PMG-Net: Electronic Music Genre Classification using Deep Neural Networks

Yuemei Tang
The Department of Music, Science and Technology College Gannan Normal University
Ganzhou, 341000, China

***Abstract*—With the rapid development of electronic music industry, how to establish a set of electronic music genre automatic classification technology has also become an urgent problem. This paper utilized neural network (NN) technology to classify electronic music genres. The basic idea of the research was to establish a deep neural network (DNN) based classification model to analyze the audio signal processing and classification feature extraction of electronic music. In this paper, 2700 different types of electronic music were selected as experimental data from the publicly available dataset of W website, and substituted into the convolutional neural network (CNN) model, PMG-Net electronic music genre classification model and traditional classification model for comparison. The results showed that the PMG-Net model had the best classification performance and the highest recognition accuracy. The classification error of PMG-Net electronic music genre classification model in each round of training was smaller than the other two classification models, and the fluctuation was small. The speed of music signal processing in each round and the feature extraction of audio samples of PMG-Net electronic music genre classification model were faster than the traditional classification model and CNN model. It can be seen that using the PMG-Net electronic music genre classification model customized based on DNN for automatic classification of electronic music genres has a better classification effect, and can achieve the goal of efficiently completing the classification in massive data.**

***Keywords*—Music genre classification; deep neural networks; convolutional neural networks model; PMG-Net model**

I. INTRODUCTION

Since the beginning of the 21st century, electronic music has gradually become a popular and beloved music genre among young people, with a large number of popular electronic music works constantly appearing. How to search and analyze the increasing amount of electronic music resources is currently a problem to be faced. Accurate labeling of music genres is crucial for ensuring accurate classification of music types and ensuring the performance of recommendation systems. Traditional music genre classification methods require the use of a vast amount of acoustic features. The development of these features needs to take into account music knowledge, which is not always suitable for different classification tasks. Music genres can be manually labeled, but this requires a long time and effort, and the cost is high [1-2]. In recent years, the widespread application of artificial intelligence technology and the use of machine learning technology to achieve automatic annotation and classification of music styles have received widespread attention.

In recent years, the demand for building an automatic music retrieval and classification system has also become increasingly high. Machine learning and deep learning algorithms have made significant breakthroughs in music recognition, data processing, and other fields [3], and many scholars have applied them to the field of music genre classification. Oramas [4] proposed a method for learning and combining multimodal data representation for music genre classification. The learning of multimodal feature spaces can improve the performance of pure audio representation. Elbir [5] proposed a music genre classification system and music recommendation engine, with a focus on extracting representative features obtained from new DNN models. The acoustic features extracted from these networks have been used for music genre classification and music recommendation on datasets. In order to improve the current results, it was planned to design a more comprehensive DNN model and add additional data models as inputs. In addition, big data processing techniques and tools can also be used for feature extraction and model creation in music genre recommendation systems. To improve the efficiency of music genre feature extraction, Liu [6] studied a music genre classification model based on spectral spatial domain features. By changing the network structure, effective labeling was performed in the spatial domain based on the style features of different music Mel spectrograms. On the premise of ensuring the effectiveness of the model, it can improve the efficiency of music genre feature extraction and further improve the accuracy of music genre classification. Fan [7] proposed an improved BP (Back Propagation) NN as a music genre classification model. Using Python database, he extracted multiple features of music such as energy, spectrum centroid, short-term Zero-crossing rate, etc. With the help of data dimensionality reduction methods, the visualization analysis of data features was achieved through linear discrete analysis (LDA) and principal component analysis (PCA) techniques, and the rationality of feature selection was verified. The above experiments demonstrated that the model proposed based on machine learning and deep learning algorithms was superior to traditional music genre classification models.

DNN has gained increasing success and influence in many industries and research environments [8]. Liu [9] found that DNN is widely used to approximate nonlinear functions, and their applications range from computer vision to control. Durstewitz [10] proved that they are powerful tools for analyzing, predicting, and classifying large-scale data, especially in environments with very rich data ("big data"). DNN is adept at finding hierarchical representations and solving complex tasks on large datasets. Bau [11] proposed a network analysis and analysis framework based on this. A

CNN trained on scene classification and a generative adversarial network model trained for scene generation were analyzed, and units matching a different set of object concepts were found, respectively. Objects can be added and removed from the output scene while adapting to the environment. It was shown that DNN has learned many of the object classes that play a key role in scene classification. Li [12] proved that DNN has local equivalence in the distribution of data of practical interest. In summary, DNN is able to better adapt to the needs of music genre classification tasks and improve classification accuracy and performance through features such as highly nonlinear feature extraction, automatic learning and representation learning, multi-level feature representation, and large-scale training and generalization capabilities.

