Hindawi

Research Article

Application of Artificial Intelligence Algorithm and VR Technology in Vocal Music Teaching

Liu Jing ID

School of Art, Chongqing College of Humanities Science & Technology, Chongqing 401524, China

Correspondence should be addressed to Liu Jing; 18404043@masu.edu.cn

Traditional intelligent algorithms mainly artificially preselect the geometric features of the image and then classify and recognize the image based on these features. The geometric features of the manually selected working images are often interfered by human factors, resulting in inaccurate feature extraction and reduced classification accuracy. Aiming at the above problems, an intelligent recognition method based on artificial intelligence algorithm is proposed. We will deeply analyze the application of the core technology of the system and introduce the virtual education function of the application through relevant content. The application of this technology can effectively improve the efficiency of vocal music teaching and enable students to obtain a richer educational experience. Then, this paper finishes the other parts of the system again, realizes the functional modules of the system one by one, and elaborates the corresponding interface diagrams. Finally, a suitable testing method can be used to test the performance modules and performance conditions of the system to ensure that the system has good reliability and availability and put it into production. The voice teaching system based on virtual reality technology is more advanced and superior than other teaching systems on the market today and can provide users with a more realistic experience. In addition, users can have more choices in the process of virtual education, thereby improving the efficiency and quality of education.

1. Introduction

In recent years, with the development of information technology, computer technology has been widely used in many fields and can be used to improve the quality of information technology. The main academic systems are academic and engineering management systems, academic affairs management systems, library management systems, etc. [1]. The conversion experiment of the virtualization model is carried out in a virtual environment. Through this alternate deployment environment design and simulation, the experimental results are simulated and the virtuality of Tadashi is emphasized. The effect of the experiment is usually equivalent and more effective in the real world. Every university needs to use its network infrastructure to simulate daily situations and problems in order to improve students' ability to find and solve problems. Internet + education has become popular, and many colleges and universities have applied VR technology [2, 3]. In actual classrooms, typical

applications are mainly VR innovation technology education, such as construction engineering, aerospace, etc. In fact, exhibition photos and videos of certain environments can only be displayed using traditional educational models, and relevant professional training is necessary. The application of virtual reality technology can combine all educational processes with actual practical projects. Students can brainstorm and create freely in the virtual environment, which helps to improve their innovative ability.

2. Related Work

By applying virtual reality technology to academic management, some studies can not only design various virtual situations for students, but also communicate with students in real time, so that students and teachers can learn from each other through the Internet [4]. Some research believes that the educational application of virtual technology is an innovation, which has solved some problems existing in the

traditional education model. In the actual teaching process, the choice of teaching methods and learning methods is very important, but the development of Internet technology has increased the difficulty and flexibility of choice, and the teaching methods have undergone great changes [5, 6]. Virtual education breaks the traditional education model. This innovative application can lead the development of education across the country and provide professional and technical personnel support for the progress and development of society. Some research proposed the concept of virtual reality technology and confirmed the world leader in this technology. In this era, the use of VR technology has entered many fields, especially in advanced technologies such as aerospace satellite technology and virtual reality technology. It is used for related training and high-end technology simulation exercises. In the military field, virtual battlefield environments and various forms of simulation training will be conducted to improve military operations capabilities and levels [7]. Some research takes into account the computer method; the existing map recognition efficiency and accuracy are very strong, so it is used for artificial intelligence calculation. By using intelligent recognition, we learned automatic features as input and achieved it through automatic failure and intelligent failure. Some research introduces the working graph recognition process of the deep learning model. In order to enter the deep learning model, the collected data must be preprocessed, including the normalization of the work graph data and the binarization and optimization of the work graph. Next, a working map label is created to finally display the details of CNN network and SSAENN network. Some research mainly introduces artificial intelligence and deep learning. Here, we will explain in detail the development process and application of artificial intelligence, the drawbacks of shallow learning to deep learning, and the advantages of using deep learning neural networks and then classify the model structures of deep learning and ordinary deep learning. This paper focuses on the model structure, algorithm principle and training process of convolution, and stacked sparse autoencoding neural networks in deep learning [8]. Some research believes that music teaching is mainly composed of classroom teaching and practical teaching. Classroom teaching is the foundation, and practical teaching is the assessment of class teaching. By discovering and solving the problems in the course, the efficiency of the course can be improved and the practical experience and expression ability can be improved. Through stage practice, students can discover and solve their own problems in time, enhance their psychological quality, and accumulate stage experience. This can enhance professional skills and increase students' stage experience [9].

3. Artificial Intelligence Algorithms and Virtual Reality Technology

3.1. Artificial Intelligence Algorithm. Deep learning is an algorithm model of deep neural networks. It is an extension of conventional neural networks and a branch of machine learning. Deep learning uses multiple nonlinear transformations to process data. It is an algorithm model with

various characteristics. They initially proposed the application of deep learning to solve the complex problems of training data, and these complex data are in the shallow network of the neural network, which can not be effectively reflected. Deep learning can imitate the cognitive principles of the powerful nervous system of the human brain. When the characteristic information input to the neuron structure is extracted hierarchically, the neurons in each level can extract the target information more deeply and decompose the extracted feature information. The characteristic information is decomposed. Assuming that the input targets of the neuron are $1x$ and $2x$, the linear transformation has the following form:

$$Z = w^T x + b. \tag{1}$$

The output expression is

$$h_{w,h}(Z) = f(Z) = f\left(w^T x + b\right). \tag{2}$$

In a deep learning network, there are many structures of such neurons, and each neuron transfers the abstract features of the data to the next layer through the nonlinear mapping of the activation function. The final function will appear as the level deepens. Finally, the learned abstract features of the deep data can be sent to the classifier to classify and recognize the data.

Since the deep learning network has multiple hidden layers, it realizes the mapping conversion from low-dimensional to high-dimensional through multilayer nonlinear transformation. Now, these data have complex structure, image rotation, translation, scale conversion, and other characteristics [10]. As shown in Figure 1, to create a $3\log(\cos(\exp(\sin(x))))$ with complex structure function, it is difficult to express this expression succinctly in a traditional shallow network, but by using a deep learning network, it can use a small number of parameters for layered expression, and the content to be expressed on each layer is simple.

In the process of shallow learning and deep learning, the input information will be nonlinearly transformed to create a mapping relationship from input to output, but when shallow learning is used to learn features, only features can be extracted manually and input into the model. Deep learning can use its own network structure to learn data features hierarchically and autonomously, but because the features it learns are too abstract, scientists are often confused and need further research.

