152 residual network models, and the model in this paper is mainly based on ResNet18 model for optimization and improvement.

In ResNet18 residual network, the depth of the network is 18 layers, which mainly includes a convolutional layer with a convolutional kernel scale of 7×7, eight base modules, a fully connected layer, and two pooling layers. Residual blocks in residual networks contain two main types, base residual modules and bottleneck residual modules, which are used in the ResNet18 network. Each residual block contains 3×3 convolutional layer, BN layer, ReLU, 3×3 convolutional layer and BN layer in turn, this type of module does not change the size of the input feature map. However, if the step parameter in the first 3×3 convolutional layer is 2, downsampling occurs in the first 3×3 convolutional layer, which will double the number of output channels and halve the feature map size size.

IV. B. Expression Emotion Recognition Model

IV. B. 1) General structure

In this paper, a dynamic expression emotion recognition model based on 3D visual spatio-temporal network is proposed. The network mainly consists of 3D-ResNet18 network, Self-Attention layer, and transformer encoder.Firstly, continuous video sequences are input into a 3D convolutional neural network, which is used to capture 3D signs of multi-frame video sequences. Immediately after the convolutional layer, a Self-Attention layer is added for self-attention computation at the feature level. The original feature sequence is also summed with the Self-Attention output through jump-joins, which preserves the information of the original features and also ensures that the fused feature representation is more comprehensive and richer. After that, the feature shapes are adapted and fed into the Transformer layer, a step designed to capture longer-range dependencies. Finally, a fully connected layer is used for classification.

IV. B. 2) Primary Feature Extraction Fusion

The Self-Attention layer pairs can better handle temporal dependencies in video data than a simple convolutional network, focusing on correlation analysis between features within a single layer, Self-Attention (SA) is used to compute the value of attention between images in a sequence with the following formula:

$$Attention\left(Q,K,V\right)=soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

In the formula, Q, K, and V are all sequences and matrices W^Q, W^K, and W^V calculated after adding position information, representing the query tensor, key tensor, and value tensor respectively. Moreover, the row vectors in Q correspond to the column vectors in K. The row vectors in K correspond to the column vectors in Q, and V calculates the weights for the attention fraction matrix that has passed through the Softmax function.

Firstly, each frame of the image after transposition is equally segmented into $x_1,x_2,\cdots x_n$ blocks, thus transforming the image into the form of a sequence. Then each sequence x_i is linearly operated with weight matrices W^Q, W^K, W^V. The query vector q_i, the key vector k_i, and the value vector v_i of the corresponding sequence are obtained by calculation. The calculation formula is as follows:

$$\begin{cases}q_i=W_Q*x_i\\k_i=W_K*x_i\\v_i=W_V*x_i\end{cases} \tag{4}$$

Then the attention mechanism will complete the calculation of the attention weights, and the initial attention value $\alpha_{1,i}$ is obtained by dot producting q_1 with all the Keys in turn. The formula is as follows:

$$\alpha_{1,i} = \frac{q_1 * k_i}{\sqrt{d_k}} \tag{5}$$

where d_k is the number of channels of the key tensor, and then $\alpha_{1,i}$ is normalized by the softmax function to obtain the final attention weight $\alpha'_{1,n}$, which is calculated as follows:

$$\alpha_{1,i'} = \frac{\exp(\alpha_{1,i})}{\sum_n \exp(\alpha_{1,i})} \tag{6}$$

Finally, the value vector v_i of each image sequence is multiplied with its corresponding attentional weight and then summed to obtain the final output Z_1, which is computed as follows:

$$Z_1 = \sum_n \alpha'_{1,i} \cdot v_i \tag{7}$$

IV. B. 3) Modeling of timing information

The Transformer model is a deep learning architecture for sequence-to-sequence tasks.The Transformer model consists of two parts: an encoder and a decoder and avoids the network design of circular connections. The input word sequence is first processed by the encoder and then passed from the encoder to the decoder.