Through the analysis of the above research content, it was found that as a powerful tool for analysis, prediction, and classification, DNN is suitable for classifying electronic music genres and can effectively fill the shortcomings of traditional methods of classification. Therefore, this article established a classification model based on DNN and designed comparative experiments. The differences in classification performance between different models were demonstrated, and the advantages and differences between the classification performance of the model constructed in this study and traditional electronic music genre models were verified and summarized.

II. Basic Knowledge of Electronic Music Genre Classification

Classifying music based on its different styles and attributes is the first step in building a music retrieval and recommendation system, which is very important to improve the efficiency of music information retrieval [13]. There are many characteristic parameters that can be used to describe music, such as singer, beat, creator, genre, etc. Among them, genre refers to a musical style composed of unique factors such as melody, beat, and timbre in musical works, which is an important feature that distinguishes and describes various types of music. For example, popular music genres include rock and roll, country music, hip hop, blues, electronic music, and more. Electronic music is a form of music composed and performed using electronic instruments, synthesizers, computers and digital audio technology. Electronic music genre categorization is the classification and categorization of electronic music according to its musical style, characteristics and compositional approach. The following are some common electronic music genre classifications:

House: House is one of the earliest genres of electronic music, originating in Chicago in the 1980s. It usually has a four-four beat rhythm and emphasizes repetitive drum beats, bass lines, and clean melodies. House music is usually dance-oriented, giving it an upbeat, dynamic feel.

Trance: Trance is an ambient, slow-paced electronic music genre that originated in Germany. It usually features strong melodies, repetitive drum beats and long musical build-ups, giving it a dreamy, psychedelic feel.

Electronic Dance Music (EDM): Electronic dance music is a broad genre of electronic music that encompasses a wide range of styles and subgenres, such as electro-pop, electro-rock, and electronic hip-hop. It usually features strong rhythms, repetitive drum beats and diverse musical elements for dance and party settings.

Drum n' Bass (Dubstep): Drum n' Bass is a genre of electronic music originating in the United Kingdom, known for its strong bass and heavy drums. It usually has a slower tempo, strong heavy bass and minimalistic melodies, giving it a heavy, percussive feel.

Ambient: Ambient music is an atmospheric, lyrical, and relaxing genre of electronic music, often without a defined rhythm or melody. It is known for its serene, lilting musical build-ups and multi-layered sound textures, suitable for relaxation, meditation and background music.

The genre of electronic music has a variety of styles: Electropop, Indie Electronica, Folktronica, Dubstep, Trip-Hop, Ambient, House, Techno, Trance, Disco, Ambient House, Deep House, Electro, Electro-Disco, pulse Glitch, and more.

Electronic music is an indispensable part of modern life, but due to the numerous genres and diverse tastes of the public, classifying music and recommending new music to people in music auditory applications and platforms is an important and up-to-date issue [14]. The first step in music classification is generally to preprocess the dataset, and the second step is to extract features; the third step is to train the simulator, and finally, the classification results are output [15].

III. Classification Methods and Models for Electronic Music Genres

A. DNN Model

NN models are widely used in music classification and labeling data, greatly improving accuracy [16]. DNN is a technology in the field of machine learning [17]. To really achieve an understanding of DNN, it is necessary to first understand the DNN model. The DNN model is expanded on the basis of the perceptron model. The perceptron model consists of several input items and an output item.

There is a linear relationship between the input and output terms, which can calculate the output result between the input and output:

$$Z = \sum_{i=1}^{n} w_i x_i + a \tag{1}$$

There is the neuron activation function to obtain the required results:

$$\operatorname{sign}(Z) = \begin{cases} 1 & Z > 0 \\ -1 & Z \le 0 \end{cases} \tag{2}$$

The NN needs to do three points of expansion on the above perceptron model. (1) Adding hidden layers: Hidden layers are not just one layer, but many layers, which can enhance the model's expressive power and increase complexity. (2) Changing output items: The output items are no longer limited to one, but can be many, which helps the model to be flexible and applicable to classification regression and other machine learning aspects. (3) Extended activation function: the neuron activation function of the above perceptron model is simple but has limited execution ability, so the NN would use other different activation function according to the situation, thus further strengthening the expression ability of the NN. The formula for the sigmoid function in logistic regression is:

$$f(Z)=\frac{1}{1+a^{-Z}} \tag{3}$$

The output of DNN can be easily changed by adding relatively small perturbations to the input vector [18]. DNN is divided according to the position of different layers. The specific situation is shown in Fig. 1.

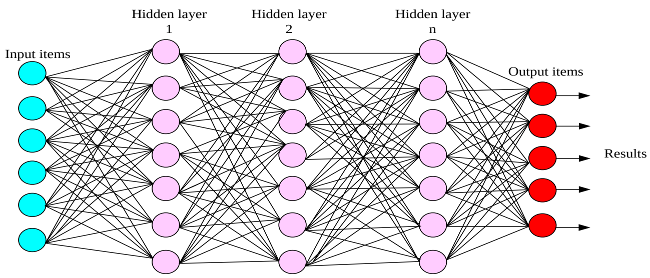

Fig. 1. Basic structure diagram of DNN.