3.1.1. Neural Network Infrastructure. Neural network is a new information processing mechanism extracted based on the basic understanding of the brain's organizational structure and activity mechanism. The neural network shows the characteristics of the human brain and has the basic functions of the human brain by simulating the activity mechanism and thinking principle of the human brain and nervous system [11]. In this network, the first-level detector processes the input data to make decisions and obtain simple results, and the second-level sensors process the first-level

$\text{log}\,(\text{cos}\,(\text{xep}\,(\text{sin}^3(\text{x}))))$

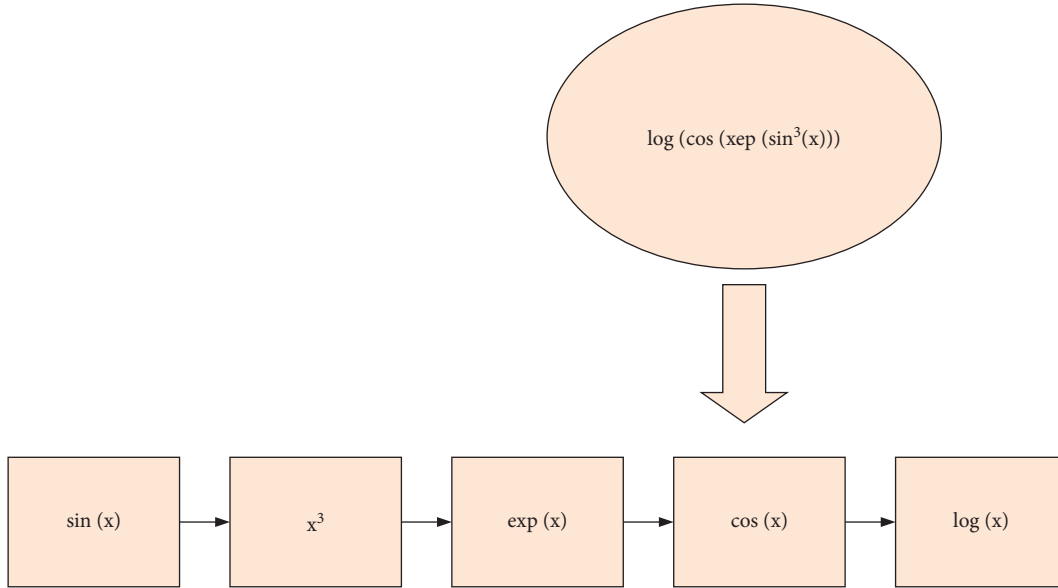| sin (x) | x^3 | exp (x) | cos (x) | log (x) |

FIGURE 1: Multilayer network structure expresses complex functions.

data to obtain the decision results. In this way, the second-layer perceptron can make more complex and abstract decisions than the first layer. Similarly, the third layer of the neural network can perform more complex data processing. In this way, a deep model of the distributed neural network is constructed recursively, and the input signal value and the corresponding weight are added to obtain the output. This is the most basic way for nerve cells to process signals, and it is also the basis for building neural network models. Let us use a function as the activation function of the neural network (here, the threshold function) to filter the results. The content so far can be abstracted and summarized with a formula:

$$\text{out} = f\left\{ \left| \sum_{i=1}^{n} w_{ij} x_i \left(t - t_{ij} \right) \right| - T_j \right\}. \tag{3}$$

3.1.2. Types of Neural Networks. The input data of the neural network is input from the input layer, and when passing through the interlayer neurons, the activation function will produce direct output data. The feedforward neural network has no feedback, so if you look at the topology of the neural network, you will find that the neural network is an undirected graph network. The output result of the feedforward neural network is the result of the interaction of each layer and the topology of the neural network, which describes the complexity of the activation function. Input data, weights, and offsets are input to the second layer after the activation function, and they eventually become the most commonly used chain structure in neural networks. The length of this chain is called the depth of the model, which is the origin of the term "deep learning" [12]. When using neural networks for training, the training samples will not directly specify the activity of neurons in other layers, but the learning algorithm will determine how to use these layers to produce the desired

output. All neuron nodes can receive input from the outside world and output it to the outside world and have the function of processing information. This loss function is an important reason for neural networks to generate parameter feedback and is the key to distinguish feedforward networks. The neural network dynamically adjusts parameters by comparing whether the loss function is minimized.

$$C = \frac{1}{2n} \sum_x \left\| y\left(x\right) - a^L\left(x\right) \right\|^2. \tag{4}$$

3.1.3. Overfitting Problem and Normalization. In practice, machine learning models are usually trained on well-trained data sets and the parameters in the model are constantly adjusted. We usually test the trained model on different test data sets after training and evaluate the quality of the model based on the test results. The error displayed by the machine learning model in the training data set is called the training error, and the expected value of the error that appears in any test data sample is called the generalized error. In order to calculate the training error and the generalization error, the aforementioned loss function can be used, for example, the square error used in linear regression and the cross-entropy loss function used in multiple logistic regression [13]. Overfitting is the main problem of neural networks, which is especially common in modern networks due to the huge network weights. In order to conduct effective training, relevant techniques are needed to detect whether overfitting occurs to avoid overtraining. If the accuracy of the test data is not improved, the training will stop. Increasing the number of training samples is one way to reduce overfitting and another way to reduce the size of the network. However, large networks have greater potential than small networks, so you can choose to apply redundancy. Here, we show the most commonly used normalization method, sometimes called weight decay or L2 normalization. The idea of L2

normalization is to add an extra term to the price function, which is called a normalization term. The normalized cross-entropy is

$$\text{loss} = -\frac{1}{n}\sum_{xj}\left[y_j\ln^{a_j^L} + 1 - y_j\ln^{\left(1-a_j^L\right)}\right] + \frac{\gamma}{2n}\sum_w w^2. \quad (5)$$

The first term is the conventional cross-entropy equation. The second item added now is the sum of the squares of the ownership weight, if we can use the quantized and adjusted factor to prohibit the parameters in the factor. Of course, other price functions, such as quadratic price functions, can be standardized. A similar standardization is as follows:

$$\text{loss} = -\frac{1}{2n}\sum_x \left\| y - a^L \right\|^2 + \frac{\gamma}{2n}\sum_w w^2. \quad (6)$$

Intuitively, the effect of normalization is that the network tends to learn smaller weights. Only when the first term of the price function can be increased sufficiently is a larger weight allowed. In other words, normalization can be seen as a manifestation between finding small weights and minimizing the original price function. These two parts are controlled by values: the smaller the value, the smaller the original price function with the smallest deviation; on the contrary, the larger the value, the larger the original price function with the largest deviation. The weight learning rules are as follows:

$$w \leftarrow w - \eta - \frac{\eta\lambda}{n}w = \left(1 - \frac{\eta\lambda}{n}\right)w - \eta\frac{(\partial C_0)}{\partial w}. \quad (7)$$

Except for adjusting the weight w by a factor, this is the same as the normal gradient descent learning rule. This adjustment reduces weight and is sometimes called weight attenuation. As a result, the weight continues to drop to zero. But this is not the case, because if the original price function falls, other terms will increase the weight. This is the principle of gradient descent. Therefore, the normalized learning rules of stochastic gradient descent are as follows:

$$w \longrightarrow \left(1 - \frac{\eta\lambda}{n}\right)w - \frac{\eta}{m}\sum_x \frac{(\partial C_x)}{\partial w}. \quad (8)$$

