Pretrained Deep Learning Model for Music Genre Classification

, Shirley Stevany Faryl A[1], Vinothini S[2], Praga deeshwari D[3], Finney Daniel Shadrach[4], Dr Anitha R[5]
[4,5]Assistant Professor, [1,2,3]UG student, *Department of ECE*
KPR Institute of Engineering and Technology
Coimbatore, Tamil Nadu
finneydaniels@gmail.com

Abstract - Classifying music files according to their genre is a challenging task in the area of Music Information Retrieval (MIR). Music genre classification is an important multimedia research domain which includes categorizing of music databases, music piece representation and distances between genres. In this paper pretrained DNN models such as alexnet and vgg are used for music classification.

Keywords—Genre, Classification, Music

INTRODUCTION

Music is a ubiquitous topic in our society as nearly everyone enjoys listening to it and many creating it. The research field of Music Information Retrieval (MIR) is primarily concerned with the extraction and inference of meaningful features from music, indexing of music using these features, and the development of different search and retrieval schemes. Companies like Soundcloud, Apple Music, Spotify, Wynk etc. use music classification, either to provide recommendations to their customers based on their previous search history, or simply as a product (like Shazam). To carry out any of the above two mentioned functions, determining music genres is the first step. Music is characterized by giving them categorical labels called genres. These genres are created by humans. A music genre is segregated by the characteristics which are commonly shared by its members. Typically, these characteristics are related to the rhythmic structure, instrumentation, and the harmonic content of the music. To categorize music files into their respective genres is a very challenging task in the area of MIR. Classification of genre can be very valuable to explain some interesting problems such as creating song references, tracking down related songs, discovering societies that will like that specific song, sometimes it can also be used for survey purposes. Automatic musical genre classification can assist humans or even replace them in this process and would be of a very valuable addition to music information retrieval systems. In addition to this, automatic classification of music into genres can provide a framework for development and evaluation of features for any type of content-based analysis of musical signals. The concept of automatic music genre classification has become very popular in recent years as a result of the rapid growth of the digital entertainment industry. Dividing music into genres is arbitrary, but there are perceptual criteria that are related to instrumentation, structure of the rhythm and texture of the music that can play a role in characterizing particular genre. Until now genre classification for digitally available music has been performed manually. Consequently, the techniques used for automatic genre classification would add value to the development of audio information retrieval systems for music.

DATASET

GTZAN Dataset

For our experiments, we used the GTZAN dataset [21] which, although it has some shortcomings [19], is a benchmark for Music Genre Classification. The GTZAN dataset consists of ten genre classes: Blues, Classical, Country, Disco, HipHop, Jazz, Metal, Pop, Reggae, and Rock. Each class consists of 100 recordings of music pieces of 30 s duration. These excerpts were taken from radio, compact disks, and MP3 compressed audio files. Each item was stored as a 22.050 kHz, 16-bit, mono audio file.

LITERATURE SURVEY

Mutiara, A. B., Refianti, R., & Mukarromah, N. R. A. (2016) classified the music genre by using audio features on several kernels of non-linear Support Vector Machine (SVM). The dataset used is GTZAN, which contains 1000 music files of 30-second-long duration recorded using 22050 Hz sample rate and 16-bit sample size in mono channel. This data set consists of ten genres namely classical, blues, country, hip pop, disco, jazz, pop, metal, rock and reggae.

There are 100 music files in each genre. An experiment is carried out for the classification of the genre by extracting features related to rhythm, timbre, LPC, and tonality from the music files. The features that are related to timbre are spectral flux, spectral centroid, spectral roll off, zero-crossing. For rhythmic features, the strength of the strongest beat and beat sum are used. Chromagram and key strength are the tonal features used. Different kernels of the SVM classifier (RBF, linear and polynomial kernels) are used to classify the music genre of the music files from the extracted musical features. The classification achieved an accuracy of 76.6%.

Song, Y., Zhang, C., & Xiang, S. (2007) proposed a semi-supervised genre based classification algorithm based on labelled as well as unlabeled music tracks. Three main features extracted are Mel frequency cepstral coefficients, Fluctuation Pattern and Spectrum Histogram. The Mel frequency cepstral coefficients and Spectrum Histogram were used to characterize the timbre feature of a music track. Fluctuation Pattern is another feature used to describe music tracks periodicity. ISMIR2004 dataset is used in the proposed work. The classification accuracy achieved for MFCC is greater than 75%.

