Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4556135, 8 pages
https://doi.org/10.1155/2022/4556135

Research Article

The Application of Minority Music Style Recognition Based on Deep Convolution Loop Neural Network

Qiaozhen Fan

School of Tourism Management, Guilin Tourism University, Guilin 541006, China

Correspondence should be addressed to Qiaozhen Fan; fqz@gltu.edu.cn

Received 10 January 2022; Revised 25 January 2022; Accepted 10 February 2022; Published 29 March 2022

Academic Editor: Xin Ning

In recent years, with the development of Internet and digital audio technology, music information retrieval has gradually become a research hotspot. Due to the rise of deep learning and machine learning in recent years, as well as the rapid improvement of computer software and hardware performance, it has laid a good foundation for identifying different genres of music. Among them, the application of minority music style recognition is also an important research direction. At present, the application performance of minority music style recognition based on deep convolution loop neural network is poor. Because convolution loop neural networks (CNNs) have strong ability to capture information features, this paper uses CNN to extract various features from music signals and classify them. Firstly, the original music signal spectrum is separated into time characteristic harmonic component and frequency characteristic impact component by using the harmonic/percussive sound separation (HPSS) algorithm. Combined with the original spectra as the input of CNN, the network structure of CNN is designed, and the influence of different parameters in the network structure on the recognition rate is studied. Experiments on minority music data sets, compared with other scholars' music recognition methods, it shows that this method can effectively improve the recognition of minority music styles using a single feature.

1. Introduction

Music recognition is a very challenging and promising task in the field of MIR (Music Information Retrieval) [1]. Because music is a developing art and there is no clear boundary between music styles, automatic recognition of music is a challenging problem. The key of music recognition is the feature extraction of music information. A variety of feature extraction and classification methods have been proposed in recent years [2, 3]. The performance of these classifiers is highly dependent on the appropriateness of manually extracted features selected by experience. Generally, feature extraction and classification in recognition tasks are two independent processing stages; in this paper, these two stages are integrated to better realize the interaction between information. Recently, deep convolution neural network CNN (convolutional neural networks-works) [4] in general recognition task [5, 6] has made remarkable progress continuously, and this aroused people's concern about CNN's classification model [7, 8]. Lee et al. [9] was the first person to apply deep learning to music recognition. Especially style and artist identification, by training a convolution deep confidence network (CDBN) with two hidden layers to try to activate the hidden layer in an unsupervised way, generate meaningful features from the preprocessed spectrum. Compared with those standard MFCC (Mel Frequency Cepstral Coefficients) features, its deep learning features have higher accuracy. For music style recognition, Li et al. [10] transformed music data into MFCC feature vectors and input them into convolution neural network (CNN) with three hidden layers. Finally, it is concluded that CNN can automatically extract image features for classification. Literature [11, 12] puts forward other applications of CNN for recognition, which shows that CNN has strong ability to capture changing image information features. Tzanetakis and Cook [13] recognize different music styles and proposes

music quality and pitch recognition technologies. Literature [14, 15] extracts acoustic features for traditional musical instruments and vocal music data to recognize music.

CNN includes multistage processing of input information, extract multilayer ,and high-level feature representations. By sharing some basic components, many manually extracted features and corresponding classification methods can be regarded as an approximate or special CNN. However, in order to retain discriminative information, these features and methods must be carefully designed and integrated. Inspired by the remarkable success of CNN in general recognition tasks, this paper applies CNN to challenging music recognition and studies the influence of the adjustment of network structure parameters on recognition rate.

2. Related Theories of Deep Convolution Neural Network

2.1. Deep Learning-Related Concepts. Deep learning has been further improved on the basis of machine learning. Because the deep learning architecture contains more layers of networks, more features can be obtained when analyzing features, which improve the learning ability of networks [16]. Compared with shallow network, deep network can use few neurons to fit the same function, which is more efficient and accurate in the learning process. At present, deep learning has been used in various fields, such as finance, security, and manufacturing [17]. Relevant knowledge of deep learning will be introduced in the following contents.