(1) Positional encoding

Transformer abandons the traditional recurrent neural network structure, which cannot obtain position information from the order of word sequences, so it needs to introduce position information into the model through position encoding. Compared to the sequential transfer of recurrent neural networks, the purpose of position encoding is to allow the model to remember the order information of words in the sequence, thus solving the problems of slow computation and large storage space occupied by RNNs when dealing with long sequences.

In Transformer, the value of position encoding is represented by calculating the sine and cosine functions. The specific calculation formula is as follows:

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{2i/d}}\right) \tag{8}$$

$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{2i/d}}\right) \tag{9}$$

where PE denotes the position encoding matrix, pos denotes the number of rows of the word in the matrix, i denotes the number of columns of the word in the matrix, d denotes the dimensionality of the word vectors, and $2i+1$ and $2i$ denote the parity, with the sine function calculating the values of the even columns and the cosine function calculating the values of the odd columns. These functions and operations enable the model to better understand the positional relationships of words in the input sequence.

(2) Encoder and Decoder

The encoder in Transformer consists of six encoder sub-layers stacked together, and its main role is to perform feature extraction and context modeling of the input sequence. Each encoding module consists of a fully connected feed-forward neural network (FFN), a multi-head self-attention mechanism (MSA), layer normalization, and residual connectivity. The decoder adds the masked multi-head subattention mechanism to the encoder. The self-attention mechanism in the encoder computes the correlation within the input sequence, while the attention mechanism in the decoder obtains the correlation of the encoder output information. Since only the encoder is used in this paper, the multi-head attention mechanism in the encoder and the feed-forward neural network will be described in detail below.

(3) Multihead self-attention

Multihead attention is an attention mechanism extended on the basis of self-attention. The calculation formula is as follows:

$$\begin{cases} MultiHead\left(Q,K,V\right) = Concat\left(head_1,\cdots,head_h\right)W^0 \\ head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \end{cases} \tag{10}$$

where W_0 is the output weight matrix, W^Q, W^K, W^V are the weight matrices corresponding to Q, K, V, h is the number of attention headers, and Concat stands for matrix splicing. In the Transformer structure, h is set to 8.

(4) Feedforward Neural Network

The feedforward neural network contains two linear layers and an activation function, the linear layer maps the output of MSK into the high dimensional space, and then filtered by the activation function and then reduced back to the original dimension. The formula is as follows:

$$FFN(x) = \max(0, XW_1 + b_1)W_2 + b_2 \tag{11}$$

where W_1 and W_2 are the weight matrices of the first and second linear layers, respectively, b_1 and b_2 are the bias terms, and X is the output matrix of MSK.

(5) Residual Connection and Layer Normalization

In Transformer, residual connectivity and layer normalization are applied to the attention sublayer and the feedforward neural network sublayer. For the attention sublayer, the residual connection directly adds the output of the attention mechanism with the input to get the final output. For the feedforward neural network sublayer, residual connections add the output of the feedforward neural network to the input to get the final output. The introduction of residual connections helps to avoid the gradient vanishing problem, allowing the network to better back-propagate the gradient and train. Layer normalization aims to normalize the features of each sample to have a mean of 0 and variance of 1 and is applied to the input of each sub-layer. With layer normalization, it can help speed up the training process and help avoid the problem of vanishing or exploding gradients. The formula is given below:

$$LayerNorm(X) = \alpha \square \frac{X-\mu}{\sigma} + \beta \tag{12}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{13}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 + \varepsilon} \tag{14}$$

where μ and σ denote the mean and variance of the input X, respectively, α and β are learnable parameters, and ε is a very small constant added to avoid divide-by-zero errors.

