

On the Gains and Losses of Multimedia-Assisted Instruction Technology in College Music Teaching Practice

Dalei Feng, Zhaoqing University, China*

ABSTRACT

Applying multimedia technology to college music teaching can directly present some abstract knowledge to students. This not only breaks the limitation of traditional teaching time and space, but also enriches the teaching content and improves the teaching quality. Therefore, it is very important to change the teaching concept, actively practice and explore, and seek a new teaching mode. This article deeply analyzes the application of multimedia in college music classroom teaching, the role and position of multimedia in music teaching, compares the advantages and limitations of multimedia technology with other educational technologies in music classroom teaching, and explains multimedia educational technology. Combined with teaching practice and courseware development, this article analyzes the application status of multimedia technology in middle schools, and demonstrates the position and role of multimedia teaching methods in music classroom from both teachers and students, so as to provide some references for the reform of music education in colleges and universities.

KEYWORDS

Convolutional Neural Network, Deep Reinforcement Learning, Multimedia, Music Classroom, Teaching

INTRODUCTION

With the promotion of quality education and the development of multimedia technology, the traditional music teaching methods have been far from meeting the needs of contemporary music education, and cannot achieve the ideal educational effect. The development of modern science and technology has brought many new technologies into the music classroom, providing a wealth of teaching resources for the music classroom. The use of multimedia technology has improved the artistry and intuition of the music classroom, providing an effective way to achieve curriculum standards, and the improvement of teaching quality has been unanimously recognized by teachers and students. The main goal of music teaching in colleges and universities is to cultivate students' ability to express, appreciate and create music. Multimedia technology integrates graphics, text and sound, and has gradually been widely used in teaching (Zereu, 2023). The traditional music teaching model leads to a relatively weakened teaching process and pays more attention to the production of results. Teachers become knowledge instillers, and students become recipients of knowledge. This approach undoubtedly restrains the development of students' creativity, imagination, discovery ability, and independent

DOI: 10.4018/IJWLTT.330646

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

inquiry ability. In addition, this approach ignores students' autonomous learning ability, and their ability to solve problems becomes relatively weak. Traditional teaching methods can no longer keep up with the rapid development of modern society, so the use of multimedia computer technology and other equipment for teaching has become an inevitable trend in the development of modern teaching technology (da Silva-Santos et al., 2023). The application of multimedia technology in music teaching in colleges and universities can first make the teaching more vivid with pictures and texts, and make the abstract image of music concrete, visualized, and intuitive, so that students can integrate into the music situation to the maximum extent and feel the music. Facing the rapid development of society, music teachers should effectively integrate modern information technology multimedia technology into music. How to improve teaching efficiency in classroom teaching, which caters to the inevitable trend of social development, is also a technology that music teachers must master (Nitin, 2023). With the rapid development of the information age, the use of multimedia computer technology and other advanced teaching equipment for auxiliary teaching has become an inevitable trend in the development of modern teaching technology.

Multimedia teaching is the preferred teaching method for educators, at present. It occupies the majority of teaching classrooms with the advantage of being more concise, intuitive, clear, and figurative. The application of multimedia teaching has become a common teaching method in college music teaching. Students are more interested in what they have learned through various displays of multimedia, so as to improve the classroom efficiency and create an efficient classroom. However, many schools have only installed the equipment in classrooms, due to their lack of understanding of multimedia equipment, and have not demonstrated the advantages of multimedia reasonably and subtly, and even conflicted with the inherent traditional teaching mode. This shows that there is unreasonable use of multimedia in the teaching process. These problems have become increasingly apparent, over time. Multimedia teaching is still in the initial stage of development. Through the continuous research and improvement of music educators, the advantages of multimedia application in music classroom teaching can be fully demonstrated (Berthaut, 2020). College music teachers use multimedia technology to make single and abstract music become multi-dimensional and specific. With the support of technology, music teaching has become dynamic and an important auxiliary means in the process of music teaching. However, in practical application, some teachers rely too much on multimedia technology, resulting in misuse and abuse. Therefore, in the process of music teaching, the use of multimedia must follow its use principles. Teachers need to be trained to ensure that they fully master the application methods of multimedia technology, so as to achieve the effectiveness of classroom teaching.

In this paper, the author deeply analyzes the application of multimedia in music teaching in colleges and universities, investigates the role and position of multimedia in music teaching, compares multimedia technology with other educational technology, and explains multimedia educational technology. Speech recognition technology is one of the core technologies of human-computer interaction, and also a frontier topic in the field of natural language processing. Multimedia is a new educational media composed of a large amount of visual and auditory information. The application of modern teaching methods can help students better learn and understand music, and can bring students intuitive and rich teaching experience. According to the basic structure rules of convolutional neural networks (CNNs), the author designed a Multilayer Residual Deep Convolutional Neural Network(MR-DCNN) model for music teaching and applied it to speech recognition. This model can better distinguish the features of different music types and how the clips change with time.

LITERATURE REVIEW

The use of multimedia in teaching has become the trend of modern teaching. The application of multimedia technology to music classroom teaching in colleges and universities makes music teaching in colleges and universities fully digital. This is reflected in the popularization of computer

music workstation, multimedia projector, portable music system, interactive personal music learning software, online interactive music creation, and other technologies. These technologies can be widely used in all aspects of music teaching to achieve the music education tasks that were difficult to complete due to technical constraints. After the development of artificial intelligence to a certain height, a smart music classroom combining various technologies has emerged. As a platform for music education, the form of music education has changed dramatically, which can maximize students' sensory experience and increase interactivity, thus improving the effect of music teaching and generally improving the quality of music teaching. Multimedia technology can do similar things to a certain extent as a substitute for artificial intelligence.