2.2. Deep Convolution Neural Network. One of the most basic units of deep convolution neural network is called neuron, which can also be called perceptron. The neural network is shown in Figure 1:

The connection weights are w_{ij}, a_i, and b_j that represent the deviation thresholds of explicit layer and implicit layer, respectively. Given the bias and weights, a mapping model of input variables and output variables can be established, in which the explicit layer is used as the input of the model for exchanging external information, and the implicit layer is used as the output of the model for extracting feature information. The probability distribution of CNN can be realized by the energy function, which is defined as formula (1) under a given state (v, h):

$$E_\theta = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{nm} b_j h_j + \sum_{i=1}^{n}\sum_{j=1}^{m} v_i w_{ij} h_j. \tag{1}$$

By visualizing and regularizing formula (1), the joint probability distribution of CNN can be obtained as shown in formula (2):

$$P(v, h\,;\theta) = \frac{1}{z(\theta)}\ \exp\left(-E_\theta(v, h)\right). \tag{2}$$

In formula (2), $z\ (\theta)$ is the normalization factor, and its value can be expressed as formula (3):

$$C_a = C_a(X_a), \forall a \in L, \tag{3}$$

where in formula (3), $\theta = \{h_m, v_n, w_{ij}, a_i, b_j\}$ is the network parameter. When the explicit layer vector v is given, the probability that the implicit element j is activated in formula (4):

$$P\left(h_j = 1|v\right) = R\left(\sum_{t=1}^{n} w_{ij} v_i + b_i\right). \tag{4}$$

Similarly, the probability that explicit I is activated in formula (5):

$$P\left(h_j = 1|v\right) = R\left(\sum_{t=1}^{n} w_{ij} v_i + b_i\right), \tag{5}$$

where R represents the ReLU activation function, which is used to activate neurons. The purpose of training CNN is to learn constantly. The optimal network parameters $\theta = \{w_{ij}, a_i, b_j\}$ are obtained to minimize the training error of samples. In the process of sample training, the maximum likelihood estimation is used to optimize the network parameter θ, the purpose of which is to make all the output sample values closest to the actual values of the samples under a certain distribution probability by optimizing the parameters, which is essentially a problem of finding the likelihood function value. The CNN model obtains the optimal fitting training data based on formula (6):

$$\theta^* = \operatorname{argmax} L(\theta) = \operatorname{argmax} \sum_{t=1}^{T} \log_i\left(v^{(t)|\theta}\right). \tag{6}$$

The parameter update method is shown in Formula (7):

$$\theta^{t+1} = \theta^t + \varepsilon \frac{\partial \mathrm{InL}(\theta)}{\partial \theta}, \tag{7}$$

where ε represents the learning rate and t the number of iterations. Partial derivative of parameters in the CNN model using the random gradient rise method to solve the problem, the partial derivative derivation process is as formula (8):

$$L(\theta) = \sum_{t=1}^{T} \log\left(P\left(v^{(t)|\theta}\right)\right) = \sum_{t=1}^{T} \log \frac{\sum_h \exp(-E(v^t, h \mid \theta))}{\sum_h \sum_h \exp(-E(v^t, h \mid \theta))}. \tag{8}$$

In the actual training of CNN, by reconstructing the distribution of explicit neurons and hidden neurons, the approximate results are obtained partial derivatives to

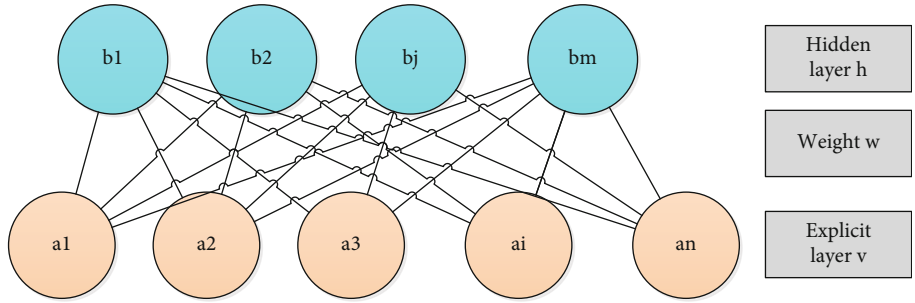

Figure 1: Neural network diagram.