B. CNN Model

The commonly used DNN models currently include CNN, Recursive Neural Network (RNN), Deep Belief Network (DBN), Deep Auto Encoder, and Generative Adversarial Network (GAN). Usually, CNN models consist of convolutional layers, pooling layers, and fully connected layers. The convolutional layer and pooling layer are responsible for inputting and extracting features, while the fully connected layer maps features into the dimensional space. Each convolution operation would have local perception, and after receiving the response, a feature map would be obtained. With a feature map, a parameter can be shared.

1) Convolutional layer: If the convolutional kernel is treated as a weight matrix, it would move according to the designed step size, and the data corresponding to the output position would be weighted and summed to become the output value in the feature map. The calculation formula is:

$$f_{o,l}=g(\sum_{n=0}^{F_{x-1}}\sum_{m=0}^{F_{G-1}}w_{m,n}X_o+n,l+m+a \tag{4}$$

Among them, $f_{o,l}$ refers to the values of row *o* and column *l* in the feature map; *w* is the weight matrix; *x* is the input matrix; *a* is the bias of convolution; *g* is the activation function; F_X and F_G are the width and height of the corresponding convolutional kernel.

2) Pooling layer: The pooling layer includes down sampling of feature maps and dimensionality reduction (reducing complexity, computational complexity, etc.) to expand the perception field and achieve invariance. Similar to convolutional layers, during the pooling process, it is necessary to set the size and step size of the pooling area and aggregate the values.

3) Fully connected layer: After the final layer of convolution or pooling, all feature maps are pieced together into a global feature formed by vectors, and then classified and judged with other probability vectors mapped by the fully connected layer.

C. PMG-Net Electronic Music Genre Classification Model

PMG-Net originated from the classification of Persian music genres. Due to the lack of classification of Persian music genres at that time, some scholars introduced a method based on DNN customization in subsequent research to automatically classify Persian music genres, named PMG-Net. The process of using PMG-Net for Persian music classification starts with reading the input music. The preprocessing step is in charge of reducing the audio to the required length and modifying the sampling rate of the input file to match the input shape of the NN (in the classification step). Next, a spectrogram of the music is created. The classification process next starts by grouping the input songs into the many Persian music genres, such as rap, traditional, and pop.

Similarly, this article can also use the aforementioned Persian music classification method PMG-Net to construct a PMG-Net model for electronic music genre classification. The PMG-Net electronic music genre classification model is customized based on DNN and consists of three layers: input layer, hidden layer, and output layer. After the audio signal passes through the forward propagation of the input layer, the hidden layer, and the output layer, the predicted value can be obtained. Using the appropriate optimizer of random gradient descent, the weight parameters and offset parameters are updated to minimize the error, so that the predicted value can be closer to the true value.

The basic logic of the PMG-Net electronic music genre classification model is to form a linear operation with the weight and offset term, and then act on the Siqmoid activation function to obtain the value of the next neuron connected to this neuron. Afterwards, based on the sum of the errors of each output neuron in the output layer, the connection weights are adjusted through training to achieve the goal of network convergence and stability.

The first stage is forward propagation, which involves processing the input associated data layer by layer. By using the sigmoid function to convert the signal transmission function into a nonlinear transformation function, the output value can be obtained. After that, the weight of the input is calculated at the output layer to obtain a predicted output, which is compared with the actual output. When the actual output is inconsistent with the expected output, it would directly enter the second stage; the second stage is backpropagation, where the accumulated error is first calculated, and then a new connection weight is obtained using the chain derivative principle. This allows for cyclic operations to minimize the error signal.

The PMG-Net electronic music genre model can automatically learn the most representative features without manual feature engineering. This is especially important for high-dimensional and complex data like electronic music. Compared to traditional feature extraction methods, the PMG-Net electronic music genre model can learn and represent rich features from raw audio data, better capturing the differences between genres. Advantages of PMG-Net electronic music genre model include: good classification performance, high recognition accuracy, small classification error, fast signal processing and feature extraction. Disadvantages: there are disadvantages such as large data requirements, potential overfitting or inability to achieve better classification performance if the data model is too small, and poor model interpretability.

IV. Electronic Music Classification Feature Extraction and Evaluation

A. Audio Signal Processing

Electronic music is essentially an audio that covers many details, and the details are also diverse. What impresses people is the main melody composed of some details, and the audio clips containing the main melody have recognizable characteristics. What most people choose to forget is the noise caused by some details. Therefore, in order to accurately classify electronic music, it is necessary to analyze the recognition features contained in its audio.

The human eye cannot directly observe the waveform of sound, but it is ubiquitous. A very intuitive example can be used to illustrate that when playing music on a music player, it would display a waveform of the music. Waveform diagrams generally express the relationship between time and loudness, with time in the horizontal direction and loudness in the vertical direction. An example waveform during music playback is shown in Fig. 2.