V. Experimental results and analysis

V. A. *Ablation experiments*

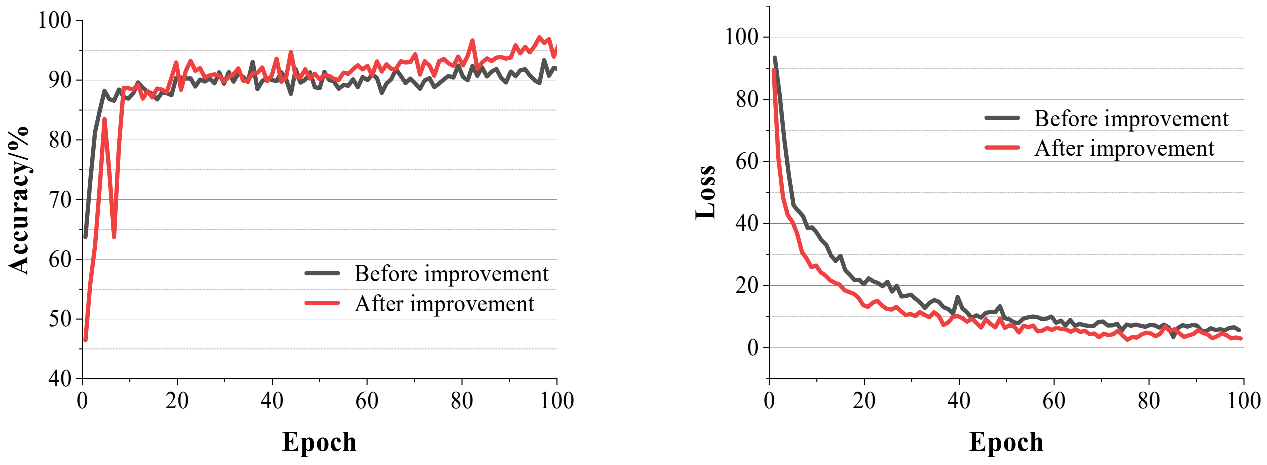
In order to verify the effect of the improved Self-Attention layer and Transformer in the 3D ResNet18 network on the performance of the emotion recognition model, in this section, the ablation experiments are conducted using the homemade dataset, and the results of the ablation experiments are shown in Table 1, where "√" indicates that this structure is adopted, and "×" indicates that this structure is adopted. After the 3D ResNet18 network is improved by using Self-Attention layer and Transformer, the accuracy of the model in recognizing the emotion of students' expressions rises from 84.58% to 94.48%, which indicates that through the optimization of the 3D ResNet18 network, the extraction of the features of the students' expressions is more adequate, and the accuracy of the emotion recognition rises significantly, which verifies the improved model's validation of the improved model.

Table 1: Ablation experiment results

Model	Self-Attention	Transformer	FLOPs/M	Params/M	Accuracy/%
3D-ResNet18	×	×	120.75	3.73	84.58
3D-ResNet18-1	√	×	10.13	0.59	90.05
3D-ResNet18-2	×	√	10.61	0.88	92.84
Our model	√	√	11.24	1.47	94.48

V. B. *Recognition accuracy*

A comparison of the accuracy and loss curves before and after the improved emotion recognition model is shown in Fig. 1, Fig. (a) shows the accuracy curve, Fig. (b) shows the loss curve, the black curve is the 3D ResNet18 model, and the red curve is the improved 3D ResNet18 model. From Fig. (a), it can be seen that the model training tends to stabilize after reaching 18 Epochs, and the improved model maintains the same high level of accuracy as 3D ResNet18 in the recognition of students' expression emotions, and in the later stage of model training, the recognition accuracy of the improved 3D ResNet18 model is significantly higher than that of the original model, and the recognition accuracy is up to about 95%. In Fig. (b), the improved model training begins to converge after 60 Epochs, the value stays above and below 5, and the improved 3D ResNet18 model loss convergence speed and loss are slightly better than the network.