parameters, as shown in Formulas (9)–(11):

$$\frac{\partial \log\left(P(v|\theta)\right)}{\partial w_{ij}} = \langle v_i h_j \rangle - \left\langle v_i{}' h_j{}' \right\rangle, \tag{9}$$

$$\frac{\partial \log\left(P(v|\theta)\right)}{\partial a_i} = \langle v_i \rangle - \left\langle v_i{}' \right\rangle, \tag{10}$$

$$\frac{\partial \log\left(P(v|\theta)\right)}{\partial b_j} = \langle h_j \rangle - \left\langle h_j{}' \right\rangle. \tag{11}$$

By constantly adjusting the parameters, the hidden layer can be used as the feature of the input data of the visual layer. The joint configuration can be expressed as formula (12):

$$E(v, h, \theta) = -\sum_{ij} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_i a_j h_j \, \theta = \{W, a, b\}. \tag{12}$$

Determine the joint probability distribution formula of a configuration by Boltzmann machine (13):

$$Z(\theta) = \sum_{h,v} \exp(-E(v, h, \theta)). \tag{13}$$

The conditions between hidden layer nodes are independent, that is, formula (14):

$$p(h|v) = p\left(h_j|v\right). \tag{14}$$

By factorizing the above formula, we can get the formula (15) about the probability that the node J of the hidden layer is 0 or 1:

$$p\left(h_j = 1|v\right) = \frac{1}{1 + \exp\left(-\sum_{ij} W_{ij} v_i - a_j\right)}. \tag{15}$$

The learning ability of CNN is several times that of the ordinary neural network. For this model, it can be abstracted as converting the preoutput into the postinput, so as to spit out the feature information layer by layer, and this way can maximize the effectiveness of this information. At present, CNN is trained by layer-by-layer calculation:

(1) Obtain the weight value of CNN by training first

(2) Based on parameter optimization, orderly adjustment is adopted. Because this weight is a refinement of the training results of the model and abandons the original weight disorder, the training purpose can be achieved without a large number of iterations

3. Music Genres and Music Characteristics

3.1. Classification of Music Genres. Under the contrast of artistry and appreciation, various music schools have been formed, and different styles have been derived from them, such as brisk, melancholy, and deep, each of which shows different artist styles [18]. There are similarities and differences between them. There are some songs whose classification labels are not defined according to people's sensory characteristics. For example, common baroque music is defined according to a historical era, which includes different regions and different styles. In addition, Indian music is a musical style type defined according to geographical location. In addition, there is a common problem in the classification of music genres; that is, the same song may belong to jazz music and rock music at the same time.

Through the characteristics of music, different minority music will be formed into many styles, each style has different characteristics, and the music styles of different ethnic minorities are also reflected through this. In the early stage of recognition technology, the classification of minority music styles mainly depends on manpower. With the rapid development of neural networks, music styles can often be classified by machine learning, and music analysis can be carried out by known music styles [19] as shown in Figure 2.

3.2. Music Feature Selection. The music characteristic is the performance music essential attribute, in order to distinguish the different ethnic minority style music, and carries on the extraction to the music characteristic is very important. Feature extraction has many methods, and feature selection also has many, if we can overselect the appropriate features, then we can make the experimental results more accurate. The classification of music features can be classified by artificial feelings, mainly through listening perception, and the characteristics of music are tone, timbre, and loudness. The other is divided into different features by Li et al. [20] and others according to music features [21, 22].

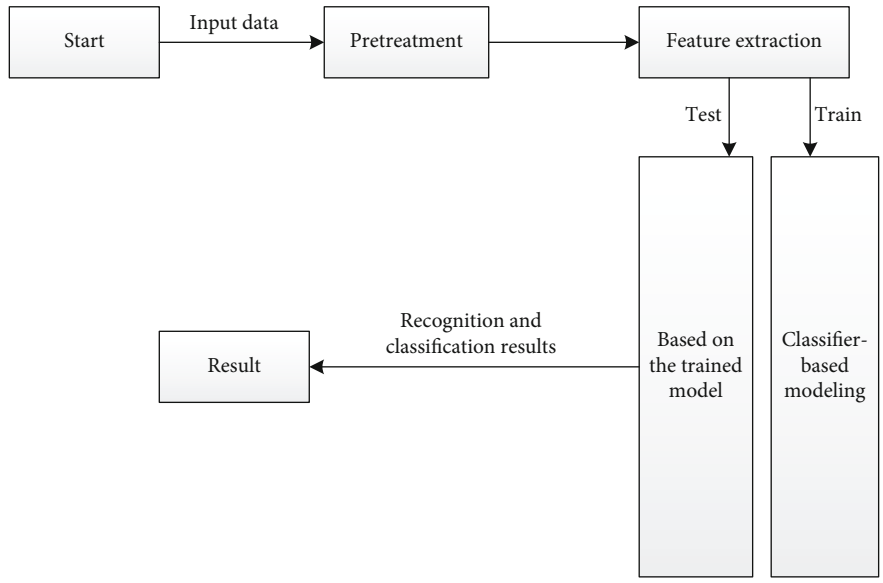

Figure 2: Flow chart of music style recognition.