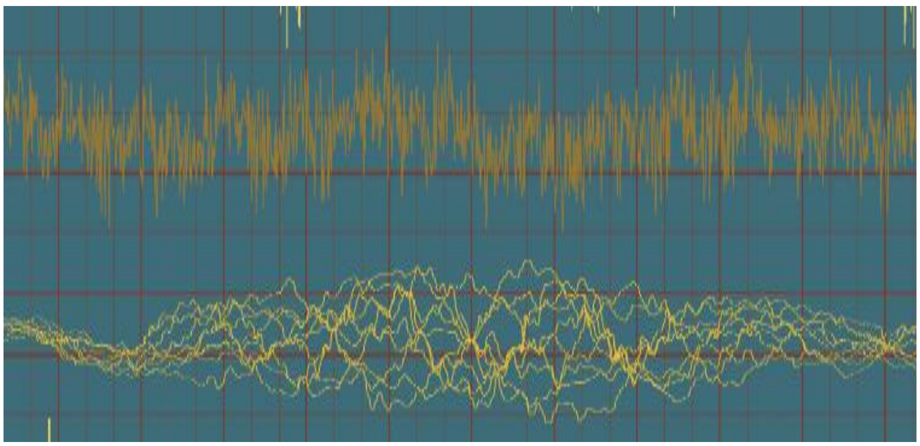

Fig. 2. Example of waveform during music playback.

Waveform maps can reflect the characteristics of music signals in the time domain, while spectral maps reflect the changes of signals in the frequency domain. A typical spectrum diagram is shown in Fig. 3.

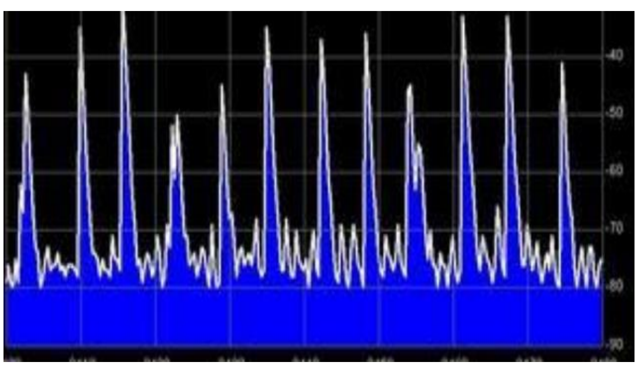

Fig. 3. Example of common spectrum diagrams.

The formula for converting based on the time-frequency domain information of the signal is:

$$\tilde{f}(X)=\int_{-\infty}^{\infty} f(a)e^{-2\Pi ia} \tag{5}$$

Among them, *X* represents any real number and represents frequency; the independent variable *a* refers to time. $\tilde{f}(X)$ is the spectrum of the corresponding signal, which is the Fourier function of the original function f(x).

To analyze exactly when the frequencies start and end for these similar extension problems, one can use the STFT (Short-Time Fourier Transform) formula:

$$X(s,f)=\int_{-\infty}^{+\infty} w(s-\tau)x(\tau)e^{-2\Pi f\tau}ds \tag{6}$$

W(s) is the window function, and $X(s,f)$ is the Fourier transform of $w(s-\tau)x(\tau)$ As soon as *s* changes, the window function shifts on the time axis. After $w(s-\tau)x(\tau)$ operation, only a part of the signal intercepted by the window function is retained as the subsequent Fourier transform to obtain a complex function, which refers to the size and phase of the signal after changing according to time and frequency.

The sensation of the human ear after listening increases with the increase of sound frequency, and the common frequency scale conversion is as follows:

$$f_{mel}=2600*\log_{10}(1+\frac{z}{600}) \tag{7}$$

z refers to the audio frequency value based on the frequency scale, and f_{mel} refers to the audio frequency value under the Mel scale.

B. Audio Feature Selection and Evaluation

Different audio signal processing methods would obtain different features. The proposed feature extraction and classification model provides higher accuracy in music classification [19]. The early development of musical features underwent three stages of classification. The first time it was divided into four parts: semantic features, short-term features, component features, and long-term features. The second time is a more detailed division of music features based on the first time, which is divided into tone, timbre, and rhythm. The third time is based on the ideas of physics, proposing two concepts: acoustic features and perceptual features. The concepts of several classified music features mentioned above provide a direction for future research. There are now many methods to classify music features in a more detailed and clear manner, which can better reflect the similarities and differences between genres. The most common features currently are time-domain features, cepstrum features that are close to human auditory perception, and frequency domain features.