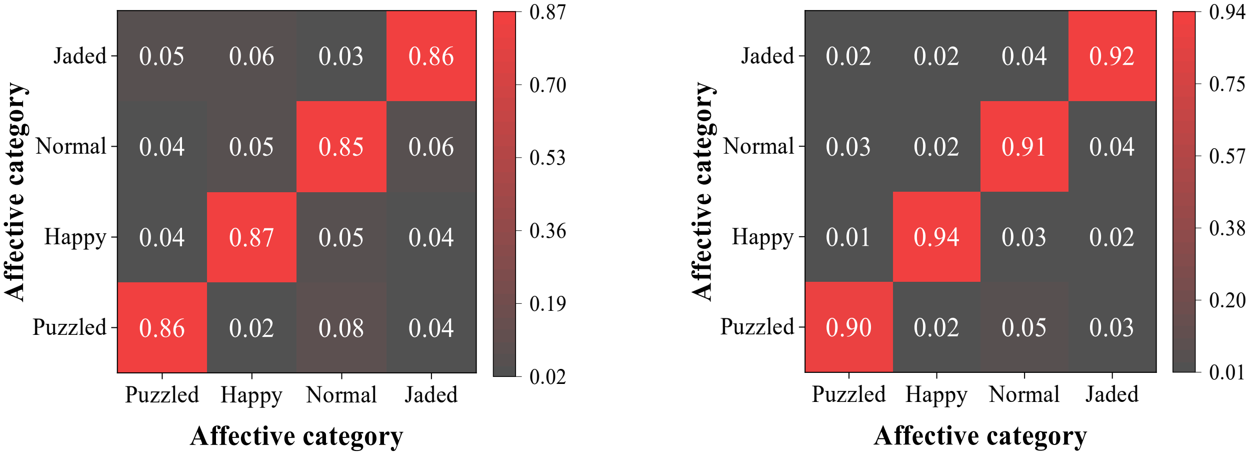


(a)Accuracy

(b)Loss

Figure 1: The accuracy and loss of the emotional recognition model before and after improvement

In order to understand the recognition accuracy of the model for each category, the confusion matrix heat map is drawn in this paper, and the results obtained from the model before and after the improvement using the homemade dataset training are shown in Fig. 2, Fig. (a) is the confusion matrix heat map of the 3D ResNet18 model, and Fig. (b) is the confusion matrix heat map of the 3D ResNet18 model after the improvement. The improved 3D ResNet18 model maintains a high level of recognition accuracy in each category, with recognition accuracies of 90%, 94%, 91%, and 92% for the four emotion types of confused, happy, normal, and fed up, respectively, which are higher than that of the 3D ResNet18 model, which is 85%~87%, and the overall emotion recognition accuracy is better than that of the original model.



(a)Before improvement

(b)After improvement

Figure 2: The comparison of the confused matrix of the homemade data set

V. C. Comparative experiments

In order to analyze the performance of the improved 3D ResNet18 model in comparison with other models, in this section, we use homemade datasets and train the network with AlexNet, VGGNet16, ResNet50, GoogLeNet, and the lightweight MobileNetV3, respectively. The comparison experimental results of different models are shown in Table 2. It can be seen that the improved 3D ResNet18 model exceeds the other network models in terms of recognition accuracy, with a sentiment recognition accuracy of 94.44%, which is an improvement of 0.51% to 13.49% over the other models, and the improved network exhibits better performance than the comparison methods in terms of FLOPs (13.58M), number of parameters (3.95M), and recognition accuracy. The effectiveness of the improved 3D ResNet18 model for recognizing students' emotions in the art education classroom was verified by comparing it with other classification models.

Table 2: Comparison result of different models

Model	FLOPs/M	Params/M	Accuracy/%
AlexNet	14.44	1.13	80.95
VGGNet16	519.55	15.76	90.88
ResNet50	189.52	24.75	93.93
GoogLeNet	42.04	6.37	92.04
MobileNetV3	14.22	5.45	88.44
3D-ResNet18	15.56	5.77	90.56
Our model	13.58	3.95	94.44

V. D. Analysis of integration applications

Ten students, including five boys and five girls, were selected from the test subjects for the applied research on the integration of AR art teaching and emotion recognition. AR technology was utilized to conduct art classroom teaching (20 min) for the subject students, and a camera was used to record the students' classroom videos, and the constructed deep learning-based emotion recognition model was used for emotion recognition.

