Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2021.020406
ISSN 2790-2692 Vol. 2, Issue 4: 41-48

Recognition of Film and Television Background Music Based on Convolutional Neural Network

Ahthasham Alassery*

University of Balochistan, Pakistan

**corresponding author*

Keywords: Convolutional Neural Network, Background Music, Music Emotion, Music Recognition

Abstract: Music is the product of consciousness and emotion, and is closely related to human life. The background music of film and television under big data has received extensive attention. Emotion is one of the main semantic information contained in music. Classification based on emotion can deeply explore music categories from multiple perspectives and improve the efficiency of music recognition. In recent years, it has gradually become a research hotspot in background music recognition. The research purpose of this paper is to identify the background music of film and television based on convolutional neural network. In the experiment, using the convolution operation, the selected eleven musical instruments basically cover a variety of music types, so as to identify the background music of film and television.

1. Introduction

With the development of digital audio technology, a large amount of music is stored in the network music database, music recognition is of great significance in video soundtrack and music information retrieval, and related research is also increasing [1]. However, existing music recognition models and techniques for extracting music features have encountered bottlenecks. Traditional classification models are difficult to extract deep music features, have poor accuracy, have poor generalization ability for music features, and cannot adapt to different data sets. Looking back at the development of music recognition, with the continuous development of deep learning technology, it has gradually been introduced into the field of music emotion recognition. A breakthrough has been made in music recognition technology at home and abroad.

With the growth of music data, music recognition based on convolutional neural networks has become imminent.Poulakis believes that cinema is often considered a modern visual medium. But despite the fact that film content consists of sound and images, the fact that the auditory dimension

of this audio medium is often overlooked in education and news commentary. On the other hand, there are also works that deal with so-called "film music", which use specific musical patterns, similes, and original words. they are from music experts [2]. Herget A showed that, despite being frequently used and widely considered in practice, background music in non-fictional media formats has shown very negative results in previous practice studies. Students were also frequently advised against using music in formats such as television news, newspapers and documentaries. Differences in the effects of background music have also been found in film and advertising studies. Consistency between music and medium has been shown to be important in predicting the effect of music. Two experiments were conducted. The first test focused on the emotions expressed and evoked by the music, the memory performance of the recipient, perceived credibility and general evaluation of the media form. The second test focused on behavioural change [3]. This paper proposes a related research on video background music recognition based on convolutional neural network.

This paper studies the overview of convolutional neural networks, including an introduction to neural networks and convolutional neural networks. Finally, the selection of the recognition method of film and television background music based on convolutional neural network is expounded. In the experiment, using the convolution operation, the selected eleven musical instruments basically cover a variety of music types, so as to identify the background music of film and television.

2. Research on Recognition of Film and Television Background Music Based on Convolutional Neural Network

2.1. Research Background and Significance

The problems related to the search of a large number of online music works have become increasingly prominent, and have attracted widespread attention in the academic community. In this context, the technology of retrieving music based on the emotional attributes of music has also been developed. A large number of music are stored in the network music database, and music information retrieval has gradually become a research hotspot [4-5]. The essence of music information retrieval is music identification and classification. The vast majority of music end users tend to like a certain type of music, and the diversity of music endows it with unique attributes. Therefore, a music identification and classification system can help people to classify music more effectively. Search and manage. MIR contains many sub-tasks, such as music emotion recognition, musical instrument recognition, genre recognition, author recognition, etc., among which music emotion recognition occupies an important position in the field of music information retrieval.