As an effective recognition model, convolutional neural network has received widespread attention. The predecessor of CNN was invented in 1980 and has since evolved into the current convolutional neural network. Convolutional neural network is mainly composed of convolutional layer (c layer), pooling layer (s layer), and fully connected layer (fc layer) [14]. There are many feature maps in each layer, and each feature map has multiple neurons arranged as shown in the figure. The main principle of CNN is to fold the partial two-dimensional image of the input layer with multiple different folding cores to generate a folded feature extraction layer c1 and then merge the folded layer c1. Obtain the converted feature mapping layer s2. The s2 layer is folded again to produce a folded layer c3 and the c3 layer is assembled again in the same manner as the s2 layer to produce a pooled layer s4. Finally, the feature map of the s4 layer is input to the fully

connected fc layer, and then the recognition and classification images are output, as shown in Figure 2.

Convolution operation fully reflects the two main attributes of CNN. Generally speaking, a local receptor area is a device in which each neuron in a hidden layer is only connected to neurons in a specific area of the layer, and the input data is window data in a local area of the image. Sharing weight means that each window data on the feature map shares a convolution kernel parameter set, that is, the same weight matrix [15]. By moving the window data around the image in fixed steps, the sum of the product of the window data and the convolution kernel is performed at each fixed step. Since it can be considered that the convolution process of the convolution kernel has been filtered by a filter, noise interference can be reduced and image characteristics can be emphasized. The selected window data is represented by JM, l jx represents the input of the j-th neuron in the feature map of the l-th convolutional layer, and $l1ix$ represents the output of the i-th neuron in the feature map of the l-th layer. f is the activation function, b is the bias parameter, $l jb$ is the bias parameter of the jth neuron of the lth convolutional layer feature map, k is the convolution kernel parameter, and $lijk$ is the lth convolutional layer feature map. Assuming it is the element of the convolution kernel parameter corresponding to the neuron, the convolution output equation is as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right). \quad (9)$$

The sigmoid function and the Tanh function are nonlinear, the saturation function has a saturation period, and the neural network training process will cause gradient diffusion. This is a diffusion problem, which has the characteristic of attaching most neurons directly to zero through a neural network, as shown in Figure 3.

After performing the convolution operation, the local features of the image can be extracted, but since the features of the image after the convolution are still many, it takes too much time to calculate when input to the classifier, and overfitting may occur. PCA must perform a merge operation, such as size reduction, to further reduce the image size. The pooling operation uses the overall characteristics of the adjacent output of the image area, rather than the network output at this location, and mainly includes the maximum pooling operation and the average pooling operation [16]. The maximum merging operation is an operation that selects the maximum value of the image area as the feature value in which the area is merged and can better extract the edge features of the image. The average pooling operation calculates the average value of the pooled image area as a feature value and better maintains the image background. Since the subject of this research is the working graph, it is necessary to extract and distinguish the edge features of the working graph, so the largest merging operation is selected. After the merge operation, the size of the feature map output to the merge layer will be reduced to the original $21/k$, and the useful features of the image will be extracted. The
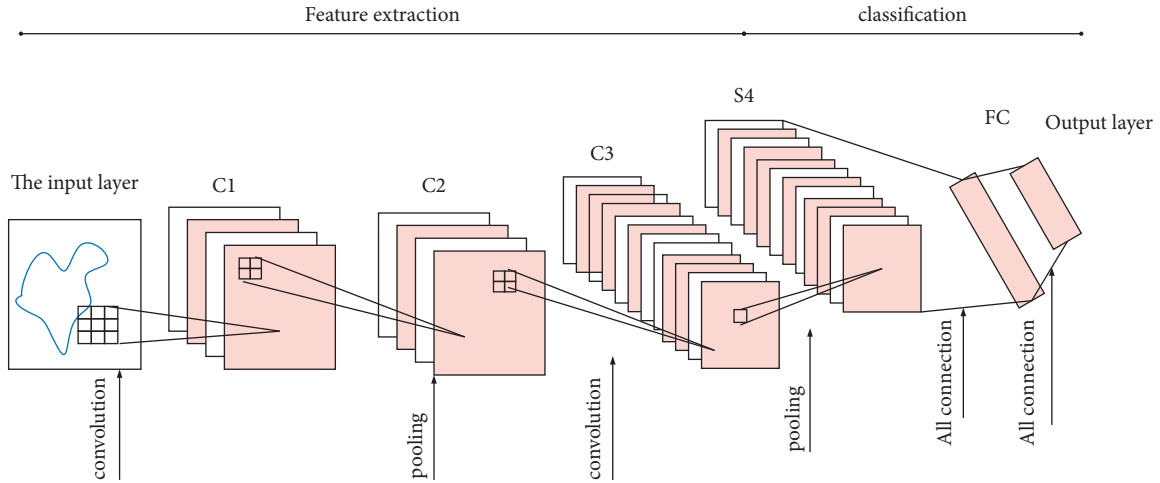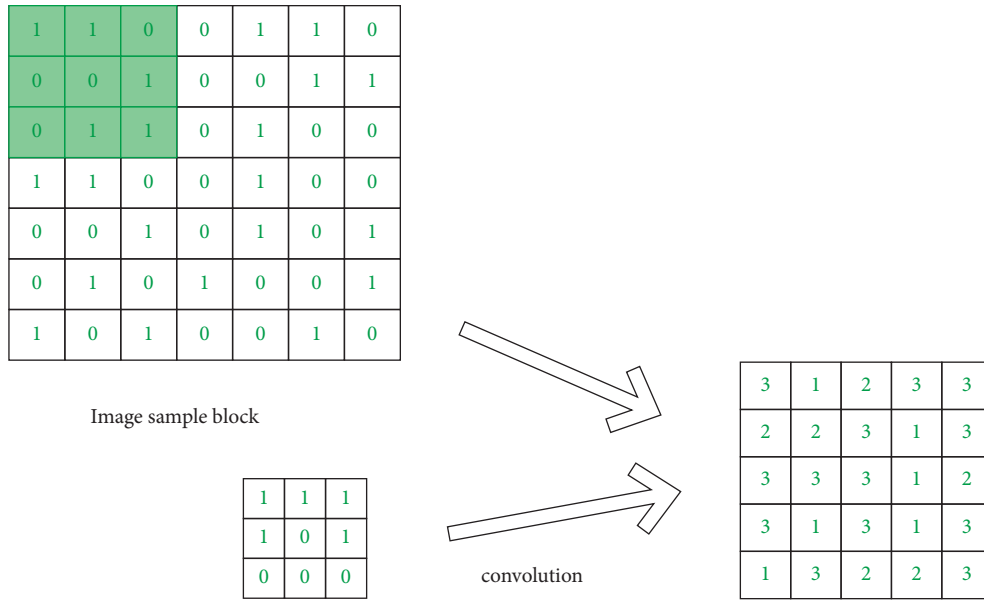
Figure 2: CNN network structure diagram.