Simsekli, U. (2010) suggested bass lines method for classification of music genres. A bass line is an instrumental melody which combines melodic, harmonic and rhythmic features and contains information for classification of genres sufficiently that is played by low-pitched instrument. The dataset used is McKay and Fujinaga's MIDI data set that contains 3 root and 9 leaf classes where Jazz, Rhythm & Blues, and Rock are the root genres where each root genre is divided into three leaf genres. The MIDI files are used to extract the bass lines. "Melodic Interval Histograms" are used as features. The classification algorithms utilized are k-nearest neighbour and the results are compared with SVMs on a standard MIDI database. A novel distance metric and perceptually weighted Euclidean distance (PWED) were proposed apart from metrics for k-nearest neighbor namely Euclidean, earth mover's, symmetric KullbackLeibler, normalized compression distances. The classification accuracy of 84% is achieved for k-nearest neighbor in general. The root labels achieved a maximum accuracy of 100% and the leaf labels achieved a maximum accuracy of 84.44%.

A. AlexNet

AlexNet is the name of a convolutional neural network which has had a large impact on the field of machine learning, specifically in the application of deep learning to machine vision. It famously won the 2012 ImageNet LSVRC-2012 competition by a large margin.

AlexNet contains 5 convolutional layers and 3 fully connected layers. Relu is applied after very convolutional and fully connected layer. Dropout is applied before the first and the second fully connected year. The network has 62.3 million parameters and needs 1.1 billion computation units in a forward pass. We can also see convolution layers, which accounts for 6% of all the parameters, consumes 95% of the computation.

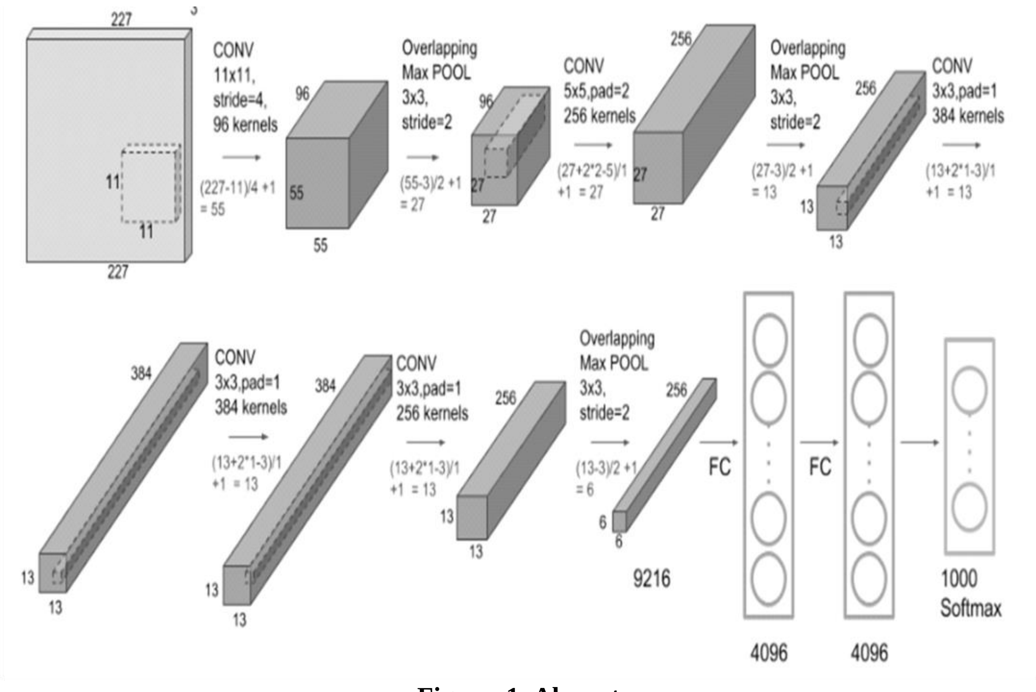

Figure 1. Alexnet

B. VGGNet

VGGNet is similar to AlexNet except withadditional convolution layers. VGGNet consists of 13 convolution, rectification, pooling and 3 fully connected layers [21]. The convolution network uses 3 × 3 windows size filter and 2 × 2 pooling network. VGGNet performs better as compared to AlexNet due to its simple architecture. The underlying architecture of VGGNet is illustrated in the Fig.