Many educators have also carried out strategic research on this. In *English Language Teaching*, Meng-yue (2020) directly summarized and analyzed the connotation of the role of modern information technology in teaching. In *International Journal of Computer and Information System*, Fitria (2023) put forward corresponding suggestions on the function and selection of modern teaching media. In terms of information teaching design and practice in the field of information teaching, Jumaboyev (2023) made a detailed discussion on the application of information technology in subject teaching. In *Technologies*, Yu et al. (2023) demonstrated the advantages of using modern multimedia teaching technology to assist music education and teaching with the example of their own practical teaching activities, which not only enriched classroom teaching methods, but also mobilized students' interest in learning music and cultural knowledge. Yu et al. (2023) mentioned that, with the development of China's modern information technology, the use of information technology in music education classrooms to assist students to experience the charm of songs, to cultivate music aesthetic ability and improve music quality will become an effective way to improve the quality of education and teaching.

The birth of the first electronic computer mainly used to solve complex calculations in 1946 marked the beginning of the modern era of computer science and technology. The birth of the first computer teaching system designed by IBM in the United States in 1958 marked the beginning of computer-aided teaching. The emergence of the first microcomputer in 1971 accelerated the rapid development of computer-aided teaching. Multimedia assisted teaching is like a chrysalis waiting to break its cocoon, quietly accumulating material energy for rapid development. Many foreign countries attach great importance to multimedia teaching. The teaching guidelines of the United States state that modern information technology should be integrated with the curriculum as a teaching tool to complete the curriculum that could not be achieved in the past. Under the requirements of the guidelines, the U.S. government established a Teaching Skills Association, which mainly uses seminars and teacher training services to assist school education, so that students can use modern information technology as a teaching tool and classroom integration as much as possible, and then improve school teaching methods to improve teaching effectiveness (Fitria, 2023). In September 1999, the *Future Music Teaching Hauswright Conference* on the integration of computer technology and modern music teaching was held in Florida, USA. The Housewright Manifesto, a plan for the development of American music education from the foresight to 2020, was released, which further accelerated the process of the integration of American music teaching and modern science and technology (Li & Huang, 2023).

In Japan, the application of modern information technology in the classrooms of primary and secondary schools has attracted increasing attention. In various disciplines, the use of information technology allows to deal with, analyze and solve real educational problems, and the overall educational effect is greatly improved (Jumaboyev, 2023). In 2005, the Japanese Education Commission clearly put forward the Educational Informatization Implementation Plan, which stipulated that teachers of various subjects in primary and secondary schools in Japan should use computers and the Internet to teach; this initiative has realized a fundamental reform of teaching methods. Its purpose is to use the teaching methods of modern information technology to change the traditional teaching mode centered on teaching materials, to use innovative methods to arouse

students' interest in learning, to guide them to think actively, and to produce good teaching effects (Meng-yue et al., 2020).

In summary, although the development progress of modern technology in different countries is different, they are aware of the importance of the application of modern technology in teaching, and are actively promoting the application of modern technology in classroom teaching. Education informatization has entered a new stage: On the one hand, the rapid development and wide application of network information technology provide strong support for education informatization; on the other hand, the exploration of educational theory and practice has been deepened and expanded, and the role of teachers has undergone profound changes, that is from the teacher to the learner and researcher.

METHODS

The Concept and Characteristics of Multimedia Teaching

Multimedia technology, also known as multimedia computer technology, is a technology that enables users to interactively convert text, graphics, audio, animation, and video in an interactive way (Bărbuleț, 2023). Multimedia technology, in short, is a system technology for real-time comprehensive processing of various media information such as voice and image (Dhiman, 2023). In this paper, the author refers to multimedia technology as a technical means that completes editing, storage, and display through the combination of two or more different types of information carriers. These information carriers cover a wide range, including text, audio, images, and video. Thus, the author refers to multimedia technology as an information science technology. People often say multimedia technology refers not to the media itself, but to the computer science and technology for information processing and application of the media itself. Moreover, these multimedia technologies are also mainly supported by computer technology (Yu et al., 2023).

Multimedia includes a new educational media formed by a large amount of visual and auditory information. Like traditional teaching methods, it can greatly broaden the learning horizons of teachers and students, and has great attraction to teachers and students (Qıslawbayevna, 2023). Using multimedia to learn can cultivate learners' creativity and imagination, and expand the depth of learning. Besides, students can use listening and seeing to truly feel the learning content, which is incomparable to teachers' language description (Cavanagh & Kiersch, 2023). In the traditional music teaching in colleges and universities, teachers mainly play audio tapes in the classroom. This teaching form is single, and the sound quality is limited. If this form is repeated for a long time, students feel bored and the result is bad.

The multimedia teaching system supports interactive teaching. In the classroom, if the teacher needs it, it can present the real content (e.g., pictures, cartoons, text, language, and music) displayed on the computer screen used by the teacher to the learners, and allow for human-based teaching. Computer interaction reflects the real interactive information content between the information content disseminator and the receiver, and, more importantly, the teaching information can be processed, converted, and decomposed (Kosch et al., 2023). This kind of interactivity enables teachers to avoid the classroom state of "full classroom," and adopt heuristic teaching methods to train students' innovative ability in thinking, by guiding students to think and explore independently. Through the interactivity of multimedia teaching, teachers can repeatedly observe students' learning, timely provide feedback, and design flexible teaching plans adaptively, at any time (Kuchenbuch et al., 2012).

The rational use of excellent multimedia courseware in teaching can enhance learners' motivation, inspire their imagination and creativity, help them improve their academic performance, and improve teaching efficiency more quickly. Multimedia can convert rich and colorful text information into simple and clear text information, graphics, animation, audio, and video, and shorten the time for students to receive information. Students can not only quickly observe the context, cause, and effect of the

learning content, but also obtain more information in a limited time, which improves the teaching quality and effect (Rulismi, 2023).