Table 1: Training hyperparameter table.

Hyperparametric	Learning rate η	Batchsize	Momentum coefficient μ	Weight attenuation coefficient λ	Dropout coefficient
Value	0.01	16	0.9	0.0005	0.5

For the same tone, music features can be distinguished by timbre, and different sound qualities can also make people distinguish different playing instruments. The size of music is reflected by loudness. Not only that but also people can also feel the tonal features.

When a feature can be expressed numerically, it means that it is a short feature, in which timbre, loudness, and tone are all short features. If the numerical value cannot be expressed, it is called time domain feature.

(a) Short feature can express the amplitude value of music vibration at a certain time point, and the calculation method is as follows Formula (16):

$$E_{\mathrm{n}} = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 = \sum_{m=n-(N-1)}^{n} [x(m)w(n-m)]^2. \tag{16}$$

The sampling time point is expressed by n, m is the expression of window function, and n is the length of window function.

(b) Probability of short feature greater than zero is an important eigenvalue in the analysis of high frequency energy. In waveform diagram, the larger the value of high frequency quantity, the more the shuttle times of zero point are. Calculation formulas such as formula (17) are as follows:

$$Zn = \frac{1}{2N} \sum_{m=n-(N-1)}^{n} \mathrm{sgn}[x(m)] - \mathrm{sgn}\,[x(m-1)]w(n-m), \tag{17}$$

where $x(m)$ is the value of the m-th sampling point in time, sgn represents a symbolic function, and sgn can be expressed as a formula (18):

$$\mathrm{sgn} = \begin{cases} 1 & x(n) >= 0, \\ 0 & x(n) < 0. \end{cases} \tag{18}$$

In addition, the frequency domain feature vector can be used for analysis, which usually contains spectrum centroid and spectrum energy. These two music feature vectors are usually used in music signal processing and analysis, and the calculation formula of spectral energy is as formula (19):

$$\mathrm{SE} = \sqrt{\frac{1}{h_0 - l_0} \sum_{w-l_0}^{h_0} |F(w)|^2}. \tag{19}$$

In the formula, the value of W is between i_0 andh_0, the minimum value of frequency is represented by $i_{0,}$ and the maximum value of frequency is represented byh_0. The calculation formula of spectrum centroid is

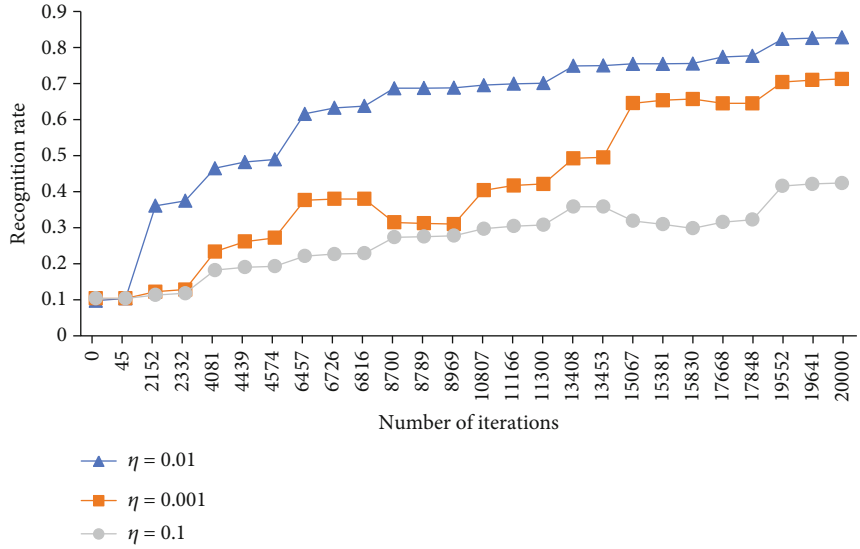

Figure 3: Effect of learning rate.