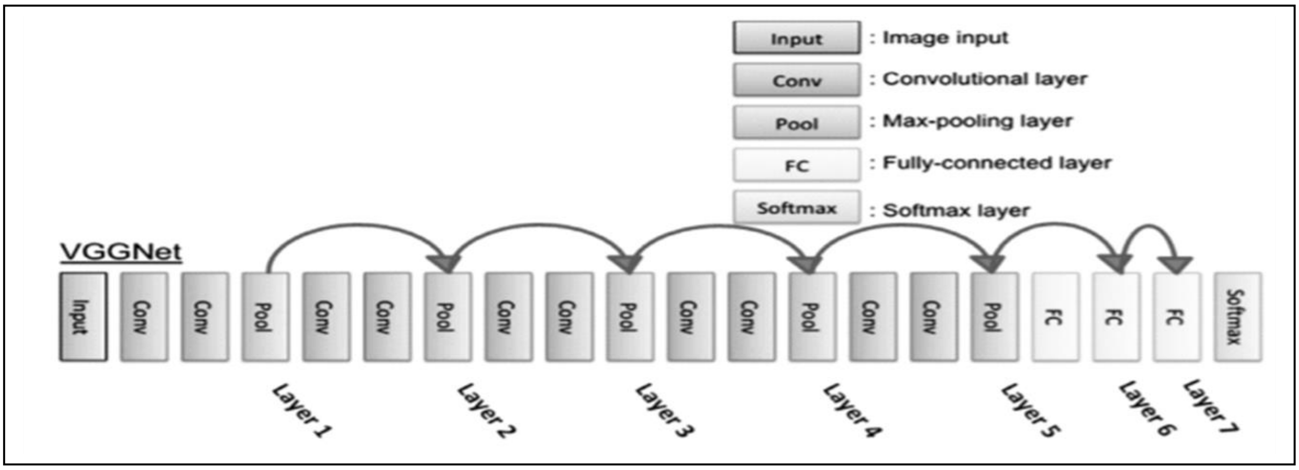

Figure 2 VGG Net

PROPOSED METHOD

The experiments are conducted on the GTZAN which is a widely used dataset for automatic music classification. The data were obtained from the following source:1. The dataset comprised 1000 30-s music files in an audio format at a sampling rate of 22050 Hz in 16-bit mono. Samples were classified into the following 10 musical genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock

Traditional machine learning approach need features to be extracted, but DNN models extract the features automatically from the music dataset.

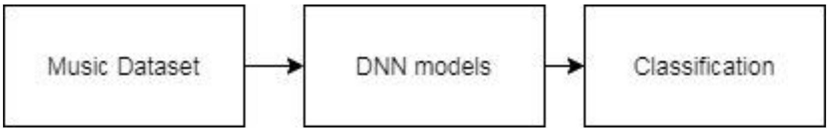

Figure 3 Proposed Method

TABLE 1 PERFORMANCE COMPARISON

Alexnet		VGG Net	
Accuracy (%)	**F-Score**	**Accuracy (%)**	**F-Score**
45%	0.6	78%	0.2

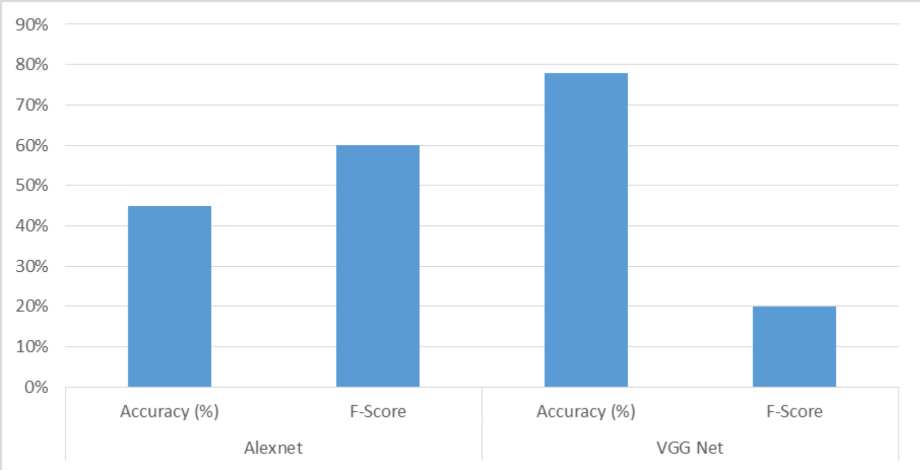

Figure 4 Performance of Alexnet and VGG net

CONCLUSION

This paper provides a analysis music data using pretrained deep neural networks in particular paying attention to latest developments. The pretThe paper critically reviews on well-established and proven methods for feature extraction and music indexing, from both the audio signal and contextual data sources about music items. These in turn permit an extensive variety of music retrieval tasks, such as semantic music search or music identification. This paper reviews current work on the important aspect of how various MIR approaches to adopted for genre classification of different types of music are evaluated and compared. Finally, a discussion about the major open challenges and future perspectives in music genre classification concludes the survey.