At present, the mainstream emotion recognition of film and television background music mainly consists of three steps: 1) Select an emotion model (continuous emotion model and discrete emotion model); 2) Preprocess music to extract useful music features and information as input; 3) Input to the recognition model for emotion recognition. The most critical part is the extraction of music emotional features and the construction of recognition and classification models [6]. Previous studies have often used single sentiment features or classifiers based on traditional machine learning models. A single emotional feature is often not uniform and cannot fully express the emotion of music. It needs to be re-extracted when performing different recognition tasks. Although the recognition effect has been improved to a certain extent, the efficiency is too low. The traditional machine learning model only has outstanding results in the recognition of unique music features, but the effect is not ideal for unfamiliar music, and the generalization ability is poor. It can be seen that the main breakthrough is in the second and third steps [7]. The basic structure of music recognition classification is shown in Figure 1:

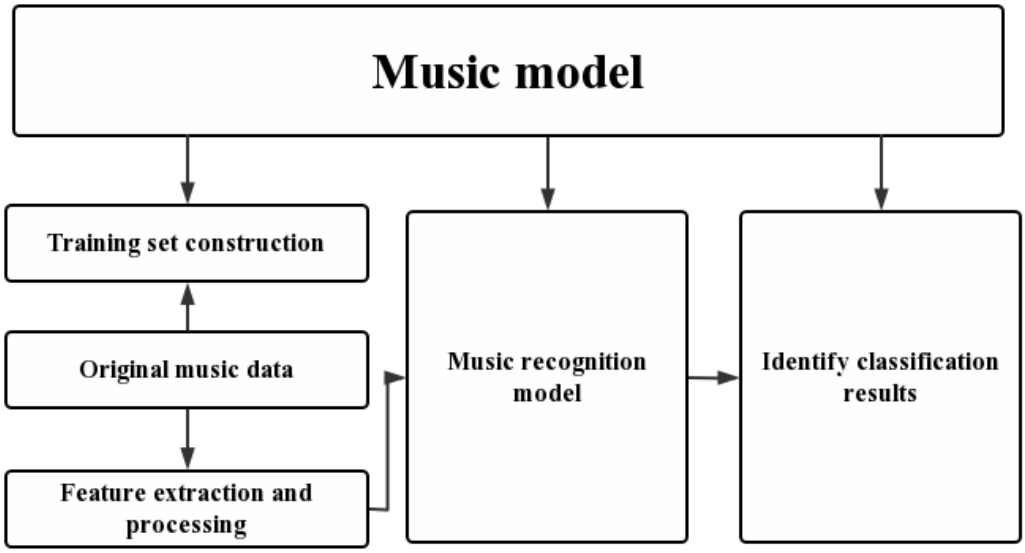

Figure 1. Basic structure of music recognition classification

2.2. Overview of Convolutional Neural Networks

As a classic feedforward neural network, the convolutional neural network is also an important part of the combined network in this paper. Inspired by the study of the cat's optic nerve, experts expounded the two ideas of convolution and pooling [8-9]. At this point, it began to lead people to imagine how to make computers observe the world like living things, and the convolutional neural network was born. Today, many powerful and classic networks are derived from variants of convolutional neural networks, which have achieved remarkable achievements in image classification, speech recognition, etc. Convolutional neural networks have become an indispensable presence in artificial neural networks. Convolutional neural network was first proposed in the field of image recognition and achieved very good results, and then it was widely used in many fields such as image classification, image segmentation, and image denoising.

Convolutional neural networks have huge advantages in feature extraction, and the ability to extract features can also be improved through training, making it a mainstream network in deep learning image processing and an irreplaceable part of deep learning [10- 11]. The main structures of convolutional neural network are convolutional layer, batch normalization layer, activation function and pooling layer. In image processing, the convolution operation mainly uses two-dimensional convolution, the purpose is to obtain the feature map of the image, extract the main features of the image, and remove redundant information. The batch normalization operation normalizes the data features so that the data obeys a normal distribution and prevents the data features from being scattered and difficult to train. The activation function is generally set after the convolutional layer to increase nonlinear factors and improve the expressiveness of the model. The pooling layer downsamples the feature map to further compress the features [12].