Music Teaching Algorithm Based on Convolutional Neural Network

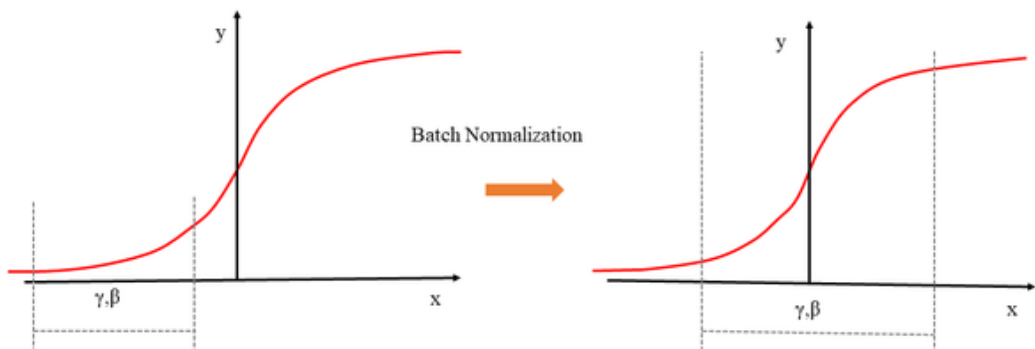
CNNs are a kind of feedforward neural networks with deep structure and convolution computation. They are one of the representative algorithms of deep learning. A CNN has the ability of representation learning and can perform shift invariant classification on input information according to its hierarchical structure. Therefore, it is also called shift invariant artificial neural networks. The basic structure of CNNs includes convolutional layers, pooling layers, and fully connected layers. At present, most of the commonly used CNNs are formed by stacking layers, and different architectures will be generated when different methods or scales are selected for stacking. In the CNN structure, in order to prevent overfitting and reduce training parameters, pooling is often used, that is, the pooling layer is used to reduce the complexity of the model, solve overfitting, and improve the robustness of features, also in order to improve its representation ability. Convolutional layers and fully connected layers often add activation functions, such as Rectified Linear Unit (ReLU) and Exponential Linear Unit (ELU) (Ley-Flores et al., 2022). According to the basic structure rules of CNNs, in this study, the author designed the MR-DCNN model for music teaching. Since the activation function ReLU will cause neurons to “die” when the parameters are poorly initialized or the learning rate is set too large, the neural network model cannot continue to learn effectively, so a batch normalization (BN) layer is introduced after the convolution layer (Dey et al., 2023). The BN is combined into a subunit structure of Conv+BN+ReLU. In addition, in order to better learn the time information, as a feature supplement to the input training, the author added two long short-term memory models before the full connectivity layer of the model, which can better understand the features of different music types and how the segments change with time (al-Ani et al., 2023).

The author’s MR-DCNN model is connected by a BN+ReLU layer between each convolutional layer and a pooling layer. The BN is mainly used to solve the problem of machine learning Independent and Identically Distributed (Suyun & Suying, 2023).

The BN changes the variance and mean by optimizing the network, so that the updated distribution is more in line with the real distribution of the input, and the nonlinear expression ability of the model is improved (Figure 1).

After adding the BN layer, the author’s MR-DCNN model not only improves the performance of the model, but also solves the problems of low training efficiency (Suyun & Suying, 2023). The BN is applied to the music teaching model in this paper without worrying

Figure 1. Schematic diagram of the role of BN



about the Dead ReLU problem in ReLU when a high learning rate is set. It can also use a lower dropout to improve the training speed and is not sensitive to weight initialization (Götz et al., 2023). Therefore, the author's model uses Xavier weights as the initialization method (Useche & Hurtado, 2019).

Model Function Selection

The related functions of the author's MR-DCNN model are divided into activation functions and loss functions. Since the activation function and the loss function play an important role in the model, the author compared different activation functions and loss functions in various aspects to provide a reference for the construction of this model (Bai et al., 2023).

Generally, the activation function of artificial neural network mainly includes Sigmoid, Tanh, ReLU, and ELU. The details are as follows:

1. Sigmoid is mostly used for classification tasks, which can map the input within (0, 1), and is monotonically continuous and optimally stable. However, it has soft saturation, that is, when its value tends to 0 or 1, the gradient is close to 0, which will cause the problem of gradient disappearance, and the output is not zero-centered, resulting in a decrease in the convergence speed, so it is less used in neural networks.
2. Tanh is also known as the hyperbolic tangent function. Compared with the Sigmoid function, the neural network tends to use Tanh. It maps the input in the interval (-1,1), the change-sensitive interval is wider, and the convergence is faster. However, it also has soft saturation and high computational cost, and Tanh is also rarely used in network models.
3. Softmax is often used for multiclassification. In this model, as the activation function of the prediction output layer, the n-dimensional real vector can be mapped to a new n-dimensional real vector between 0 and 1, as Equation 1 shows:

$$f(x)_j = softmax(x)_j = \frac{e^{z_j}}{\sum_{i=1}^k e^{z_i}} \quad (1)$$

Equation 2 shows the corresponding derivative:

$$\frac{\partial f(x)_j}{\partial z_i} = f(x)_i \begin{cases} 1, & i = j \\ 0, & \text{other} \end{cases} - f(x)_i f(x)_j \quad (2)$$

Among them, the denominator of Equation 1 has the function of normalization, which converts the output into probability form.