As the preferred choice for audio feature processing, time-domain features require less computation and the extraction steps are not complex. Common time-domain characteristics include short-time Zero-crossing rate (ZCR), and the calculation formula is:

$$\mathrm{ZCR} = \frac{1}{2}\sum_{i=1}^{I-1}\left|\mathrm{sgn}\left[y(m+1)\right] - \mathrm{sgn}\left[y(m)\right]\right| \tag{8}$$

Among them, $y(m)$ refers to the discrete signal, and the $\mathrm{sgn}(y)$ unction is as:

$$\mathrm{sgn}(y) = \begin{cases} 1 & y > 0 \\ 0 & y = 0 \\ -1 & y < 0 \end{cases} \tag{9}$$

e frequency domain feature acquisition step first utilizes Fourier transform to transform music signals from time domain to frequency domain, and then performs statistical analysis and calculation on the signals within the frequency domain. The commonly used parameters for frequency domain features include SE (Spectral Entropy), SF (Spectral Flux), Spectral Centroid (SC), and Spectral Rolloff (SR).

The spectral entropy SE represents the relationship between power spectrum and entropy rate, and the formula is:

$$\mathrm{SE} = -\sum_{F=0}^{\frac{Fw}{2}} Q(f)\log_2\left[Q(f)\right] \tag{10}$$

Among them, $Q(f)$ refers to the power spectral density and Fw refers to the sampling frequency.

The spectrum represents the frequency distribution of an audio signal, and the spectral centroid is a way to represent the center of the spectrum. The number of high-frequency components of an audio signal is directly proportional to the value of the spectral centroid. The formula for the spectral centroid is:

$$\mathrm{SC} = \frac{\sum_{w=i}^{h} w\left|F(w)\right|^2}{\sum_{w=i}^{h}\left|F(w)\right|^2} \tag{11}$$

The formula for spectral flux SF is:

$$\mathrm{SF} = \frac{1}{h-i}\sum_{w=i}^{h}\left|F(w+1) - F(w)\right| \tag{12}$$

V. Experiment on Classifying Electronic Music Genres

In order to verify the genre classification performance of each model, a dataset publicly available on W website was selected, and 2700 different types of electronic music were selected as the experimental data for this article. Among them, there were 300 tracks each for Dubstep, Trip-Hop, Ambient, House, Techno, Trance, Disco, Electro, and Glitch. Each training sample was monophonic, 100s in duration, and sampled at 20,050 Hz. After the experimental data was prepared, the samples were input into the CNN model, the PMG-Net electronic music genre classification model and the traditional classification model for training. The optimizer selected Adam (Adam Optimizer), and the loss function selected Cross Entry. The learning rate was set to 0.0003; the training round was set to 12, and the batch size was set to 128. Finally, the graphics processing unit (GPU) was responsible for accelerating the training. The traditional classification model, CNN model, and PMG-Net electronic music genre classification model were set as T1, T2, and T3, respectively. The classification performance of each model is shown in Fig. 4.

Fig. 4 shows the classification performance results of the three models after 12 training sessions, which is equivalent to the corresponding model accuracy. The higher the accuracy, the better the model performs in classifying electronic music genres. Among them, the accuracy distribution of the T1 model was between 0.3 and 0.53, and the best classification performance occurred during the 6th training session; the accuracy distribution of T2's model was between 0.3 and 0.62, and the best classification performance occurred during the first training; the accuracy distribution of the T3 model was between 0.542 and 0.75, and the best classification performance occurred during the 8th training. In addition, in 12 model training sessions, if the accuracy is greater than or equal to 0.5, T1 would be trained 2 times; T2 would be trained 4 times, and T3 would be trained 12 times. The training results indicated that among these three models, T3 had the best classification performance and the highest accuracy for electronic music genres.

Fig. 5 reflects the classification error changes of the three models with different training rounds. During the training process, the errors in the previous rounds of training were generally relatively large. T1 and T2 approached convergence after six rounds of training, and the classification error gradually decreased to less than 1. T3 approached convergence after four rounds of training, and the classification error gradually decreased to less than 1. Moreover, the error of each round of T3 training was smaller than T1 and T2. The fluctuation range of T3 error was relatively small, and the classification was relatively stable.

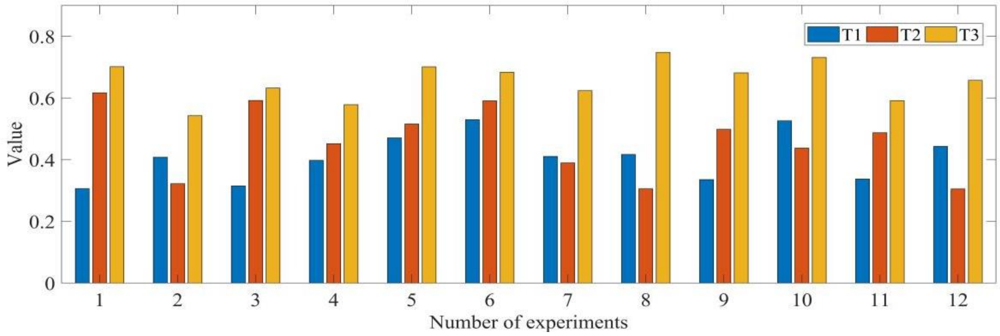

Fig. 4. Classification performance of three models trained 12 times.