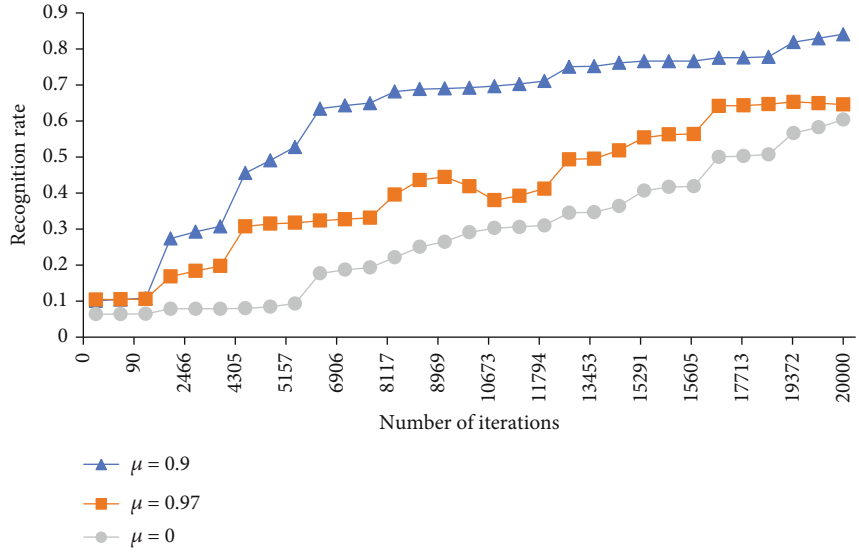

Figure 4: Influence of momentum coefficient μ.

shown in formula (20):

$$\mathrm{sc} = \frac{\sum_{w=l_0}^{h_0} w|F(w)|^2}{\sum_{w=l_0}^{h_0} |F(w)|^2}. \tag{20}$$

3.3. Music Style Recognition Algorithm

3.3.1. HPSS Algorithm. Using the HPSS algorithm to separate music tracks, the original tracks are separated into harmonic sound source and impact sound source; then, these two kinds of sound sources and original tracks are transformed by short-time Fourier transform, the transformed spectra are input into CNN network for learning, training, and prediction, and the final output result is the final recognition rate.

3.3.2. Harmonic/Percussive Separation Algorithm. Music signals are usually composed of harmonic sound components and impact sound components, and it has very different characteristics [23]. In this paper, the harmonic/percussion separation algorithm is used to separate harmonic and impact sound components of music signals, and the key of this method is to focus on the difference of the continuous direction of harmonic spectrum and impulse spectrum. Harmonic spectrum is usually continuous in time direction, while impulse spectrum is continuous in frequency.

3.3.3. Web-Based Training and Learning Methods. The network structure of this paper is a layered deep convolution neural network, it extracts local features by convolving input images and a set of kernel filters, and the convolution layer generates a feature map through a linear convolution filter and a nonlinear activation function (ReLU). The output of

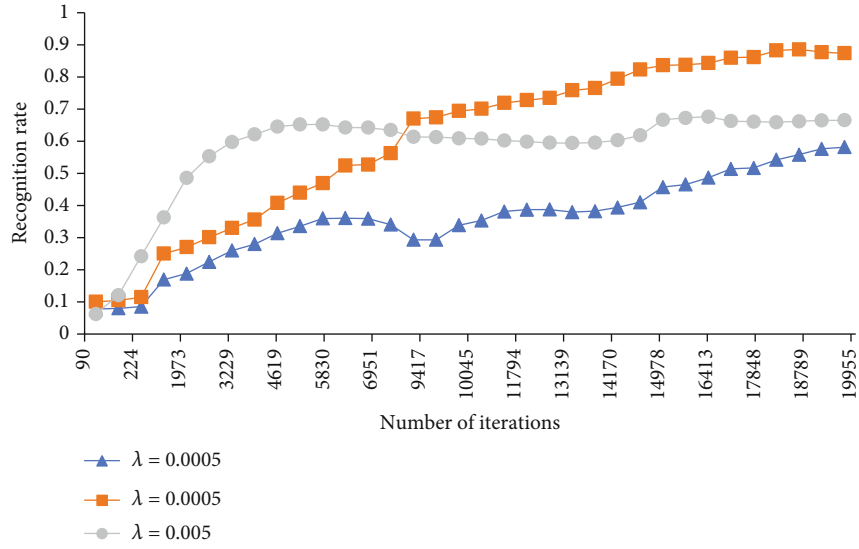

Figure 5: Effect of weight attenuation λ.