Figure 3: Convolution operation.

expression of the pooling feature output of the pooling layer is expressed as follows:

$$x_j^l = f\left(\max\left(x_i^{l-1}\right) * w_j^l + b_j^l\right). \tag{10}$$

By arranging in a one-dimensional feature vector, data is output from all connected layers and then classified. The formula for all connection outputs is expressed as follows:

$$x^l = f\left(w^l x^{l-1} + b^l\right). \tag{11}$$

A set of test data is input into the neural network, and the convolutional neural network may not provide very good recognition accuracy. In other words, the network model will overfit the training set data and therefore cannot be well generalized to the test set. To avoid overfitting the network model, a dropout method is added to the network model to optimize the model. It is essentially a regularization method

applied to all connection layers of the CNN network. That is, the process of restoring the neurons in the next training to the original state by using only a part of the weight parameters and randomly selecting and operating until the end of the training. The test process embodies the idea of model averaging, which is equivalent to a network of different structures that has been activated and can operate all nodes. Each training will change the network structure. The adaptability to functions will be enhanced, and the network will have more generalization capabilities. Figure 4 shows the failure mechanism of some missing nodes in the network.

The user-based collaborative filtering algorithm finds products that users like from the relationship between different users' attitudes and preferences for the same products and content, calculates gender, and recommends users with the same taste. User-based collaborative filtering algorithm mainly includes two steps.
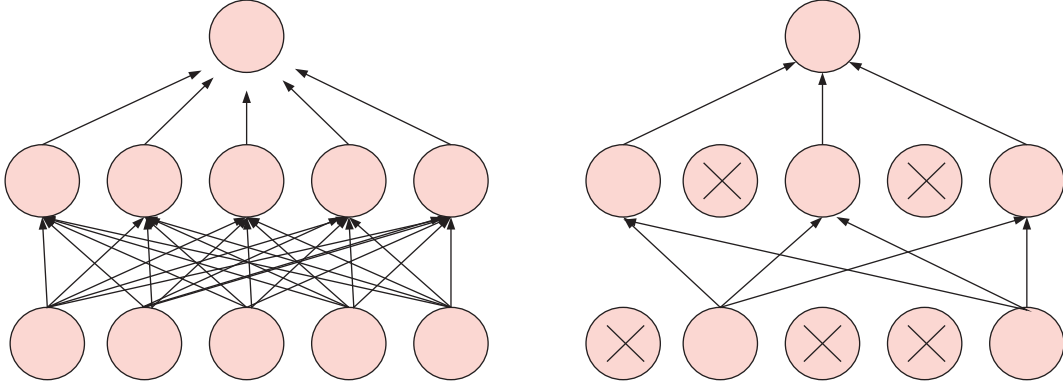
Figure 4: Dropout node failure mechanism.

(1) Calculate the similarity between users to obtain a similarity list.

(2) Search the similarity set, and find and recommend products that the user does not have among the users with the highest similarity.

The main point of step (1) is to calculate the similarity of interest of two users. Collaborative filtering algorithms mainly use the similarity of behavior. The interest similarity is calculated in the case of a given user u and user v, $n(u)$ represents a group of articles for which user u has positive feedback, and $n(v)$ user v represents an article for which positive feedback has been collected. Then, the interest similarity between u and v can be calculated by the following formula:

$$w_{uv} = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}. \tag{12}$$

Or it can be calculated by cosine similarity:

$$w_{uv} = \frac{|N(u) \cap N(v)|}{\sqrt{N(u)}\sqrt{N(v)}}. \tag{13}$$

Assuming that the interest list of user a is $\{b, c\}$ and the interest list of user c is $\{a, b, d\}$, the similarity between user a and user c can be calculated.

$$w_{AC} = \frac{|(b,c) \cap (a,b,d)|}{\sqrt{(b,c)}\sqrt{(a,b,d)}} = \frac{1}{\sqrt{6}}. \tag{14}$$

The user cf algorithm recommends products that users did not intend to buy in the past. For example, step (3) measures the user's interest in the product.

$$P(u,i) = \sum_{v \in S(u,k) \cap N(i)} w_{uv} r_{vi}. \tag{15}$$

The similarity between the collaborative recommendation algorithm and the calculation will cause errors, and this error will be applied in practice. For example, listing popular products of common interest of users will affect the calculation of similarity.

When designing a recommendation system, there is usually more than one recommendation algorithm. In order to choose a suitable recommendation algorithm, it is necessary to analyze the data and solve the problem of information overload according to user needs.

As shown in the figure, the autoencoder uses unsupervised learning to perform network pretraining, encodes and reconstructs the input data, minimizes reconstruction errors, and characterizes the hidden layer of the input data. The data encoding process is from the input layer to the hidden layer, and the reconstruction process is from the hidden layer to the output layer. Here, the +1 in the circle is called the offset coefficient. w and b are autoencoders, including the following: (1) w and (1) b connect the input layer and the hidden layer, and (2) w and (2) b connect the hidden layer and the output layer. It is the weight matrix and the deviation matrix. It can be seen that the number of input layer nodes is the same as the number of output layer nodes. In other words, the autoencoder tries to learn the features in the hidden layer, optimizes the output Servlet x to make it as close to x as possible, and minimizes the reconstruction error. Finally, the feature extracted by the hidden layer is the input layer, as shown in Figure 5.

The hidden layer elements extracted by the autoencoder cannot effectively represent the data in the input layer. The sparse limit means that the output is activated when the value of the output function of the value neuron is as close to 1 as possible, and when the value of the output function of the value neuron is as close to 0 as possible, most of the output is suppressed. The completed operation is called the sparse limit.

$$\widehat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}\left[a_j(x')\right]. \tag{16}$$

Such dilution constraint is realized by adding a constraint factor to the autoencoder loss function by the neuron. The disciplinary factors are expressed as follows:

$$KL\left(\rho\|\widehat{\rho}_j\right) = \rho\,\log\frac{\rho}{\widehat{\rho}_j} + (1-\rho)\mathrm{loglog}\frac{1-\rho}{1-\widehat{\rho}_j}. \tag{17}$$