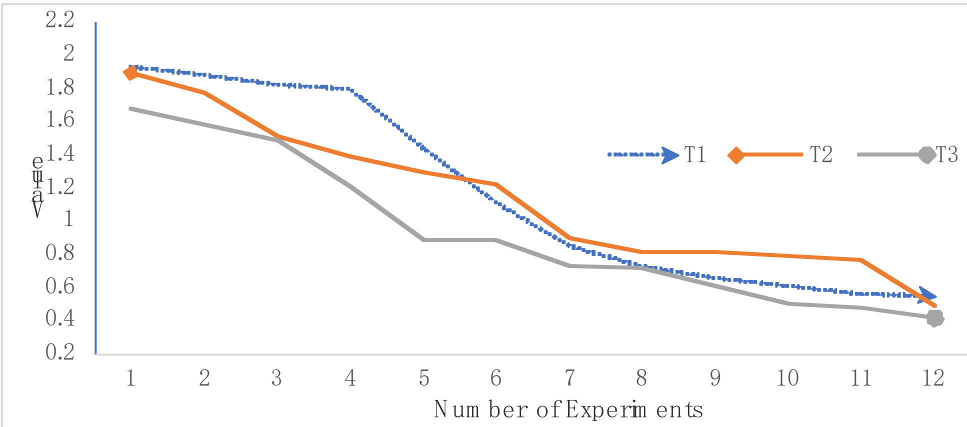

Fig. 5. Changes in classification error with training rounds.

To classify electronic music genres, it is not only necessary to consider classification accuracy, but also to evaluate the model's signal processing speed and feature recognition speed for audio samples. Signal processing was conducted using 2700 pre experimental mono audio samples. It was known that each sample had a duration of 100s and a sampling frequency of 20050 Hz. 12 batches of training were conducted, and the average speed of the total number of times was taken as the processing result, as shown in Table I.

In Table I, T1's audio signal processing duration was 53.92s-59.94s, and the fastest electronic music type was Techno; T2's audio signal processing duration was 45.99s-54.88s, and the fastest electronic music type is the Trip-Hop; the audio signal processing duration of T3 was 30.22s-39.02s, and the fastest electronic music type was Glitch. The experimental results showed that T1 and T2 were not as fast as T3 in audio signal processing, and T3 had good processing performance.

Audio signals have short-term stationarity. Generally, the processing and analysis of audio signals are based on "short-term", which indicates that their characteristics remain unchanged within a short time range (10s-30s). Therefore, when extracting the features of audio signals, the signal is segmented and windowed, and the results at each frame are actually the return value of the feature extraction function. Using the experimental data and conditions of the audio signal mentioned above to collect feature extraction speed, the results are shown in Fig. 6.

TABLE I. SIGNAL PROCESSING SPEED OF THREE MODELS FOR AUDIO SAMPLES (IN SECONDS)

	Dubstep	Trip-Hop	Ambient	House	Techno	Trance	Disco	Electro	Glitch
T1	59.83	54.33	59.74	58.3	53.92	59.94	57.49	55.63	55.89
T2	46.81	45.99	53.46	53.65	48.24	47.65	47.18	54.04	54.88
T3	30.73	32.57	32.79	34.67	30.69	31.96	37.62	39.02	30.22

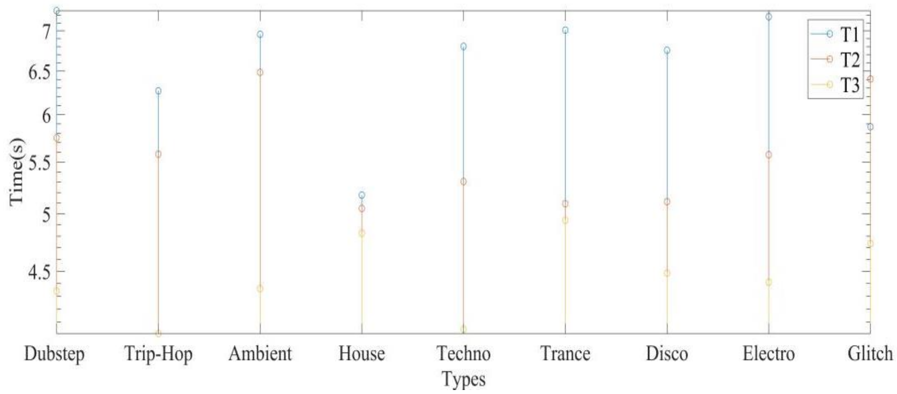

Fig. 6. Audio feature extraction speed under three models.