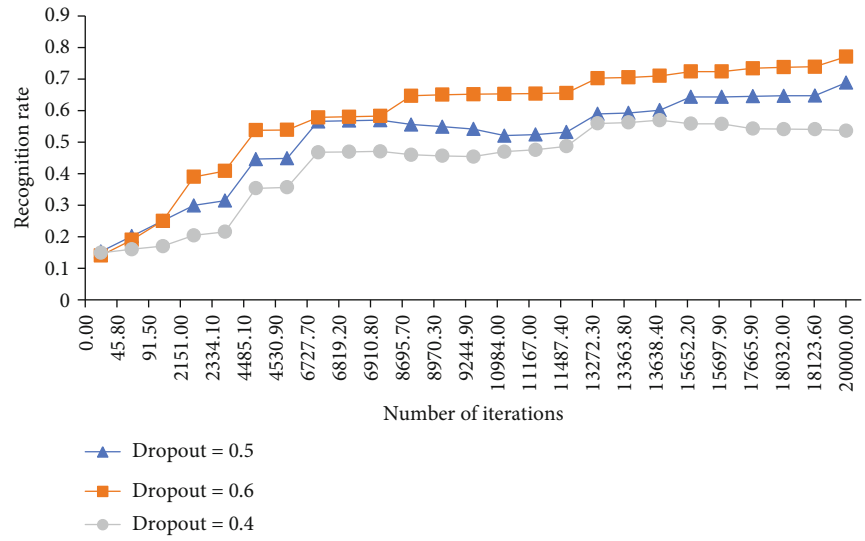

Figure 6: Effect of dropout value.

neurons in the same layer forms a plane, it is called a feature map, and then the convolution feature map is obtained by pooling and filtered to the next layer. Different feature maps are obtained by setting different kernel filters in the local receptive field. Given X_1^q represents the p-th feature map at the l-level, the convolution of the whole feature map and the activation function used are shown in formula (21):

$$x_t^q = \max\left(0, \sum x_{t-1}^q\right). \tag{21}$$

The activation function is used after the convolution operation. Since it is found that the normalization of local response is helpful to the generalization of the network, this response normalization achieves a form of lateral inhibition found in real neurons, and the effect of this lateral inhibition is to make the output values of neurons calculated by different convolution kernels more sensitive to the activity of neurons with larger calculated values.

4. Experimental Simulation Analysis

4.1. Data Preparation. For the whole experiment, we collected 10 different ethnic minority styles of music from the Internet, and each style category contained 100 audio recordings, which lasted for 30 seconds and had a total of 1,000 music festival selections. For the whole data set, we use recognition rate as a performance index to help us recognize music more efficiently.

4.2. Parameter Settings. Experiments in this paper show the importance of correctly adjusting hyperparameters, which can be divided into two types: model-related

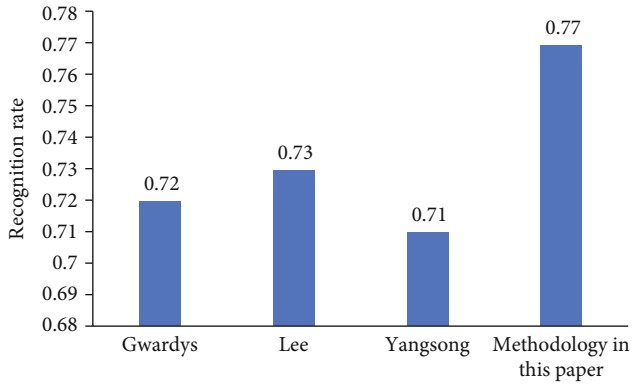

Figure 7: Comparison of this method with other methods.

hyperparameters and training-related hyperparameters, as shown in Section 2.2 and Table 1.