2.3 Selection of Recognition Method of Film and Television Background Music Based on Convolutional Neural Network

Music emotion recognition is mainly divided into discrete emotion recognition and continuous

emotion recognition [13-14]. The continuous emotion recognition method is to cut the whole piece of music into several segments, make an emotion label value for each segment, and generate the corresponding regression value based on the dimensional space model. The choice of method therefore determines the flow and results of the experiment. Continuous emotion recognition mainly relies on the general continuous emotional space model. 16]. However, it is significantly different from the commonly used emotional semantics. For example, the emotional state represented by two-dimensional coordinates is not intuitive, and it is also significantly different from cognitive habits. A new architecture is set up to extract and chromatogram-related feature information from the original signal during processing, and then send it to the convolutional neural network synchronously. Meet the recognition performance requirements [17-18]. In the process of emotional feature extraction of music, specific recognition research is carried out through convolutional neural network, and the recognition effect of different models is analyzed. Due to the complex data set of continuous emotion recognition, the characteristics of audio must be huge and cumbersome, and how to design an effective recognition model is the key; discrete emotion recognition has a relatively small amount of data, but due to the particularity and subjectivity of music emotion, its Feature selection is difficult to grasp and innovate. Combining the above two points, the two identification methods can make breakthroughs on the original basis [19].

3. Investigation and Research on Recognition of Film and Television Background Music Based on Convolutional Neural Network

3.1. Research Content

In order to identify the background music of film and television, the eleven musical instruments selected in this paper basically cover four types of music: rock music, classical music, pop music, and jazz music. And the training set and the test set all have a certain proportion of the four types of music. On the one hand, the influence of music type on instrument recognition cannot be ruled out; on the other hand, when the training set lacks a certain type of music, if this type of music appears in the test set, the recognition effect will not be ideal.

3.2. Convolution Operation

For the initial features of the input, intermediate features are learned in the multi-layer convolution process, and finally advanced features that are conducive to classification and recognition are learned.The expression for the convolution operation is:

In CNN, $x(t)$ is the input feature, $w(t)$ is the convolution kernel. When dealing with two-dimensional matrix data, the above formula can be written as:

$$s(t) = x(t) * w(t) = \sum_{\tau=-\infty}^{\tau=+\infty} x(\tau) w(t-\tau) \tag{1}$$

The size of the convolution kernel is $M \times N$, and the inner product operation is performed between the convolution kernel and the corresponding area of the input feature matrix.

$$s(i,j) = \sum_{m=0}^{M} \sum_{n=0}^{N} \left(w_{m,n}, x_{i+m} + w_b \right) \tag{2}$$

4. Analysis and Research on Recognition of Film and Television Background Music Based on Convolutional Neural Network

4.1. Music Recognition under Estimated Pitch

The object identified here is the type of musical instrument in the background music of the film and television. The overall accuracy is defined as the ratio of the number of correct identifications of all instruments to the total number of identifications of all instruments, including the number of correct identifications, the number of incorrect identifications, and the number of unidentified times. The specific score results are shown in Table 1 and Figure 2:

Table 1. Identification data table of film and television music background musical instruments

Type	Harmonic mapping matrix order			
	1	2	3	4
Violin	0.844	0.871	0.712	0.698
Viola	0.872	0.701	0.617	0.691
Saxophone	0.876	0.867	0.821	0.751
Kettledrum	0.871	0.841	0.791	0.783
Side drum	0.807	0.834	0.830	0.847
Xylophone	0.860	0.819	0.871	0.691
Piano	0.789	0.791	0.701	0.887
Gu he	0.791	0.794	0.858	0.737
Bass	0.793	0.756	0.904	0.857
Overall accuracy	0.792	0.718	0.893	0.658