The detected faces were input into the trained and improved 3D ResNet18 model, and the results of sentiment classification were output. For a sample of students' expressions, the scores of the three positive emotions of "normality", "doubt" and "happiness" are 7, 8, and 9, and the score of "boredom" is -10. The higher the score, the higher the concentration of the students at this moment, and the score is relative to the class as a whole, and a score above 0.5 can be considered as a higher class mood than the class average. Finally, the average score of all students in the whole art course can be used to obtain the overall student emotional score, so as to evaluate the teaching quality of art classroom.

One student (denoted as "Student 1") was randomly selected from 10 students as a sample of the mood scoring experiment. The emotional score of Student 1 is shown in Figure 3, except for the part where no face is detected, that is, when Score=0, the emotional score of the student at other times is basically between 0.40~0.65, which is at the average level, which is a relatively positive classroom emotion.

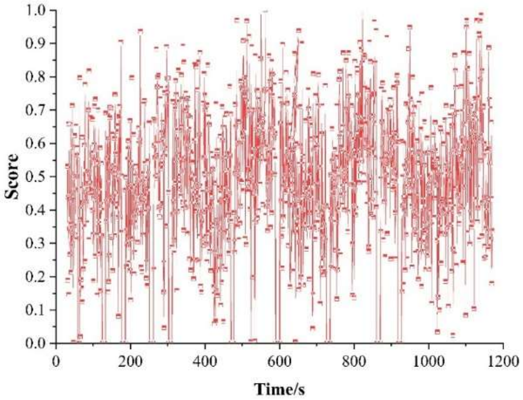


Figure 3: Student 1 mood ratings

Also, a comparison was made between STUDENT1 and the average ratings of all the students who received emotion ratings. The results of the students' mood ratings are shown in Figure 4. It can be calculated that the average emotion rating of these 10 students is around 0.65, which is a relatively active listening state. Through the results of students' emotion recognition, teachers can judge the students' individual and classroom learning, adjust the teaching content and teaching method accordingly, and promote the quality of art education.

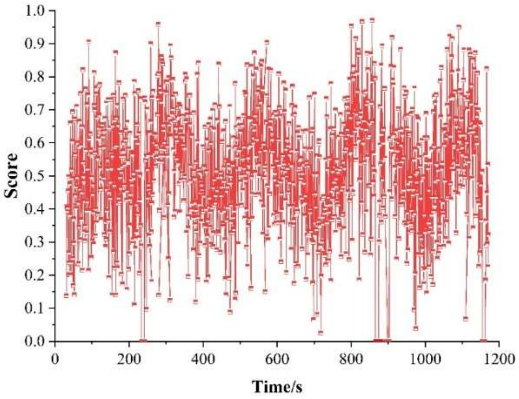


Figure 4: Student mood scores

V. E. Emotion Recognition-Driven AR Art Instruction

There are still many problems to be overcome in the integration of AR and emotion recognition in art education towards large-scale application landing. Discussed from the three aspects of emotion recognition data source, adaptive adjustment of AR teaching course content, and AR teaching safety plan, emotion recognition-driven AR art teaching is shown in Fig. 5, with a view to realizing improvements in future research.

With the increasing number of scenarios in art teaching, the analysis of datasets focusing only on external behavioral actions can no longer meet the many application scenarios in the field of AR teaching, and it is necessary to combine the actual situation to select psychological, physiological, behavioral, and other aspects of the data source to identify and calculate the learner's emotions. For example, AR glasses can be used to sense the dynamic changes of learners' facial muscles to obtain more accurate facial expression data, AR handles can be squeezed to determine learners' grip strength, and wearable pulse measurement bracelets can be used to record learners' pulse data. Although this study did not mention the use of psychological or physiological data to analyze learners' emotions, future research work will study the emotional state of learners in AR art teaching from the perspective of fusion of different data sources.