To adjust these hyperparameters, press 4 : 1 is randomly divided into two subsets; that is, 800 music tracks are used for training, and 200 music tracks are used for testing. Training-related hyperparameters can significantly affect the convergence and learning rate of the network, and their influence is illustrated by the recognition rate curve, as shown in Figures 3–6. In each diagram, we focus on one hyperparameter, while the others are set to the best values in Table 1.

4.2.1. Influence of Learning Rates on Recognition Rate. Figure 3 shows that for 20000 iterations of training samples, as long as the rate is below 0.001, the process of learning will be very slow, and the recognition rate is not stable. If it is too large such as 0.1, the learning process will be unstable, and the classification performance will be reduced.

4.2.2. Effect of Momentum on Recognition Rate μ. Figure 4 illustrates the influence of momentum on recognition rate μ. It shows that using momentum μ can speed up the learning process well. At the same time, if it is too large, such as 0.97, it will cause oscillation in the initial stage and slow convergence. In addition, it reduces the recognition performance in the later stage.

4.2.3. Influence of Weight Attenuation λ on Recognition Rate. Figure 5 illustrates the effect of weight attenuationλ, indicating that a smaller λ seems to be a safer choice, while a larger λ, such as 0.005, can undermine the stability of the learning process.

4.2.4. Effect of the Dropout Value. Dropout is a technique to prevent overfitting in the process of training neural network. In this experiment, the dropout is fine-tuned to 0.4 or 0.5 and 0.6. When the dropout value is increased, the training time is slightly longer, and the convergence is slow. The training takes 20000 iterations. The classifier based on CNN can produce good classification performance after 20000 iterations.

4.3. Comparison of Accuracy with Other Methods. At present, many scholars have put forward different research methods for music style recognition. As shown in Figure 7, Gwardys and Grzywczak [24] also uses the HPSS algorithm to obtain spectrogram and then fine-tuned an 8-layer network. Finally, the accuracy rate is 72%. This paper uses this spectrogram to train this 8-layer network. The accuracy has improved. EOL spatiotemporal [25] trained a CDBN with only two layers. The depth of recognition model is shallower than that in this paper, and the amount of data is also less than that in this paper, but the experimental results are very close to the accuracy before expanding the data in this paper. It can be seen that small data sets can also have good results in shallow networks. Yang and Yu [26] use K-means clustering in traditional machine learning methods to recognize music, and its recognition rate is 71%. Compared with the deep learning method proposed in this paper, the classification accuracy is improved to some extent.

5. Conclusion

In this paper, a method of music style recognition based on convolution neural network is proposed, the network framework used in this method is designed in detail, and some key factors affecting its classification performance are studied. When the original spectrogram is used for experiments at first, the recognition rate is only 67%. In order to improve the results, harmonic/percussive separation experiment was carried out, and finally, the recognition rate is improved by 6%. In addition, this experiment shows the importance and effectiveness of data expansion, especially when the training data is insufficient, when the experimental data are expanded, the experimental result is improved by 3%. The experiment in this paper has achieved a certain recognition rate, using the framework of convolution neural network. Coupled with data expansion and other means, make it possible to apply deep learning to small data integration and has achieved certain discrimination ability. In the future work, on the one hand, the recognition rate is improved by collecting more music tracks of each style. On the other hand, a hybrid network structure can be constructed. For example, convolution neural network and cyclic neural network can be used to extract local features, and cyclic neural network can integrate these extracted features in time. Because cyclic neural network has memory function for the previous information, the extracted features can be more integrated and coherent, and the recognition results can be improved.

Data Availability

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

Conflicts of Interest

The author declared that he/she has no conflicts of interest regarding this work.

Acknowledgments

Project supported by "2021 Improvement Project in Guangxi University of Basic Scientific Research Ability for Young and Middle-Aged Teachers,". title: "A Study on the Smart Mechanism Development of the Protection of Guangxi Music Intangible Cultural Heritage under the Background of Urbanization" (2021KY0130).

References

[1] N. M. Patil and M. U. Nemade, "Content-based audio classification and retrieval: a novel approach," in *International Conference on Global Trends in Signal Processing, Information Computing and Commu-nication(ICGTSPICC), IEEE, International Conference on Global Trends in Signal Processing*, pp. 599–606, Jalgaon, India, 2017.