4. The full name of ReLU is rectified linear unit, which is the mainstream activation function in the current network model. It is a max function, which is equivalent to $f(x) = \max(0, x)$, as Equation 3 shows:

$$f(x) = relu(x) = \begin{cases} x, & x > 0 \\ 0, & \text{other} \end{cases} \quad (3)$$

The derivative of ReLU is relatively simple, as Equation 4 indicates:

$$f'(x) = \text{relu}'(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{other} \end{cases} \quad (4)$$

ReLU has no negative value. When $x < 0$, it has hard saturation. When $x > 0$, its derivative is 1, that is, the gradient does not decay in the positive direction, so the gradient does not disappear (Figure 2). Compared with Sigmoid and Tanh, which require exponential operations, ReLU can quickly obtain the activation value through a simple threshold, and the convergence speed in Stochastic Gradient Descent is also faster.

5. ELU solves some problems of ReLU and has excellent characteristics. The ELU function generally needs to select a value of α , and the common value is between 0.1-0.3.

The ELU function has soft saturation on the side of $x < 0$, and no saturation on the side of $x > 0$, that is, the ELU has a certain antiinterference ability. Its output can take a negative value, and the mean value is close to zero, so the convergence speed is faster, but ELU still has the problem of exponential operation, the calculation amount is large, and the gradient explosion cannot be avoided. The neural network does not learn the value of α , so that the value depends on experience.

After the above analysis and comparison, the MR-DCNN model the author proposes in this paper selects ReLU as its training activation function, and selects Softmax as its output layer activation function.

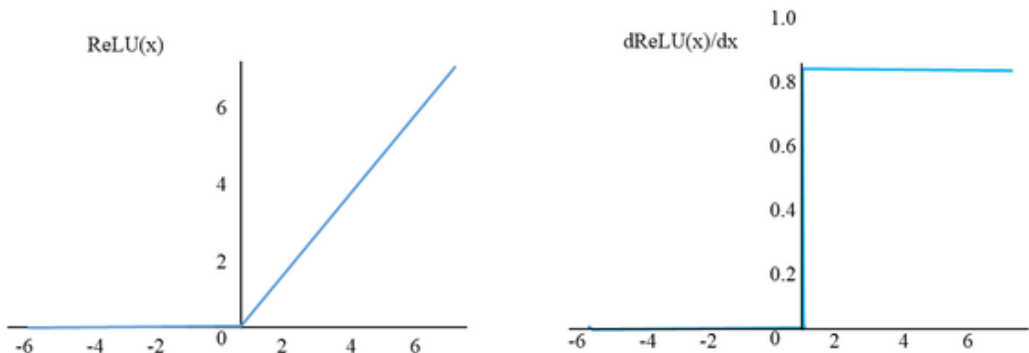
Choice of Loss Function

The loss function is the objective function for evaluating and optimizing the effect of the network model in deep learning. It guides the network parameter learning by back-propagating the error between the model prediction result and the real result. In general, it is a nonnegative function, and the smaller the function value, the better the model robustness. In application, the loss function is solved and evaluated by minimizing the loss function. For example, it is used for parameter estimation of models in statistics and machine learning, for risk management and decision-making in macroeconomics, and for optimal control theory in control theory.

Mean Square Error Loss

The mean square error (MSE) loss is the expected value of the square of the difference between the predicted value and the true value; the smaller the value, the higher the model accuracy (Equation 5):

Figure 2. ReLU function and its derivative diagram



$$L(Y, \hat{Y}) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (5)$$

Mean Absolute Error Loss

The mean absolute error (MAE) loss can better reflect the actual situation of prediction error (Equation 6):

$$L(Y, \hat{Y}) = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

Cross Entropy Loss

The cross entropy (CE) loss is generally used to evaluate the difference between the probability distribution obtained by the current training and the real distribution. The smaller the value, the closer the distribution of the two. Commonly used loss functions include binary_crossentropy (BCE), categorical_crossentropy (CCE), and sparse_categorical_crossentropy (SCE).

Binary Cross Entropy

The BCE is commonly used in binary classification tasks, and it is usually necessary to add a sigmoid function to the last layer of the network (Equation 7):

$$BCE(x) = -\sum_{i=1}^2 y_i \log f_i(x) + (1 - y_i) \log(1 - f_i(x)) \quad (7)$$

Categorical Cross Entropy

The CCE is commonly used in multiclassification tasks where the activation function of the output layer is the Softmax function. When CE performs single-label multiclassification, the y value needs to be one-hot encoded (Equation 8).

$$CE(x) = -\sum_{i=1}^C y_i \log f_i(x) \quad (8)$$

Sparse Categorical Cross Entropy

The SCE is similar to CCE, except that the way of calculating the value of y is different. In the SCE, y is the original integer form, such as 2, 0, 1. In the above formula, x represents the input, C represents the number of categories, and y_i represents the true label corresponding to the i th category.

In order to better cooperate with the whole model of this experiment, after analysis and comparison, the CCE is selected as the loss function to measure the effect of network training and prediction.

Since the author's MR-DCNN model proposed belongs to the category of CNN models, its time complexity analysis is actually the analysis of each layer of the network, that is, the total time complexity of the MR-DCNN model is the sum of all convolutional layers. It is summation, that is, multiplication within layers and accumulation between layers. Next, at the theoretical level, the time complexity of a single convolutional layer, is used as an entry point to analyze its overall time complexity:

$$T \sim O(M^2 \cdot K^2 \cdot C_{in} \cdot C_{out}) \quad (9)$$

In Equation 9, the output feature map M is determined by the input matrix X , convolution kernel K , padding P , and stride S , as Equation 10 shows:

$$M = \frac{x - K + 2 * P}{s} + 1 \quad (10)$$

$$T \sim O\left(\sum_{l=1}^N M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l\right) \quad (11)$$

In Equation 11, N denotes the depth of CNN, l denotes the l th convolutional layer, C_l denotes the number of output channels of the l th convolutional layer, and C_{out} , C_l-1 denotes the number of input channels of the l th convolutional layer C_{in} .