In Fig. 6, for nine types of audio samples, including Dubstep, Trip-Hop, Ambient, House, Techno, Trance, Disco, Electro, and Glitch, the feature extraction speed of T1 was the fastest at 5.18 seconds, with most feature extraction speeds maintained between 6-7 seconds; the feature extraction speed of T2 for these nine types of audio samples was the fastest at 5.05 seconds, with most feature extraction speeds maintained between 5-6 seconds; the fastest feature extraction speed for these nine types of audio samples in T3 was 4.01 seconds, and the overall feature extraction speed remained between 4-5 seconds. The feature extraction speed of T3 was faster than that of T1 and T2.

The traditional electronic music genre classification model had unstable performance, low classification accuracy, and single music signal features. The CNN model took the second place, and the PMG-Net electronic music genre classification model had the best classification effect on electronic music genres.

In order to further illustrate the classification effect of the PMG-Net electronic music genre classification model chosen in this paper, firstly, without any comparable results, 100 electronic music samples of different genres were randomly selected from the publicly available dataset of W network for 10 experiments. Among them, 20 samples each of reverberating heavy beat Dubstep, godly dance Trip-Hop, ambient Ambient, hoedown House, and techno were selected. The results are shown in Table II.

According to the results in Table II, it is found that the PMG-Net electronic music genre classification model classifies the accurate number very close to the standard classification number. The high degree of accuracy close to the standard number of classifications means that the model is more capable of generalization. Even when confronted with new, unseen music samples, the model is able to accurately categorize them into the correct genre. This suggests that the model does not just memorize the training data, but learns the universal music genre characteristics.

Secondly PMG-Net electronic music genre classification model in the case of comparing with convolutional neural network model and traditional classification model, the classification results of convolutional neural network model and traditional classification model are shown in Table III and Table IV.

TABLE II. PMG-NET ELECTRONIC MUSIC GENRE CLASSIFICATION MODEL NUMBER OF CORRECT CLASSIFICATIONS (IN SAMPLES)

	Dubstep	Trip-Hop	Ambient	House	Techno
1	20	20	20	19	19
2	20	20	20	20	20
3	20	20	20	20	20
4	20	19	20	20	19
5	19	20	20	20	20
6	20	19	19	20	19
7	20	20	20	20	20
8	19	20	20	20	20
9	20	20	20	19	20
10	20	20	19	20	19

TABLE III. TRADITIONAL CLASSIFICATION MODEL NUMBER OF CORRECT CLASSIFICATIONS (IN SAMPLES)

	Dubstep	Trip-Hop	Ambient	House	Techno
1	11	11	16	15	14
2	12	10	12	14	16
3	11	11	17	11	16
4	11	14	13	16	12
5	11	17	16	15	10
6	10	11	16	12	15
7	11	14	12	12	10
8	16	14	11	15	17
9	16	12	13	16	16
10	17	11	11	10	16

According to the results in Table III, it was found that the number of correct classifications for the traditional classification model ranged from 10 to 17. There are errors in every music classification.

TABLE IV. CNN CLASSIFICATION MODEL NUMBER OF CORRECT CLASSIFICATIONS (IN SAMPLES)

	Dubstep	Trip-Hop	Ambient	House	Techno
1	11	18	16	15	14
2	12	16	12	14	16
3	18	11	17	18	16
4	11	14	18	16	12
5	18	17	16	15	17
6	13	18	16	12	15
7	11	14	12	12	17
8	16	14	11	15	17
9	16	18	13	16	16
10	18	11	11	11	16

According to the results in Table IV, it was found that the number of correct classifications of the convolutional neural network model ranged from 11 to 18. It can be seen that the PMG-Net electronic music genre classification model has particularly obvious classification advantages both in comparison with other existing classification models and in direct experiments.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, a classification model was constructed based on deep neural network, and the publicly available dataset of W network was selected as the experimental sample. The sample data were input into the convolutional neural network model, the PMG-Net electronic music genre classification model, and the traditional classification model for training, respectively. The experiments were conducted in the following four aspects: in the aspect of the model classification performance, the classification performance of PMG-Net electronic music genre classification model was the best, and the training accuracy was high in every round. In terms of model classification performance, the PMG-Net electronic music genre classification model had the best classification performance, with a high accuracy of more than 0.5 in each round of training; in terms of model classification error, the PMG-Net electronic music genre classification model had a small classification error and was stable; in terms of audio signal processing, the PMG-Net electronic music genre classification model had the fastest processing speed, with a processing time of less than 40 seconds for each round of processing; in terms of feature extraction for audio samples, both the convolutional neural network model and the traditional classification model had the fastest feature extraction speed. Classification models had feature extraction speeds of more than five seconds, and only the PMG-Net electronic music genre classification model had a feature extraction speed of less than five seconds.

In conclusion, using deep neural networks to customize the PMG-Net model for electronic music genre classification provides higher classification accuracy, automated processing, flexibility, scalability, real-time performance, and personalized classification results. These advantages make the PMG-Net model a powerful tool for handling electronic music genre classification tasks.