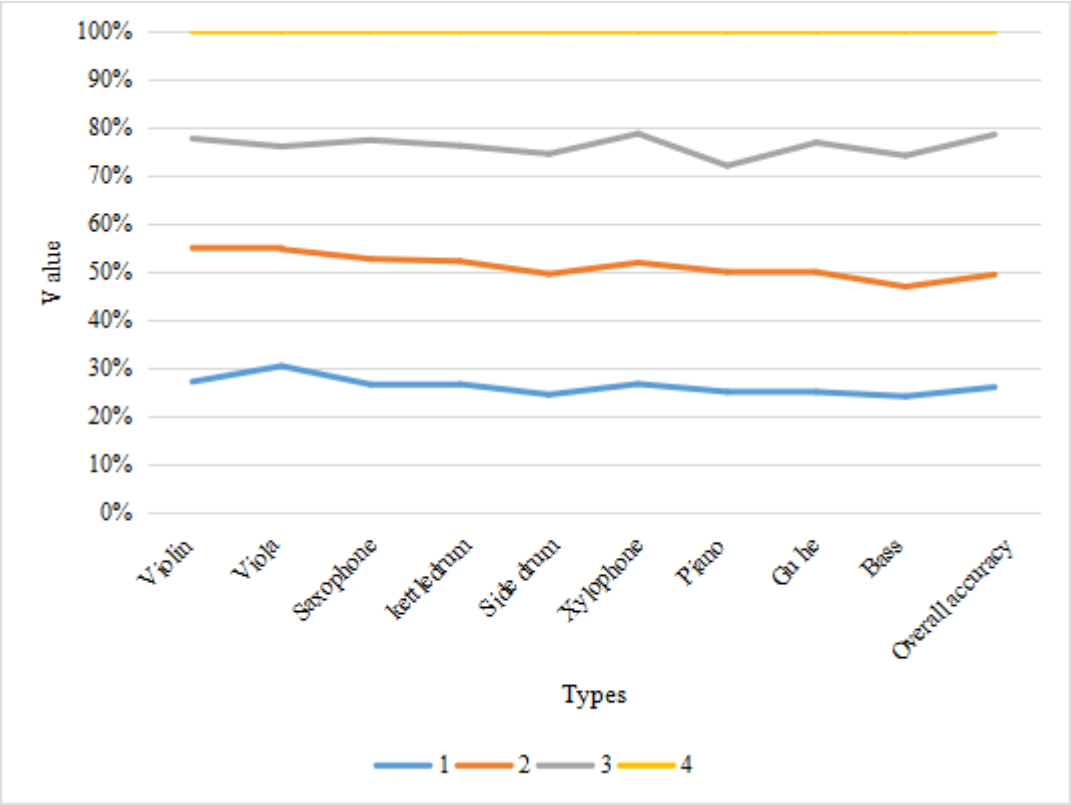

Figure 2. The actions and overall accuracy of the eleven instruments under using estimated pitch

4.2. Music Recognition at Real Pitch

To demonstrate the effectiveness of adding pitch as an input feature, we use the true pitch label matrix to form the harmonic mapping matrix (n is 1 to 4), which means that the values in the pitch feature matrix are all correct values. The scores and overall accuracy of the eleven instruments using real pitch are shown in Table 2 and Figure 3:

Table 2. Music identification data for real high notes

Type	1	2	3	4
Violin	0.970	0.963	0.871	0.905
Viola	0.837	0.879	0.874	0.917
Saxophone	0.965	0.963	0.904	0.928
kettledrum	0.871	0.951	0.954	0.936
Sidedrum	0.864	0.964	0.987	0.915
Xylophone	0.871	0.934	0.986	0.978
Piano	0.961	0.961	0.998	0.936
Guhe	0.871	0.955	0.991	0.984
Bass	0.845	0.896	0.890	0.981
Overallaccuracy	0.984	0.982	0.984	0.987