Monte Carlo Method

The Monte Carlo method is a model-free calculation method based on random numbers. In general, the Monte Carlo method can be regarded as a statistical simulation method that equates the average return value with the value, and the more samples, the closer to the optimal solution. The basic idea of Monte Carlo method is: In order to solve the problem, firstly, a probability model or stochastic process is established so that its parameters or numerical characteristics are equal to the solution of the problem. Then, these parameters or numerical characteristics are calculated through the observation or sampling test of the model or process, and finally the approximate values are given. The accuracy of the solution is expressed by the standard error of the estimated value. The main theoretical basis of the Monte Carlo method is the theory of probability and statistics, and the main means are random sampling and statistical experiments.

Compared with the dynamic programming method, the Monte Carlo method is a model-free reinforcement learning method, that is, it does not need to provide a complete model environment and only learns through multiple sampling or heuristics to form experience. However, it usually requires many segments, and each segment must be a complete sample. When a segment is long, and $at \neq \pi(st)$, it needs to be truncated, so the learning rate is not high and may cause problems such as large variance.

The Monte Carlo method computes the average of the future decaying cumulative returns G as the expectation of the state value $v(s)$:

$$G = \frac{1}{T} \sum_{t=1}^T G_t \quad (12)$$

Assuming that the empirical trajectories under policy π are $S_1, A_1, R_2, S_2, A_2, R_3, \dots, S_{T-1}, A_{T-1}, R_T, S_T \sim \pi$. Equations 13 and 14, respectively, show the value function update:

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots + \gamma^{T-1} R_T \quad (13)$$

$$V(S_t) \leftarrow V(S_t) + \alpha (G_t - V(S_t)) \quad (14)$$

Given the strategy π , as the number of samples of the complete state sequence increases, that is, when the sampling tends to infinity, the obtained state value $v\pi(s)$ also approximates the real state value. Equation 15 shows the value function:

$$v_{\pi}(s) = \mathbb{E}_{\pi} [G_t | S_t = s] \quad (15)$$

The Monte Carlo method has the following characteristics: 1) It can directly learn the state sequence or experience trajectory from the environment, that is, the Monte Carlo method only needs experience; 2) it can use the experienced state sequence for off-line learning; 3) in the absence of the task of the model does not depend on state transition probabilities.

RESULT ANALYSIS AND DISCUSSION

Experimental Results and Analysis

According to the experimental plan, the author compared the model proposed in this paper with the comparison model on three datasets for evaluation indicators such as MAE, RMSE, and Top@K accuracy, as well as the parameters of learning rate, activation function, and feature generation.

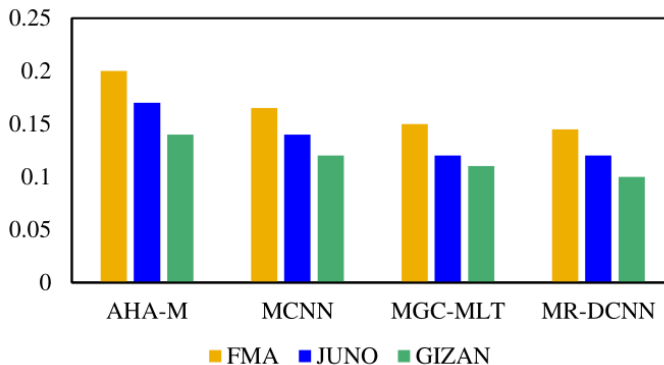
In order to comprehensively and objectively prove that the performance of the model in this paper is better than that of the comparison model, the researcher conducted the experiments on the three datasets of FMA, JUNO, and GTZAN to compare the MAE, RMSE, and Top@K prediction accuracy with the comparison model, respectively. The specific experimental process and results were as follows. The public audio dataset GTZAN contains 10 different styles of music, each of which contains 100 audio files, and each of which is 30 seconds. The JUNO dataset includes feature analysis and metadata of one million songs. This dataset does not contain audio, but only features. FMA dataset is a dataset for music analysis, which consists of full length and HQ audio, precalculated features, and audio track and user level metadata.

First, the researcher compared the MAE results of the MR-DCNN model proposed and the AHA-M, MCNN and MGC-MLT models and analyzed them on the three datasets of FMA, JUNO, and GTZAN. Figure 3 illustrates the experimental comparisons.

As Figure 3 shows, the performance of the MR-DCNN model is significantly better than that of other comparison models, that is, the MAE results of the MR-DCNN model on the three datasets are 0.146, 0.113, and 0.099, respectively, which are smaller than those of other comparison models.

Specifically, the MAE result values of the MR-DCNN model compared with the AHA-M model on the FMA dataset are 0.146 and 0.204, respectively; the MAE result values on the JUNO dataset are 0.113 and 0.165, respectively; on the GTZAN dataset, the MAE results on the above are 0.009 and 0.153, respectively, indicating that the deep neural network model is significantly better than the general neural network model. Compared with the MCNN and MGC-MLT models on the FMA

Figure 3. MAE comparison chart of three datasets



dataset, the MR-DCNN model improved MAE by 12.6% and 2.67%; on the JUNO dataset, the model performance improved by 15.0% and 1.74%, respectively. The model performance on the GTZAN dataset improved by 18.2% and 9.17%, respectively, indicating that the MR-DCNN model has better feature extraction capabilities and that the combined representation ability of DCNN and long short-term memory will be greatly improved. For the author's MR-DCNN model, the performance on the GTZAN dataset is better than that on the FMA dataset and the JUNO dataset. The possible reason is that the distribution of audio categories in GTZAN is relatively balanced, so that, in the same category, the existence of a large number of similar audios (i.e., there will be no cross features of the same category anomalies) is more conducive to the training and learning of the model in this paper. This may also be a potential factor for the poor performance of the author's MR-DCNN model on the FMA dataset.