The loss function of the model is expressed as

$$J(W,b) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left\|h_{W,b}(x^i) - x^i\right\|^2. \tag{18}$$
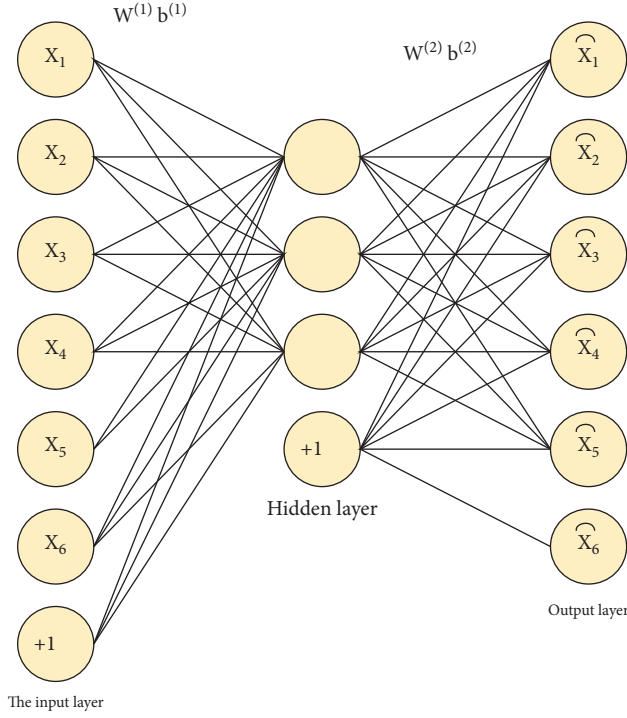
Figure 5: Autoencoder structure.

The loss function of the sparse autoencoder increases the discipline factor and limits the sparsity, which is expressed as follows:

$$J_{\text{square}}(W, b) = J(W, b) + \beta \sum_{i=1}^{n} KL(\rho \| \hat{\rho}_j). \tag{19}$$

The size of the reconstruction error indirectly indicates the effectiveness of the encoding process used to extract sparse features. Since the size of the reconstruction error is evaluated by the loss function of the sparse autoencoder, the optimal weight matrix w offset matrix b can be obtained by using the backpropagation algorithm to minimize the equation, and finally is the sparse representation of the hidden layer of the input data characteristics.

Soft max is used as the input of the classifier in the hidden layer of the stacked sparse autoencoder, so that the samples can be finally classified. Multiclassification problems usually use the Soft max classifier to classify the features learned by the stacked sparse autoencoder. The Soft max classifier is a generalization of the logistic regression classifier for multiclassification problems. Its hypothesis function is expressed as follows:

$$h_\theta[x(i)] = \begin{bmatrix} p(y(i)) = 1 | x(i); \theta \\ p(y(i)) = 2 | x(i); \theta \\ \vdots \\ p(y(i)) = k | x(i); \theta \end{bmatrix} = \frac{1}{\sum_{i=1}^{k} e^{\theta_j^T x(j)}} \begin{bmatrix} e^{\theta_1^T x(i)} \\ e^{\theta_2^T x(i)} \\ \vdots \\ e^{\theta_k^T x(i)} \end{bmatrix}. \tag{20}$$

Using the maximum entropy model, the loss function is expressed as follows:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y(i) = j\} \log \frac{e^{\theta_j^T x(i)}}{\sum_{i==1}^{k} e^{\theta_j^T} x(i)} \right]. \tag{21}$$

The weight attenuation term is added to the loss function to reduce overfitting. This attenuation term has too many parameter values. At this time, the loss function is as follows:

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y(i) = j\} \log \frac{e^{\theta_j^T x(i)}}{\sum_{i==1}^{k} e^{\theta_j^T} x(i)} \right]$$
$$+ \frac{\lambda}{2} \sum_{i=1}^{k} \sum_{j=0}^{n} \theta_{ij}^2. \tag{22}$$

In the decentralized process, that is, the data processing process, the goal is to extract and process information about the accuracy of system design, which is characterized by the data that determines the user's purchase of products and the information in the user's own context. The information data system obtains the ideal information by reducing the interference to the data information. We aggregate the purchase status of customers for each product and obtain information about customers' purchase intentions. The accuracy and efficiency of neural network training are generally low, and as a result of classification, severe deviation is essential. In addition to data cleaning, data preprocessing must also consider the differences between data units and quantities. Therefore, a series of processes should be performed, such as data integration, conversion, and standardization. Data preprocessing aims to improve data quality while making it more suitable for specific mining techniques or tools. Before training a neural network, it is usually necessary to preprocess the input data. Due to the different units of the input data, some of the data is very large and some of the data is very small, which may cause the training time of the neural network to be too long and the convergence speed to be too slow. For a data set with a large input data distribution, the effect of the neural network classification algorithm may be too large, and the input effect at a point with a small data range may be small. Since the value range of the activation function of the output layer of the neural network is limited, it is necessary to map the target data of the network training to the value range of the activation function. The usual processing method is the normalization method of mapping the data to [0, 1] or [−1, 1]. The normalization algorithm is as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{23}$$

So far, the source code is carried out from the source site where the product exists through the crawling algorithm. Therefore, the analysis module needs to crawl the source code according to the analysis rules. The data obtained has various defects such as omissions and ambiguities. The neural network processes it first. The collected data is shown in Table 1.

User number	Product code	Property 1	Property 2	...	Property n	Category
19127	1	0.81472	0.16218		0.059619	1
19127	2	0.90579	0.79428		0.68197	1
19127	3	0.12699			0.042431	1
19127	4	0.91338	0.16565		0.071445	1

When dealing with missing data in a table, we show two ways to delete or insert data with columns. Since we are using a search engine to obtain n-dimensional product attribute data, the amount of data is very small. In order to ensure the integrity of product information, we interpolate missing data. The interpolation methods often used in data preprocessing include Lagrange interpolation, Newton interpolation, and average calculation. Substitute the coordinates of n points into the polymorphic function:

$$L(x) = \sum_{i=0}^{n} y_i \prod_{j=0, j \neq i}^{n} \frac{x - x_j}{x_i - x_j}. \qquad (24)$$

Lagrange's interpolation formula is very useful for theoretical analysis and is widely used in data processing, but the following formula changes as the number of missing values increases or decreases. However, Newton interpolation is rarely used in practice. Newton interpolation is a formula of n known points $(x_1, y_1), (x_2, y_2)...(x_n, y_n)$.

$$f[x_{rv}x_{n-1}, \ldots, x_1, x]$$
$$= \frac{f[x_{n-1}, \ldots, x_1, x] - f[x_n, x_{n-1}, \ldots, x_1]}{x - x_n}. \qquad (25)$$

Due to too much product feature data, it is necessary to reduce the data size. The commonly used data is replaced by a small number of new variables to reflect as much of the original variable data information as possible. In addition, since the new variables are orthogonal to each other, information that duplicates the original variables can be removed. The Lagrangian function is constructed as follows:

$$L = t_1'Vt - \lambda_1(t_1't_1 - 1). \qquad (26)$$

Its λ is the Lagrangian coefficient. Calculate the partial derivative of L with respect to λ, and set it to zero; then:

$$\begin{cases} \dfrac{\partial L}{\partial t_1} = 2Vt_1 - 2\lambda_1 t_1 = 0 \\[2mm] \dfrac{\partial L}{\partial \lambda_1} = -(t_1't_1 - 1) = 0 \end{cases}, \quad Vt_1 = \lambda_1 t_1. \qquad (27)$$