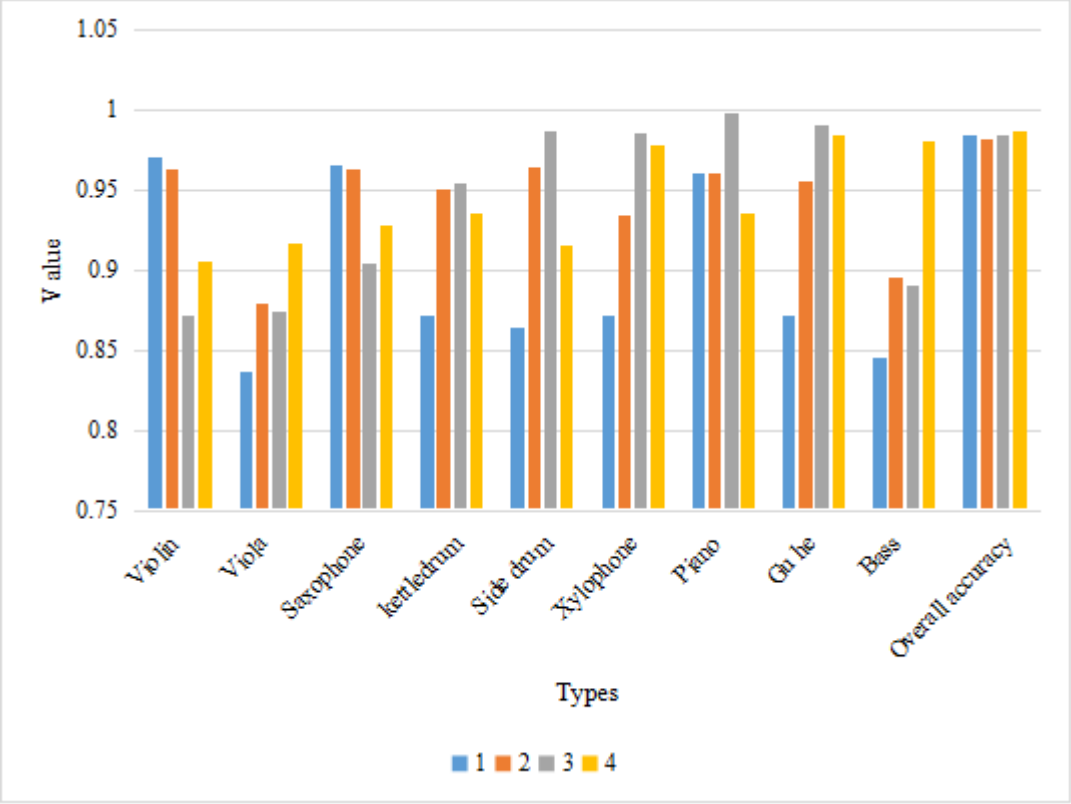

Figure 3. Fractions and overall accuracy of the eleven instruments using the true pitch

The results show that using the harmonic mapping matrix composed of real pitch labels as input features, the scores and overall accuracy of various instruments are higher than using the harmonic mapping matrix composed of estimated pitch labels as input features. It shows that there is a positive correlation between the pitch estimation effect and the musical instrument recognition effect. In addition, we can also observe from the comparison of the two tables that compared with

the timpani and xylophone, the recognition score of the snare drum is not much different in the comparison of using the real pitch or the estimated pitch, which may be because the snare drum is a percussion without a fixed pitch. instrument, while the other two are fixed-pitch. And we are based on the pitch feature. The recognition effect of musical instruments with clear pitch and obvious overtones is better, so the recognition score of orchestral instruments is significantly higher than that of percussion instruments. From this perspective, the effectiveness of selecting pitch features can also be seen.

5. Conclusion

Convolutional neural networks have achieved remarkable success in image recognition (such as handwritten digit recognition), sound signal processing, and medical diagnosis. These problems seem to be different types of problems, but their characteristics are actually the same. The essence is to use the convolutional neural network to learn the internal characteristics of the data model in a large amount of data, so as to construct a model that can almost accurately describe the data. , and then generalize to the same kind of unknown data, and give the estimation result to realize the transition from known to unknown. As an art form, music expresses the author's thoughts and moods through the combination of "rhythm", "melody" and "harmony", thereby arousing people's emotional resonance. Music emotion has the characteristics of strong subjectivity and rich style. Background music of film and television has cultural and artistic value and commercial value. Therefore, the background music of film and television based on convolutional neural network has gradually attracted the attention of the majority of researchers.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Jennifer L. Improving passions: sentimental aesthetics and American film, by Charles Burnetts, Edinburgh, Edinburgh University Press, 2017, 192 pp. $105.00 (hardback), ISBN 9780748698196. New Review of Film & Television Studies, 2018, 16(3):355-359. https://doi.org/10.1080/17400309.2018.1481900

[2] Poulakis N. Review of "Understanding Sound Tracks through Film Theory" by Elsie Walker and "Sound: Dialogue, Music, and Effects" edited by Kathryn Kalinak. Film & History An Interdisciplinary Journal of Film and Television Studies, 2018, 48(2):30-34.