Impact of Learning Rate

The MR-DCNN model involves several parameters, and different learning rates are used on the GTZAN dataset to conduct experiments on the impact of the MR-DCNN model. Among them, the learning rate parameter value is set from $1.00\text{E-}06$ to $1.00\text{E-}02$, and Figure 4 shows the specific experimental results.

As Figure 4 evidences, the learning rate rises sharply in the interval $[1.00\text{E-}06, 5.00\text{E-}05]$, and drops sharply in the interval $[5.00\text{E-}03, 1.00\text{E-}02]$, indicating that the learning rate is too small and too large, and the model does not converge and learn well, and the model is relatively stable in the middle interval. For Top@1, the accuracy shows large fluctuations with the change of the learning rate, but the changes in Top@2 and Top@3 accuracy are relatively stable, indicating that the author's model has a good ability to express large K values. The comparison of the trends of MAE and RMSE on the learning rate highlight that the model performs best when the learning rate is $1.00\text{E-}04$.

Commonly used activation functions are Sigmoid, Tanh, ReLU, and ELU. The neural network model can perform nonlinear operations on neurons. At this time, the neural network model can be regarded as a corresponding complex nonlinear function to solve complex problems. Therefore, different activation functions will have different effects on model performance. In this experiment, the author chose to use evaluation indicators such as MAE, RMS, and Top@K on the GTZAN dataset to analyze the influence of different activation functions on the MR-DCNN model proposed.

Figure 4. The effect of learning rate on the model

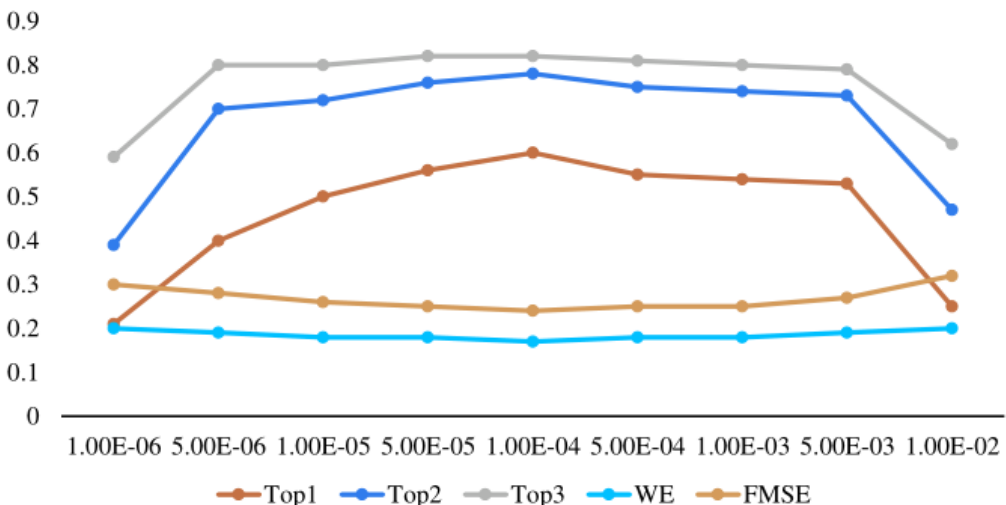


Figure 5 and Figure 6 show that, when ReLU is selected as the activation function of the MR-DCNN model, the accuracy of the model is higher than that of other activation functions, and the MAE and RMSE are lower than those of other activation functions. The result value shows that choosing ReLU can get the best effect.

Parameter Sensitivity Experiment

The MR-DCNN model proposed in this paper includes parameters such as the learning rate. On the R3-Yahoo dataset, different learning rate parameter values are taken to conduct experiments on the impact of the MR-DCNN model. Among them, the learning rate parameter value is set from 1e-04 to 5e-02, and Figure 7 shows the specific experimental results.

Figure 7 evidences that the learning rate increases sharply when the hit rate is in the interval [1.00E-04, 5.00E-04], and decreases sharply in the interval [1.00E-02, 5.00E-02], indicating