However again, because there may be some fuzzy boundaries between electronic music genres, that is, certain music may be characterized by more than one genre at the same time, it is difficult to categorize them unambiguously. This requires the PMG-Net model to have some flexibility and robustness to handle diversity and ambiguity.

VII. CONCLUSION

Electronic music is particularly popular in the world today due to its vibrant rhythm and diverse forms of expression, and is increasingly favored by the public. Therefore, it is very necessary to classify electronic music genres, which can achieve comprehensive retrieval and meet the music needs of different people. People can experience the ultimate charm brought by electronic music. The PMG-Net electronic music genre classification model customized by using deep neural network had better results than the convolutional neural network model and traditional classification model in the experiments of model classification performance, classification error, audio signal processing and feature extraction speed. It can be seen that the PMG-Net electronic music genre classification model customized by deep neural network is very suitable for classifying electronic music genres, fast and accurate.

The research approach of using DNN for electronic music genre classification has certain reference value for future research on automatic classification of other music genres. The drawback of this article is that the experimental electronic music audio sample size is relatively small. Therefore, future work needs to continue to collect more available electronic music samples and input them into the model to improve the model's generalization ability. The architecture and parameter settings of the deep neural network-based PMG-Net classification model have been explored and improved to improve classification performance. Different network architectures, activation functions, loss functions, etc., or techniques such as transfer learning and integrated learning can be tried to further improve the model performance.

ACKNOWLEDGMENT

2020 science and technology research fund project of Jiangxi Provincial Department of education "Research on LBG electronic music detection method based on computer-aided technology" (Project No: GJJ209912).

REFERENCES

[1] Zhang, K. Music Style Classification Algorithm Based on Music Feature Extraction and Deep Neural Network. Wireless Communications and Mobile Computing, 2021, 2021(4): 1-7.

[2] Pelchat, N., Gelowitz, C. M. Neural network music genre classification. Canadian Journal of Electrical and Computer Engineering, 2020, 43(3): 170-173.

[3] Dong, A. M., Liu, Z. Y., Yu, J. G., Han, Y. B., Zhou, Q. Automatic classification of music genres based on visual transformation networks. Computer Applications, 2022, 42(S01): 54-58.

[4] Oramas, S., Barbieri, F., Nieto, Caballero O., Serra, X. Multimodal deep learning for music genre classification. Transactions of the International Society for Music Information Retrieval. 2018, 1 (1): 4-21.

[5] Elbir, A., Aydin, N. Music genre classification and music recommendation by using deep learning. Electronics Letters, 2020, 56(12): 627-629.

[6] Liu, W., Wang, J., Qu, H., Dong, L., Cao, X. Music genre classification algorithm based on spectral space domain feature attention. Computer Applications, 2022, 42(7):2072-2077.

[7] Fan, S. H. Music genre classification based on improved BP neural network. Software Engineering, 2021, 24(9): 17-20.

[8] Geirhos, R., Jacobsen, J. H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., et al. Shortcut learning in deep neural networks. Nature Machine Intelligence, 2020, 2(11): 665-673.

[9] Liu, C., Arnon, T., Lazarus, C., Strong, C., Clark, B., Kochenderfer, M. J. Algorithms for verifying deep neural networks. Foundations and Trends® in Optimization, 2021, 4(3-4): 244-404.

[10] Durstewitz, D., Koppe, G., Meyer-Lindenberg, A. Deep neural networks in psychiatry. Molecular psychiatry, 2019, 24(11): 1583-1598.

[11] Bau, D., Zhu, J. Y., Strobelt, H., Lapedriza, A., Zhou, B., Torralba, A. Understanding the role of individual units in a deep neural network. Proceedings of the National Academy of Sciences, 2020, 117(48): 30071-30078.

[12] Li, S., Jia, K., Wen, Y., Liu, T., Tao, D. Orthogonal deep neural networks. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(4): 1352-1368.

[13] Wen, Z. Y. Research on music style classification model based on swarm intelligence optimization neural network. Modern electronics, 2019, 42(21): 82-85.

[14] Elbir, A., Aydin, N. Music Genre Classification and Music Recommendation by Using Deep Learning. Electronics Letters, 2020, 56(12): 627-629.

[15] Du, W., Lin, X., Sun, J. W., Yu, B., Yao, K. F. An automatic music classification method based on hierarchical structure. Miniature Microcomputer Systems, 2018, 39(5): 888-892.

[16] Nam, J., Choi, K., Lee, J., Chou, Y. S., Yang, H. Y. Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach. IEEE signal processing magazine, 2018, 36(1): 41-51.

[17] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Muller, K. R. Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 2021, 109(3): 247-278.

[18] Su, J., Vargas, D. V., Sakurai, K. One pixel attack for fooling deep neural networks. IEEE Transactions on Evolutionary Computation, 2019, 23(5): 828-841.

[19] Thiruven, G. R. Speech/music classification using PLP and SVM. International Journal of Engineering and Computer Science, 2019, 8(2): 24469-24472.