It can be seen that λ is its corresponding eigenvalue. At this time

$$Var(P1) = t_1'Vt = t_1'\lambda_1 t_1 = \lambda_1 t_1 t_1' = \lambda_1. \qquad (28)$$

Principal component regression analysis = principal component analysis + multiple linear regression analysis. The product attribute data selected here is 100*400. Through analyzing the formula, it is found that the quasi-component of the data is determined, and appropriate neural network input data can be obtained. Principal component analysis mainly includes the following two steps:

(1) Standardize the data in all product attributes. In standardization, the average value of each attribute is set to 0, and the standard deviation is set to 1, to eliminate the quantitative difference between attributes. The sample standardized input variable matrix is

$$X = \{x_{i,j} | i = 1, \ldots, N, j = 1, \ldots, M\}. \qquad (29)$$

(2) Obtain the correlation matrix between product attributes. The correlation matrix allows highly correlated indicators, and the covariance between these indicators can be referred to as another variable of the first component. After removing the first component, the residual correlation matrix is calculated. This residual correlation matrix gives the second set of related variables. These codispersions can be replaced by a second component, where the second component and the first component are orthogonal. In this way, it continues until all the distributions of the original product attribute data are extracted.

$$P_1 = X * t_1, \|t_1\| = 1. \qquad (30)$$

Similarly, by inputting the information of the variable matrix, it can be seen from the perspective of probability statistics that the greater the variance of the variable, the more information the variable contains. Therefore, the above problem requires the variance of the variable p to be the largest. The dispersion of p is expressed by the following formula:

$$\text{Var}(P_1) = \frac{1}{n}\|P_1\|^2 = \frac{1}{n}t_1'X'Xt_1 = t_1'Vt_1, \quad V = \frac{1}{n}X'X. \qquad (31)$$

Usually the amount of data information in the first principal component is very large, and it is inferred in descending order. After obtaining the result graph of the component, the data dimension of the component is projected into the original space, which is the process of expanding the data sample.

3.2. Design of Vocal Music Teaching System Based on Virtual Reality Technology. Virtual reality education system is not only software development technology, network, and
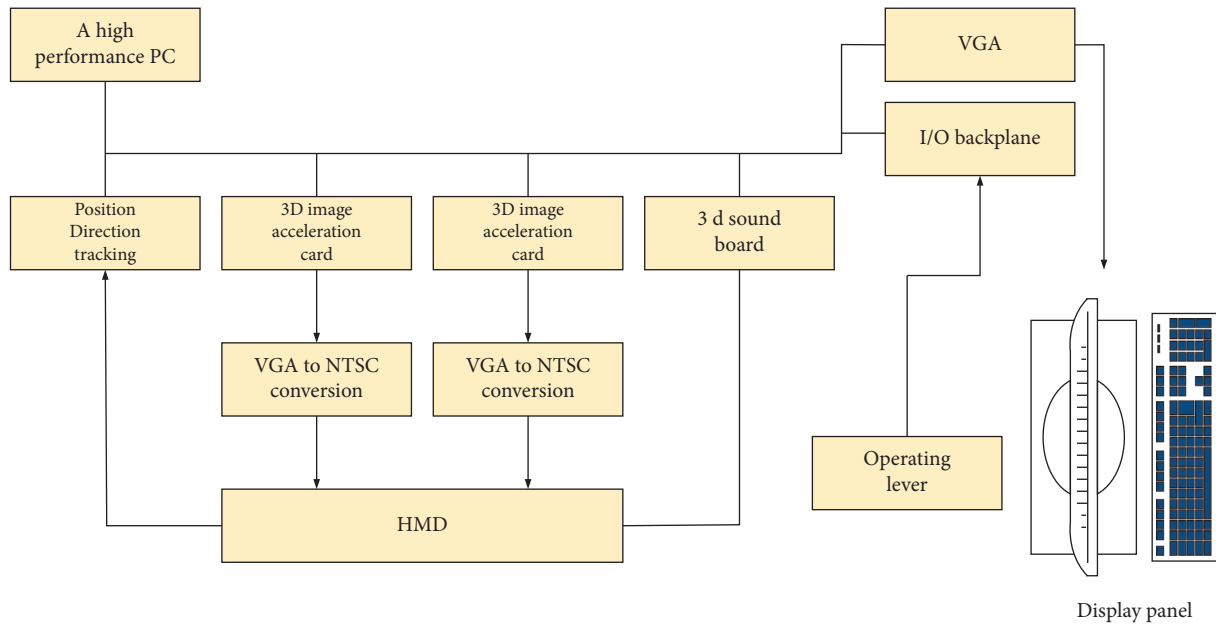
FIGURE 6: System architecture design diagram.

TABLE 2: Student information form.

Numbering	Field name	Field type	/F	Is null	Field depict
1	jx_zhbh	Int	Primary key	No	Account number
2	jx_yhc	Varchar (20)		No	User's nickname
3	jx_xb	Varchar (20)		No	Gender
4	jx_nl	Varchar (20)		No	Age
5	jx_txdz	Varchar (20)		No	Contact address
6	jx_jtzz	Varchar (20)		No	Home address
7	jx_yx	Varchar (20)		No	Mailbox
8	jx_sjhm	Int		No	Mobile phone number
9	jx_xh	Int		No	Student id
10	jx_sxzy	Varchar (20)		No	Major
11	jx_tcah	Varchar (20)		No	Special hobby

communication technology, but also many technologies. It is related to the hardware infrastructure of the profession. The virtual reality education system not only has basic education functions, but also needs to realize real-time interaction between humans and machines, which requires high system performance. With the development of science and technology, the cost performance of personal computers has been improved, and many application systems use high-performance PCs to support the system. The figure below intuitively illustrates the design of deploying a virtual reality system on a high-performance PC and the principle of use.

As can be seen from Figure 6, a complete virtual education system requires not only basic equipment, but also special hardware configuration, such as location tracking equipment, 3D image acceleration card, 3D sound board, etc.

When designing a database, information table design is the foundation. The information table consists of an information table and multiple pieces of information in the information table and is usually field type, field name, primary key information, etc. Because there are many functional modules in the system, each functional module displays only one information table, especially student information table, homework information table, scene information table, educational resource information table, examination question information table, and interactive information table.

Table 2 mainly contains information such as account number, nickname, gender, age, communication destination, major, special skills, hobbies, and majors.

(1) Job information table: The job information table mainly includes job number, subject, faculty, placement time, deadline, class, job details, and other information. For detailed information about job information, please refer to Table 3.