Figure 5. Comparison of MAE and RMSE for different activation functions

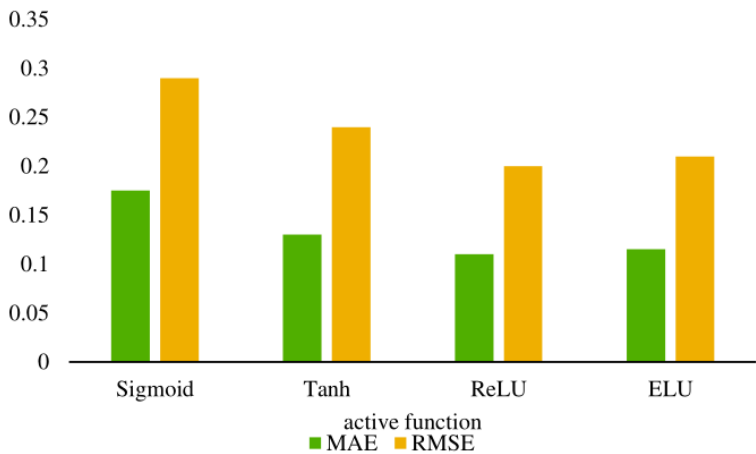


Figure 6. Comparison of the accuracy of different activation functions

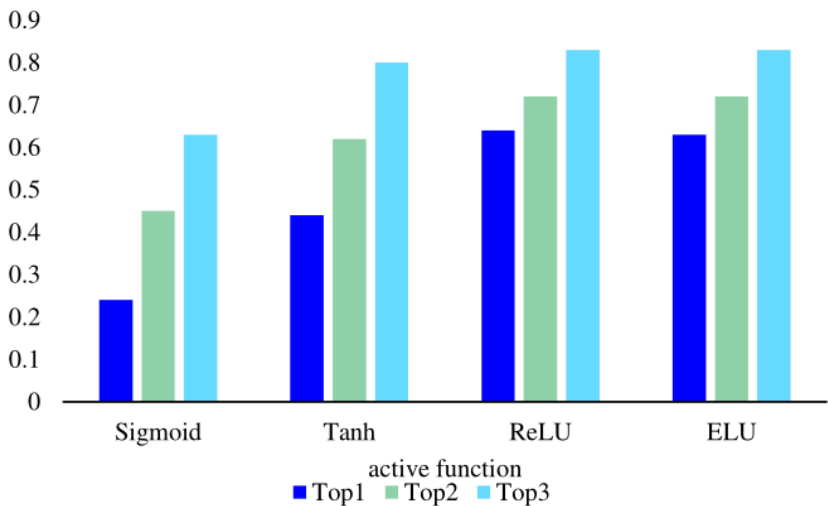
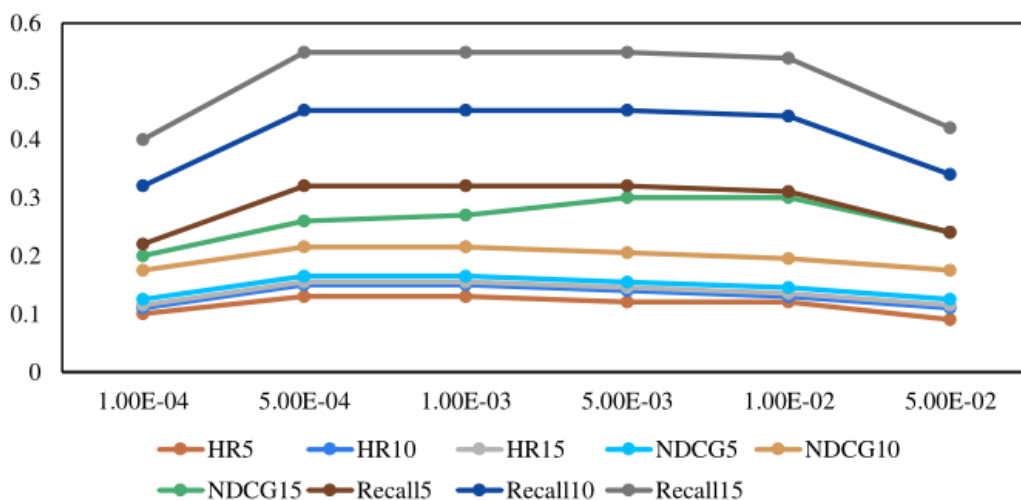


Figure 7. The effect of learning rate on the model



that the learning rate is too small or too high. If it is large, the model will not be able to train or converge well, and the model is relatively stable in the interval (5.00E-04, 1.00E-02]. For the recall rate and Normalized Discounted Cumulative Gain, there is no large fluctuation with the change of the learning rate. When the learning rate is 1.00E-03, the highest value of the corresponding index is reached, indicating that the model can reflect better teaching performance at this time. According to the basic structure rules of CNNs, the author designed a MR-DCNN model for music teaching and applied it to speech recognition. This model can better identify the features of different music types.

CONCLUSION

Through collecting and consulting a large number of documents related to multimedia music teaching, in this paper the author systematically analyzed the necessity and importance of the application of multimedia technology in music teaching in middle schools in China, analyzed the advantages of multimedia teaching, and expounded the application of multimedia teaching technology in various music teaching links in music teaching. As an auxiliary means of music classroom teaching, multimedia application technology provides an effective means for the reform of music education. Music appreciation is one of the components of music education. If in the teaching of music appreciation instructors can skilfully use multimedia for teaching, they will achieve many positive results. Singing teaching is the primary task of music teaching. In multimedia teaching, students are not only taught how to sing songs, but also how to sing songs well and sing out their understanding of songs. In the teaching of basic knowledge of music theory, multimedia is used to help abstract into images. Expanding teaching is the extension of music activities and the comprehensive embodiment of aesthetic characteristics. The application of multimedia in music teaching has greatly expanded the space and capacity of music teaching, and greatly enriched the teaching methods and resources. Of course, multimedia technology is not omnipotent, and teachers cannot blindly exaggerate its role. The production and use of multimedia courseware for music appreciation cannot simply replace teachers' teaching with multimedia. They must follow the principle of subsidiarity, pay attention to the leading role of teachers, and regulate the two-way interaction with students.

DATA AVAILABILITY

The Figures the author used to support the findings of this study are included in the paper.

CONFLICTS OF INTEREST

The author declares that he has no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGEMENT

The author would like to show sincere thanks to those technicians who have contributed to this research.