[3] Herget A K, Albrecht J. Soundtrack for reality? How to use music effectively in non-fictional media formats. Psychology of Music, 2021, 50(2):508-529. https://doi.org/10.1177/0305735621999091

[4] Venczel P. A Brief Analysis of the Functionality and Dramaturgy of the Soundtrack in Film and

Theater (Part I.). Theatrical Colloquia, 2020, 10(2):138-165. https://doi.org/10.2478/tco-2020-0027

[5] Florya A V. Intertextual Elements in the Poetics of the Film "CARGO 200" by A. Balabanov. Bulletin of Udmurt University Series History and Philology, 2019, 29(6):1071-1080. https://doi.org/10.35634/2412-9534-2019-29-6-1071-1080

[6] Larson J. Improving passions: sentimental aesthetics and American film. New Review of Film and Television Studies, 2018, 16(3):1-5. https://doi.org/10.1080/17400309.2018.1481900

[7] Neupert R. Agnès Varda between film, photography, and art. New Review of Film and Television Studies, 2018, 16(3):1-5. https://doi.org/10.1080/17400309.2018.1480577

[8] Crim B E. "I got no problem killing my kin": Fury (2014) and the evolution of the World War II combat film. Film & History an Interdisciplinary Journal of Film and Television Studies, 2018, 48(1):4-14.

[9] Knox S. Shameless, the push-pull of transatlantic fiction format adaptation, and star casting. New Review of Film and Television Studies, 2018, 16(3):1-29. https://doi.org/10.1080/17400309.2018.1487130

[10] Wilks L. Her stories: daytime soap opera & US television history: by Elana Levine, Durham and London, Duke University Press, 2020, 400 pp. $29.95 (paperback), ISBN 9781478008019. New Review of Film and Television Studies, 2021, 19(2):242-244. https://doi.org/10.1080/17400309.2021.1918494

[11] Carter L. Unsettled Scores: Politics, Hollywood, and the Film Music of Aaron Copland and Hanns Eisler. By Sally Bick. Music and Letters, 2021, 101(4):807-810. https://doi.org/10.1093/ml/gcaa064

[12] Pett E, Warner H. The Invisible Institution? Reconstructing the History of BAFTA and the 1958 Merger of the British Film Academy with the Guild of Television Producers and Directors. Journal of British Cinema and Television, 2020, 17(4):449-472. https://doi.org/10.3366/jbctv.2020.0542

[13] Sticchi F. European cinema and continental philosophy: film as thought experiment. New Review of Film and Television Studies, 2019, 17(2):1-5. https://doi.org/10.1080/17400309.2019.1587220

[14] Walton S. Cruising the unknown: film as rhythm and embodied apprehension in L'Inconnu du lac/Stranger by the Lake (2013). New Review of Film and Television Studies, 2018, 16(3):1-26. https://doi.org/10.1080/17400309.2018.1479183

[15] Guimares J P. Beyond the Pulp Vanguard: Bruce Andrews's Film Noir Series and the Dead-End of Escapist Experimentalism. CounterText, 2021, 7(3):467-500. https://doi.org/10.3366/count.2021.0247

[16] Novakovi M. Zoran simjanovi's Balkan Ekspres mix: The status of a song between the archival and the original film music. Zbornik Akademije umetnosti, 2019, 2019(7):120-132. https://doi.org/10.5937/ZbAkUm1907120N

[17] Ahn S H. Reconsidering the Locus of Film Music: Diegesis, Non-Diegesis, and In-between Spaces. Music Theory Forum, 2019, 26(1):197-216. https://doi.org/10.16940/YMR.2019.26.1.197

[18] Justus T. Toward a naturalized aesthetics of film music: An interdisciplinary exploration of intramusical and extramusical meaning. Projections, 2019, 13(3):1–22. https://doi.org/10.3167/proj.2019.130302

[19] Rajaobelina P L, Dusseault P, Ricard L. The mediating role of place attachment in experience and word of mouth: The case of music and film festivals. International Journal of Arts Management, 2019, 21(2):43-54.