Adaptive adjustment of AR art teaching content according to the results of emotion recognition.The purpose of applying AR technology to teaching is to innovate cultural education pathways, to realize the adaptation of practice and teaching content, and to achieve the optimized teaching effect and teaching goals. In the large-scale popularization and application of AR teaching in the future, the recognition of learners' emotions will provide services for personalized learning content pushing and dynamic adjustment of classroom content. According to the real-time emotion recognition results for the learner to push the appropriate personalized art learning content, so as to meet the learner's personal learning expectations, to achieve better educational results.

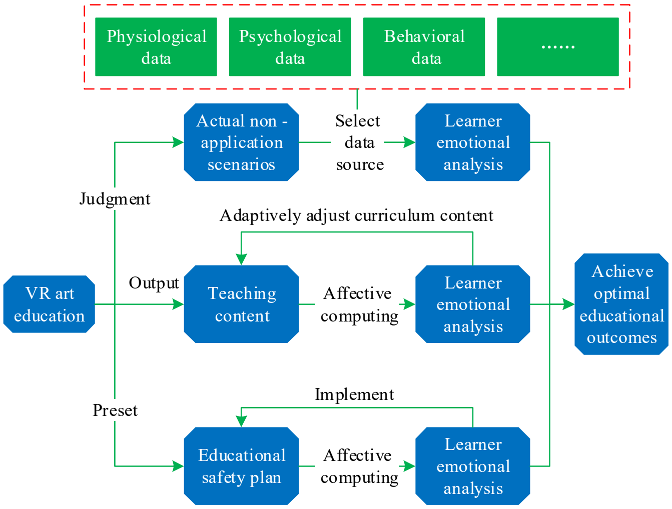


Figure 5: The AR Art teaching driven by emotional recognition

VI. Conclusion

Introducing augmented reality technology into art education has become an important direction of focus in current art education. In this paper, emotion recognition is integrated into AR art teaching, 3D-ResNet18 network is improved, and deep learning-based emotion recognition model is constructed to assist in improving the quality of art teaching courses. Facial image data of students in AR art teaching is collected as a dataset and recognized using the improved 3D-ResNet18 network model, and it is found that the improved model has improved in accuracy and loss value compared with the original 3D ResNet18 model, and the results of the two are stable at around 95% and 5 after convergence. The accuracy of this paper's method is greater than 90% for recognizing the four emotions of students' confusion, happiness, normality and boredom, and its accuracy for emotion recognition is 0.51% to 13.49% higher than that of other methods in the comparison test, which illustrates the superiority of this paper's method for recognizing students' emotions. In addition, the model is used in the actual art classroom, and the overall emotion score of the sample students is obtained to be about 0.65, showing a more positive classroom emotion, which confirms the feasibility of integrating the emotion recognition model with AR technology in art teaching. Augmented reality technology, as a highly representative and continuously developing advanced technology in the digital era, breaks the spatial limitation of the teaching environment. In this paper, augmented reality technology and emotion recognition model are jointly applied to art education, which to a large extent expands the development path of art teaching, and as an emerging technological means to improve the quality of teaching courses and bring driving force for the development of art teaching.

References

[1] Li, E. (2024). Intervention of Art Education on College Students' Aesthetic Mood Based on Emotion Recognition Algorithm. Frontiers in Art Research, 6(9).

[2] Panciroli, C., Fabbri, M., Luigini, A., Macauda, A., Corazza, L., & Russo, V. (2023). Augmented reality in arts education. In Springer Handbook of Augmented Reality (pp. 305-333). Cham: Springer International Publishing.

[3] Dong, Z., An, J., & Liu, L. (2023, December). Leveraging Emotion Recognition for Enhancing Arts Education: A Classroom Behavior Analysis Approach. In Proceedings of the 2023 International Conference on Information Education and Artificial Intelligence (pp. 212-217).