REFERENCE

- al-Ani, O., Das, S., & Wu, H. (2023). Imitation Learning with Deep Attentive Tabular Neural Networks for Environmental Prediction and Control in Smart Home. *Energies*, 16(13), 5091. doi:10.3390/en16135091
- Bai, J., Rabczuk, T., Gupta, A., Alzubaidi, L., & Gu, Y. (2023). A physics-informed neural network technique based on a modified loss function for computational 2D and 3D solid mechanics. *Computational Mechanics*, 71(3), 543–562. doi:10.1007/s00466-022-02252-0
- Bărbuleț, G. (2023). The use of multimedia in language teaching. *Swedish Journal of Romanian Studies*, 6(1), 191–201. doi:10.35824/sjrs.v6i1.24967
- Berthaut, F. (2020). 3D interaction techniques for musical expression. *Journal of New Music Research*, 49(1), 60–72. doi:10.1080/09298215.2019.1706584
- Cavanagh, T. M., & Kiersch, C. (2023). Using commonly-available technologies to create online multimedia lessons through the application of the Cognitive Theory of Multimedia Learning. *Educational Technology Research and Development*, 71(3), 1033–1053. doi:10.1007/s11423-022-10181-1 PMID:36570341
- da Silva-Santos, C. H., Goulart, P. R., Bertelli, F., Garcia, A., & Cheung, N. (2023). Nature inspired algorithms for the solution of inverse heat transfer problems applied to distinct unsteady heat flux orientations in cylindrical castings. *Journal of Intelligent Manufacturing*, 34(5), 2407–2430. doi:10.1007/s10845-022-01935-y
- Dey, K., Kalita, K., & Chakraborty, S. (2023). Prediction performance analysis of neural network models for an electrical discharge turning process. [IJDeM]. *International Journal on Interactive Design and Manufacturing*, 17(2), 827–845. doi:10.1007/s12008-022-01003-y
- Dhiman, D. B. (2023). Key Issues and New Challenges in New Media Technology in 2023: A Critical Review. *Journal of Media & Management*, 5(1), 1–4.
- Fitria, T. N. (2023). Augmented Reality (AR) and Virtual Reality (VR) Technology in Education: Media of Teaching and Learning: A Review. [IJCIS]. *International Journal of Computer and Information System*, 4(1), 14–25.
- Götz, P., Tuna, C., Walther, A., & Habets, E. A. (2023). Online reverberation time and clarity estimation in dynamic acoustic conditions. *The Journal of the Acoustical Society of America*, 153(6), 3532–3542. doi:10.1121/10.0019804 PMID:37387542
- Jumaboyev, R. M. (2023). Methodology for Teaching Information Culture to Students in the System of Higher Education. *Академические исследования в современной науке*, 2(8), 126–134.
- Kosch, T., Welsch, R., Chuang, L., & Schmidt, A. (2023). The Placebo Effect of Artificial Intelligence in Human–Computer Interaction. *ACM Transactions on Computer-Human Interaction*, 29(6), 1–32. doi:10.1145/3529225
- Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., & Pantev, C. (2012). Electromagnetic correlates of musical expertise in processing of tone patterns. *PLoS One*, 7(1), e30171. doi:10.1371/journal.pone.0030171 PMID:22279568
- Ley-Flores, J., Alshami, E., Singh, A., Bevilacqua, F., Bianchi-Berthouze, N., Deroy, O., & Tajadura-Jiménez, A. (2022). Effects of pitch and musical sounds on body-representations when moving with sound. *Scientific Reports*, 12(1), 1–20. doi:10.1038/s41598-022-06210-x PMID:35177677
- Li, L. Y., & Huang, W. L. (2023). Effects of Undergraduate Student Reviewers' Ability on Comments Provided, Reviewing Behavior, and Performance in an Online Video Peer Assessment Activity. *Journal of Educational Technology & Society*, 26(2), 76–93.
- Meng-yue, C., Dan, L., & Jun, W. (2020). A Study of College English Culture Intelligence-Aided Teaching System and Teaching Pattern. *English Language Teaching*, 13(3), 77–83. doi:10.5539/elt.v13n3p77
- Nitin, R., Gustavson, D. E., Aaron, A. S., Boorom, O. A., Bush, C. T., Wiens, N., Vaughan, C., Persici, V., Blain, S. D., Soman, U., Hambrick, D. Z., Camarata, S. M., McAuley, J. D., & Gordon, R. L. (2023). Exploring individual differences in musical rhythm and grammar skills in school-aged children with typically developing language. *Scientific Reports*, 13(1), 2201. doi:10.1038/s41598-022-21902-0 PMID:36750727

Qislawbayevna, O. A. (2023). THE ROLE OF MUSIC IN THE EVOLUTION OF MANKIND. *INNOVATIVE DEVELOPMENTS AND RESEARCH IN EDUCATION*, 2(20), 48–51.

Rulismi, D. (2023). CIPp Model In Utilizing Interactive Multimedia-Based Teaching Materials (Evaluation Study On Pertiwi 1). *Indonesian Journal of Basic Education*, 5(3), 231–235.

Suyun, W., & Suying, Z. (2023). Application of big data classification effects based on neural network in video English course and relevant optimization suggestions. *Soft Computing*, 27(11), 7615–7625. doi:10.1007/s00500-023-08123-x PMID:37192890

Suyun, W., & Suying, Z. (2023). Application of big data classification effects based on neural network in video English course and relevant optimization suggestions. *Soft Computing*, 27(11), 7615–7625. doi:10.1007/s00500-023-08123-x PMID:37192890

Useche, J., & Hurtado, R. (2019). Melodies as maximally disordered systems under macroscopic constraints with musical meaning. *Entropy (Basel, Switzerland)*, 21(5), 532. doi:10.3390/e21050532 PMID:33267246

Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and applications of artificial intelligence in music education. *Technologies*, 11(2), 42. doi:10.3390/technologies11020042

Yu, X., Ma, N., Zheng, L., Wang, L., & Wang, K. (2023). Developments and applications of artificial intelligence in music education. *Technologies*, 11(2), 42. doi:10.3390/technologies11020042

Zereu, P. R. S. (2023). The Musical Language and Interpretation of Poulenc's Concerto in D Minor for Two Pianos and Orchestra. *Rangsit Music Journal*, 18(1), 155–169.

Dalei Feng graduated from the Central Conservatory of Music and was admitted to Ohio State University for a master's degree with a full scholarship. He is now a lecturer of vocal music in the Conservatory of Music of Zhaoqing College, Guangdong Province.