(2) Scene information is the main component of the 3D model. It is necessary to design a larger scene based on a larger teaching plan and teaching content and systematically record relevant information about the scene in detail. For details of the scene information table, please refer to Table 4.

TABLE 3: Work information sheet.

Numbering	Field name	Field type	/F	Is null	Field depict
1	jx_zhbh	Int	Primary key	No	Job number
2	jx_km	Varchar (20)		No	Subject
3	jx_krls	Varchar (20)		No	Subject teacher
4	jx_bzsj	Datetime		No	Set up time
5	jx_jssj	Datetime		No	Deadline
6	jx_bi	Varchar (20)		No	Class
7	jx_yx	Varchar (20)		No	Job details

TABLE 4: Scene information table.

Numbering	Fie Id name	Field type	P/F	Is null	Field depict
1	jx-tjbh	Int	Primary key	No	Field number
2		Datetime		No	Build time
3	jx_yssl	Int		No	Number of elements
4	jx_mrs	Varchar (20)		No	Capacity
5	jx_cjlx	Varchar (20)		No	Scene type
6	jx_sysj	Dttetime		No	Usage time
7	jx_sbsl	Int		No	Equipment quantity
8	jx_sykm	Varchar (20)		No	Use subjects

TABLE 5: Educational resource information form.

Numbering	Field name	Field type	/F	Is null	Field depict
1	jx_zybh	Int	Primary key	No	Resource number
2	jx_zykx	Varchar (20)		No	Resource type
3	jx_scsj	Datetime		No	Upload time
4	jx_xscs	Int		No	Download times
5	jx_scry	Varchar (20)		No	Uploader
6	jx_sskm	Varchar (20)		No	Subject

TABLE 6: An interactive information sheet.

Numbering	Field name	Field type	/F	Is null	Field depict
1	jx_zybh	Int	Primary key	No	Message number
2	jx_zykx	Varchar (20)		No	Publisher
3	jx_scsj	Varchar (20)		No	Interactive staff
4	jx_xscs	Varchar (20)		No	Type of interaction
5	jx_scry	Datetime		No	Interaction time
6	jx_sskm	Datetime		No	Release time
7	jx_fbdd	Varchar (20)		No	Release location

(3) The teacher resource information table mainly contains information such as resource number, resource type, download time, download times, uploader, and affiliation, as shown in Table 5.

(4) The interactive information table mainly includes information number, caller, distributor, delivery type, delivery time, delivery location, etc. For more information, see Table 6.

This article is based on the voice teaching system of virtual reality technology to carry out related elaboration. There are relatively many types of voice music involved, and the design only for teaching is introduced in detail. It involves modeling and final realization of people. The complexity of 3D modeling may affect the realization effect of VR technology to some extent. Improving the efficiency and level of 3D modeling is an important part of virtual reality. In a specific implementation, it is first necessary to clearly analyze the environment of the scene. The BNF-based paradigm rules provide a modeling language for the scene, create a dance model, and make it possible. They can demonstrate works on the stage and finally meet the students' independent learning needs, so that the entire system has good interactivity, including three main structures of the 3D model, one is the shape of the human body, the second is the action, and the third is the facial action. The basic principle of the model's behavior in this process is based on the changes in human bones and muscles. Therefore, based on biological principles, these expressions are expressed by a computer system that builds the surface model of the actor

by modeling using 3D modeling technology and finally realizes the work of ballet. It is lively and can stimulate students' interest in learning sounds. The complexity of the 3D model is inversely proportional to the speed of the system. That is, the larger the data of the constructed model, the more complex the model, and the longer the drawing time, but the more realistic the display effect, and vice versa. Creating models requires dedicated 3D modeling software, but the most popular software today is the 2013 version of 3Dmax, which has been used in various design fields and has a very good user experience. The software is also used for the formation of surface models.

In order to model the details, polygon modeling technology must be used to fine-tune the head and body of the model, and the details can be processed and adjusted by manipulating the surface of points and lines. The human face has a natural three-dimensional shape, and recognition based on the 3D data of the human face is a common method to solve the pose problem. Therefore, it is necessary to start with the smallest details. In particular, eye treatment is an important part of a person's vitality and has a major impact on the postrendering effect. At the same time, you need to pay attention to the beginning of the eye pupil. As you can see, hard-edging is not always possible. Otherwise, the actual effect in the later stage will seriously affect the effect. Body processing is also very important, especially in the processing of each joint, to control the number of model faces to prevent the model from stretching due to changes in future work.

The optimization and improvement of 3D deformation has also become one of the hottest issues in this field, especially by building a complete head model to improve modeling efficiency and further improve angles such as facial elements. Here, the model formed by the UV editor is used so that the visual effects of the actors can be reflected, and the texture, color, and texture effects of the points corresponding to the actual actors can be presented. After rendering the 3D model, choose the characteristics of a real person.

In order for an actor to perform a movement, he must have a skeletal system. The key to this process is to build a bone model and use the movement of the bone to represent the movement of the person. After running the program, execute skeleton processing through "Biped" in the "Create Panel" function menu. This process is the detailed processing part of the design. If each joint is not connected properly, it will have an adverse effect on the performer's subsequent movement. After completing this process, go through 3D max to check whether bone changes and muscle stretching will affect the model composition. At the same time, it is necessary to set the weight value for the model, which is a masking task, and the processing of the model joints is very important. The arrangement status of the weight parameters is displayed in the weight value on the left. If you want to adjust these parameters, you can adjust the value on the right. On the other hand, when managing the intensity of the model and the hue of heating and cooling, the heating and cooling can be adjusted through the overcolor adjustment of the model surface to express the real effect of the actors.

Since the design of the 3D model is based on different scenes, it is necessary to construct the scene before obtaining the basic motion data. In order to collect the movement data of each level, 10 shots of Eagle-4 were placed, and the complete movement was collected in Eagle, two of which were placed on both ends of the Eagle-4 lens. Divide into four groups, each group of eight, one of which stays in the corner of the triangle guard, and then capture the route from the angle of the hand. The other group can be placed at a higher position, that is, at the four corners above the stage, and the stage can be used in all directions. After deploying all the lenses, power them up uniformly and connect them to the PC desktop in the interface format, and complete the integration process of the basic work data of each Eagle lens through the PC. In order to easily manage the basic data, each shot needs to be numbered, and when searching for motion data, only the shot number is required to search for motions in a specific direction, which is very convenient. One is to put a "field"-shaped label on the ground to identify each area of the field. Secondly, in order to facilitate the later synthesis process and enhance the capture effect, it is necessary to identify obstacles in the field and capture data from all angles. There are always 30 dancing models, wearing special dance costumes, and marking points at their designated locations. In this way, the operating data can be reflected on the collector. In order to facilitate the rule management of these markers, the parameter setting part of ModelEdit will rename each marker according to scientific rules.