[4] Miralay, F. (2022). Examination of educational situations related to augmented reality in art education. International Journal of Arts and Technology, 14(2), 141-157.

[5] Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., & Li, B. (2024). Development and application of emotion recognition technology—a systematic literature review. BMC psychology, 12(1), 95.

[6] Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. Library & Information Science Research, 33(3), 244-255.

[7] Katirai, A. (2024). Ethical considerations in emotion recognition technologies: a review of the literature. AI and Ethics, 4(4), 927-948.

[8] Li, G., & Wang, Y. (2018, October). Research on leamer's emotion recognition for intelligent education system. In 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (pp. 754-758). IEEE.

[9] Cui, Z. (2025). Expression Recognition Algorithm Based on Fusion Features for Students' Emotional Analysis on Art Education Platform. IEIE Transactions on Smart Processing & Computing, 14(1), 22-32.

[10] Salloum, S. A., Alomari, K. M., Alfaisal, A. M., Aljanada, R. A., & Basiouni, A. (2025). Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods. Smart Learning Environments, 12(1), 21.

[11] Barron-Estrada, M. L., Zatarain-Cabada, R., & Bustillos, R. O. (2019). Emotion Recognition for Education using Sentiment Analysis. Res. Comput. Sci., 148(5), 71-80.

[12] Yu, H. (2021). Online teaching quality evaluation based on emotion recognition and improved AprioriTid algorithm. Journal of Intelligent & Fuzzy Systems, 40(4), 7037-7047.

[13] Unciti, O., Ballesté, A., & Palau, R. (2024). Real-Time Emotion Recognition and its Effects in a Learning Environment. Interact. Des. Archit, 60, 85-102.

[14] Du, Y., Crespo, R. G., & Martínez, O. S. (2023). Human emotion recognition for enhanced performance evaluation in e-learning. Progress in Artificial Intelligence, 12(2), 199-211.

[15] Arena, F., Collotta, M., Pau, G., & Termine, F. (2022). An overview of augmented reality. Computers, 11(2), 28.

[16] Yılmaz, R. M., & Göktaş, Y. (2018). Using augmented reality technology in education. Cukurova University Faculty of Education Journal, 47(2), 510-537.

[17] Cabero-Almenara, J., Fernández-Batanero, J. M., & Barroso-Osuna, J. (2019). Adoption of augmented reality technology by university students. Heliyon, 5(5).

[18] Cabero-Almenara, J., Llorente-Cejudo, C., & Martinez-Roig, R. (2022). The use of mixed, augmented and virtual reality in history of art teaching: A case study. Applied System Innovation, 5(3), 44.

[19] Miralay, F. (2024). Use of Artificial Intelligence and Augmented Reality Tools in Art Education Course. Pegem Journal of Education and Instruction, 14(3), 44-50.

[20] Song, B. (2021). Virtual reality and augmented reality technologies for art education: The perceptions and responses of undergraduate students. Visual Inquiry: Learning & Teaching Art, 10(3), 361-369.

[21] Maraza-Quispe, B., Alejandro-Oviedo, O. M., Llanos-Talavera, K. S., Choquehuanca-Quispe, W., Choquehuayta-Palomino, S. A., & Caytuiro-Silva, N. E. (2023). Towards the development of emotions through the use of augmented reality for the improvement of teaching-learning processes. International Journal of Information and Education Technology, 13(1), 56-63.

[22] Shomoye, M., & Zhao, R. (2024). Automated emotion recognition of students in virtual reality classrooms. Computers & Education: X Reality, 5, 100082.

[23] van der Haar, D. (2020). Student emotion recognition using computer vision as an assistive technology for education. In Information Science and Applications: ICISA 2019 (pp. 183-192). Springer Singapore.

[24] Llurba, C., Fretes, G., & Palau, R. (2022). Pilot study of real-time Emotional Recognition technology for Secondary school students. IxD&A, 52, 61-80.