REFERENCES

[1] Mahardhika, F., Warnars, H. L. H. S., Heryadi, Y., & Lukas. (2019). "Indonesian's Dangdut Music Classification Based on Audio Features",In 1st 2018 Indonesian Association for Pattern Recognition International Conference, INAPR 2018 - Proceedings (pp. 99–103). Institute of Electrical and Electronics Engineers Inc.

[2] Mutiara, A. B., Refianti, R., & Mukarromah, N. R. A. (2016). Musical genre classification using support vector machines and audio features. Telkomnika (Telecommunication Computing Electronics and Control), 14(3), 1024–1034.

[3] Markov, K., & Matsui, T. (2012). Music genre classification using self-taught learning via sparse coding. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[4] Medhat, F., Chesmore, D., & Robinson, J. (2017). Automatic Classification of Music Genre Using Masked Conditional Neural Networks. 2017 IEEE International Conference on Data Mining (ICDM).

[5] Muhammad Asim Ali and Zain Ahmed Siddiqui, "AutomaticMusic Genres Classification using Machine Learning" International Journal of Advanced Computer Science and Applications(ijacsa), 8(8), 2017.

[6] Murauer, B., & Specht, G. (2018). Detecting Music Genre Using Extreme Gradient Boosting (pp. 1923–1927). Association for Computing Machinery (ACM).

[7] Pereira, R. M., Costa, Y. M. G., Aguiar, R. L., Britto, A. S., Oliveira, L. E. S., & Silla, C. N. (2019) "Representation Learning vs. Handcrafted Features for Music Genre Classification" , In Proceedings of the International Joint Conference on Neural Networks (Vol. 2019-July). Institute of Electrical and Electronics Engineers Inc.

[8] Panagakis, Y., & Kotropoulos, C. (2010). Music genre classification via Topology Preserving Non-Negative Tensor Factorization and sparse representations. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing.

[9] Prabhu, N. R., Andro-Vasko, J., Bein, D., & Bein, W. (2018). Music Genre Classification Using Data Mining and Machine Learning. Information Technology – New Generations, 397–403.

[10] Panwar, S., Das, A., Roopaei, M., & Rad, P. (2017). A deep learning approach for mapping music genres. In 2017 12th System of Systems Engineering Conference, SoSE 2017. Institute of Electrical and Electronics Engineers Inc.

[11] Quinto, R. J. M., Atienza, R. O., & Tiglao, N. M. C. (2017). "Jazz music sub-genre classification using deep learning", In IEEE Region 10 Annual International Conference, Proceedings/TENCON (Vol. 2017-December, pp. 3111–3116). Institute of Electrical and Electronics Engineers Inc.

[12] Rosner, A., Schuller, B., & Kostek, B. (2015). Classification of Music Genres Based on Music Separation into Harmonic and Drum Components. Archives of Acoustics, 39(4), 629–638.

[13] Senac, C., Pellegrini, T., Mouret, F., & Pinquier, J. (2017) "Music Feature Maps with Convolutional Neural Networks for Music Genre Classification".Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing - CBMI '17.

[14] Sarkar, R., & Saha, S. K. (2015). Music genre classification using EMD and pitch based feature. In ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition. Institute of Electrical and Electronics Engineers Inc.

[15] Sigtia, S., & Dixon, S. (2014). Improved music feature learning with deep neural networks. In ICASSP,IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings(pp. 6959–6963). Institute of Electrical and Electronics Engineers Inc.

[16] Song, Y., Zhang, C., & Xiang, S. (2007). Semi-Supervised Music Genre Classification. 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07.

[17] Simsekli, U. (2010). Automatic Music Genre Classification Using Bass Lines. 2010 20th International Conference on Pattern Recognition.

[18] Salamon, J., Rocha, B., & Gomez, E. (2012). Musical genre classification using melody features extracted from polyphonic music signals. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[19] Silla, C. N., Kaestner, C. A. A., & Koerich, A. L. (2007). Automatic music genre classification using ensemble of classifiers. In Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics (pp. 1687–1692).

[20] Song, G., Wang, Z., Han, F., & Ding, S. (2017). Transfer learning for music genre classification. In IFIP Advances in Information and Communication Technology (Vol. 510, pp. 183–190). Springer New York LLC.