The performance on stage can objectively express students' understanding of knowledge. Performers will process and reorganize all life experiences expressed on stage based on their own real life experiences and emotions, combined with the professional knowledge gained in the learning process. After creation, it will appear in front of the audience in the form of voice. When singing onstage, pay attention to the matching of work. It is not suitable for clumsy singing and complicated work. We are looking for "Kamigata combination." On the stage, the actors and singing are integrated, the works come from the heart, and the pursuit of its beauty cannot be ignored. For this, we must develop regular practice, habits, and professional performance skills.

The speech course with stage practice is incomplete. Without the speech course, the stage practice will not be able to play its original role. In addition to regular courses, practice is also very important. Practice can be used as an effective method to verify the quality of education and can verify whether the quality of education affects the results of practice. Performers need to spread professional theoretical knowledge and advanced music skills to a larger audience. Voice exercises are a form of classroom knowledge. Students can quickly discover their own shortcomings in the practice process, so as to make effective corrections and improve the level of theoretical knowledge in the correction process. It not only effectively improves the professional level of vocal music, but also further expands the practice circle. The experience continuously summed up in practice will benefit the sound development of vocal music education in the future.

Good psychological quality is the foundation of student stage performance. Students must not only learn the skills of voice music, but also fully demonstrate their talents on the stage, enhance their expressiveness on the stage, and fully demonstrate their strengths.

(1) To cultivate stage expression ability in the classroom, teachers are required to actively cultivate students' performance ability and improve their stage expression ability. For the works familiar to students, let them create their own feelings and behaviors, and let them express better with words and gestures. In class, invite other students to practice singing and acting. Of course, it is best to record the performance of students, let them observe for themselves, find out what is good or bad for them, and let them correct immediately. Only in this way can we train the stage adaptability and psychological background.

(2) Training for overcoming stage phobia, the position, lighting, audience, etc. are different from those during practice, so it will deepen tension and fear. The desire to get high scores intensifies the psychological pressure and makes students feel emotionally nervous both physically and mentally. In order to solve this psychological problem, students must have a good mental state and the ability to overcome stage fear. However, this kind of courage and attitude cannot be strengthened through regular practice, only through practice. Students need self-confidence to succeed, and they need basic skills to gain self-confidence. Therefore, students should be prepared before performing, walk onto the stage with a good attitude, be familiar with the venue, and suggest that they can succeed.

(3) According to the actual situation of the students, choose songs suitable for stage practice and master the ability and psychological state of the songs to choose the more difficult songs. According to actual problems, according to the differences of the students' tone, singing voice, and intonation, choose appropriate and reasonable songs. At the beginner stage, the choice of songs should generally be slightly lower than the actual level of the students, which not only achieves the purpose of adequate practice in this way, but also helps to enhance students' self-confidence.

4. Conclusion

In this article, we compare deep learning and its improved model with other demonstration images of shallow network recognition, showing that deep learning and its improved model perform better in the recognition of demonstration images. After in-depth update and perfect neural network, the model can avoid the difficulty of extracting geometric features of some images and the phenomenon that the extracted features are unreasonable. The application of virtual reality technology in the education field can break the traditional education model and provide students with special virtual scenes, so that students can deal with specific scenes through their own knowledge. It can improve students' practical ability. The traditional education model cannot stimulate students' interest in learning, but more importantly, they cannot use the professional knowledge they have learned. In this article, we have developed and designed an art education system based on virtual reality technology to make up for these shortcomings. Aesthetic education occupies a very important position in phonetic education in colleges and universities, and it is particularly important to inculcate aesthetic education into phonetic education. To instill aesthetic education into high school phonetics courses, it can be mainly manifested as follows: first, improve teachers' professional skills and enrich teachers' teaching methods; second, improve students' understanding of musical instruments; third, carry out some music performance activities and enhance students' practical ability. Finally, the integration of emotional experience into speech can enhance the beauty of students and promote their healthy development.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] A. O. Ojo, M. Raman, and A. G. Downe, "Toward green computing practices: a Malaysian study of green belief and attitude among Information Technology professionals," *Journal of Cleaner Production*, vol. 224, no. 7, pp. 246–255, 2019.

[2] V. Fratto, M. G. Sava, and G. J. Krivacek, "The impact of an online homework management system on student performance and course satisfaction in introductory financial accounting," *International Journal of Information and Communication Technology Education*, vol. 12, no. 3, pp. 76–87, 2016.

[3] U. Hüseyin, T. Murat, and P. Y. Ezgi, "The effects of the authentic learning approach with a course management system (moodle) on students mathematics success and online authentic learning self-efficacy," *Educational Research and Reviews*, vol. 15, no. 11, pp. 679–689, 2020.

[4] Z. Lv, "Virtual reality in the context of internet of things," *Neural Computing & Applications*, vol. 32, no. 13, pp. 9593–9602, 2019.

[5] K. Ibraheem Arif, "Building an online course management system for Iraqi education colleges," *Indian Journal of Science and Technology*, vol. 11, no. 7, pp. 1–6, 2018.

[6] L. W. Wardana, "Paper airplane and talking stick learning methods to increase students understanding about management information system courses," *IOSR Journal of Business and Management*, vol. 18, no. 09, pp. 164–169, 2016.

[7] D. Zhou, "The development status of VR technology and its application field research practice," *Electronic Technology and Software Engineering*, vol. 25, no. 17, pp. 147-148, 2018.

[8] Z. Lv, X. Li, and W. Li, "Virtual reality geographical interactive scene semantics research for immersive geography learning," *Neurocomputing*, vol. 254, pp. 71–78, 2017.

[9] A. N. Alkhaldi and A. M. Abualkishik, "Predictive factors for the intention to adopt a mobile blackboard course management system: the case study of university of hail in Saudi Arabia," *Indian Journal of Science and Technology*, vol. 12, no. 19, pp. 1–12, 2019.

[10] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2218–2234, 2020.

[11] Y. Gao, L. Gao, X. Li, and X. Yan, "A semi-supervised convolutional neural network-based method for steel surface defect recognition," *Robotics and Computer-Integrated Manufacturing*, vol. 61, Article ID 101825, 2020.

[12] H. Zhao, H. Liu, W. Hu, and X. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network," *Renewable Energy*, vol. 127, pp. 825–834, 2018.

[13] A. Caggiano, J. Zhang, V. Alfieri, F. Caiazzo, R. Gao, and R. Teti, "Machine learning-based image processing for on-line defect recognition in additive manufacturing," *CIRP Annals*, vol. 68, no. 1, pp. 451–454, 2019.

[14] X. Li and S. Wang, "Object detection using convolutional neural networks in a coarse-to-fine manner," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2037–2041, 2017.

[15] T. Chen, S. Lu, and J. Fan, "S-CNN: subcategory-Aware convolutional networks for Object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2522–2528, 2018.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, IEEE, Honolulu, HI, USA, July 2017.