[2] Y. M. G. Costa, L. S. Oliveira, A. L. Koerich, F. Gouyon, and J. G. Martins, "Music genre classification using LBP textural features," *Signal Processing*, vol. 92, no. 11, pp. 2723–2737, 2012.

[3] L. Meng, S. Ding, and Y. Xue, "Research on denoising sparse autoencoder," *International Journal of Machine Learning & Cybernetics*, vol. 8, no. 5, pp. 1719–1729, 2017.

[4] Z. Kai and D. Shi-fei, "Advances in image super-resolution reconstruction," *Computer Engineering and Applications*, vol. 53, no. 16, pp. 29–35, 2017.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural netw orks," *Advances in Neural Infor-mation Processing Systems (NIPS)*, vol. 25, pp. 1097–1105, 2012.

[6] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision- based data-driven pavement distress detection," *Construction & Building Materials*, vol. 157, pp. 322–330, 2017.

[7] A. Sharif Razavian, H. Azizpour, and J. Sullivan, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedingsof the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 806–813, Columbus, OH, USA, 2014.

[8] M. Ling-heng and D. Shi-fei, "Depth perceptual model based on the single image," *Journal of Shandong University*, vol. 46, no. 3, pp. 37–43, 2016.

[9] H. Lee, P. Pham, Y. Largman, and A. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in Neural Information Processing Systems(NIPS)*, pp. 1096–1104, British Columbia, Canada, 2009.

[10] T. L. Li, A. B. Chan, and A. H. Chun, "Automatic musical pattern feature extraction using convolutional neural network," *Lecture Notes in Engineering & Computer Science*, vol. 2180, no. 1, 2010.

[11] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rating image aesthetics using deep learning," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2021–2034, 2015.

[12] G. E. Hinton, N. Srivastava, and A. Krizhevsky, "Improving neural networks by preventing co-adaptation of feature detectors," *Computer Science*, vol. 3, no. 4, pp. 212–223, 2012.

[13] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[14] S. A. Lashari, R. Ibrahim, and N. Senan, "Soft set theory for automatic classification of traditional Pakistani musical instruments sounds," in *IEEE International Conference on Computer Information Science*, pp. 94–99, Kuala Lumpur, Malaysia, 2012.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, and S. Reed, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, 2015.

[16] G. Sun, C. C. Chen, and S. Bin, "Study of cascading failure in multisubnet composite complex networks," *Symmetry*, vol. 13, no. 3, p. 523, 2021.

[17] G. Widmer, "Discovering simple rules in complex data: a meta-learning algorithm and some surprising musical discoveries," *Artificial Intelligence*, vol. 146, no. 2, pp. 129–148, 2003.

[18] G. E. Hinton, S. Osindero, and Y. W. Teh, "a fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[19] M. Bretan, G. Weinberg, and L. Heck, "A unit selection methodology for music generation using deep neural networks," in *Proceedings of the 8th International Conference on Computational Creativity (ICCC 2017)*, pp. 72–79, Atlanta, GA, USA, June 2017.

[20] Y. Li, J. Zhang, and D. Pan, "A study of muscic recognition based on RBM model," *Journal of Computer Research & Development*, vol. 51, no. 9, pp. 1936–1944, 2017.

[21] G. Chen and S. Li, "Network on chip for enterprise information management and integration in intelligent physical systems," *Enterprise Information Systems*, vol. 15, no. 7, pp. 935–950, 2021.

[22] M. Fan, "Application of music industry based on the deep neural network," *Scientific Programming*, vol. 2022, Article ID 4068207, 6 pages, 2022.

[23] Y. H. Yang, Y. C. Lin, Y. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008.

[24] G. Gwardys and D. Grzywczak, "Deep image features in music information retrieval," *International Journal of Electronics and Telecom-munications (IJET)*, vol. 60, no. 4, pp. 321–326, 2014.

[25] EOL Spatiotemporal and SOR Related, "AI context unsupervised feature learning for audio classification using convolutional deep belief networks," in *International Conference on Neural Information Processing Systems*, pp. 1096–1104, Curran Associates Inc, Boston, MA, USA, 2009.

[26] S. Yang and F.-q. Yu, "Speech/music discriminator based on sample entropy," *Computer Engineering and Applications*, vol. 48, no. 23, pp. 125–127, 2012.