Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 1): pp. 4932-4941
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.01.527

Music Genre Prediction Using Convolutional Recurrent Neural Network

[1]Lavanya S , [2]Taruniyaa S , [3]Balamurugan M

[1,2]Student, Department of Computer Science and Engineering Sri Sairam Engineering College

[3]Assistant Professor, Department of Computer Science and Engineering Sri Sairam Engineering College

ABSTRACT : Music is becoming increasingly easier to consume by way of apps on the internet and songs. Perhaps the most popular feature of music is the musical form. A complex task in the field is grouping music tracks according to their criteria for the structured arrangement of audio files and for the increasing interest in genre grouping in automated songs. Additionally, a critical aspect of the identification and aggregation of music in related genres is the recommended method for song and album generator. In this project, we adapt the transfer learning techniques to train a custom music genre classification system with customized genres and data. The model takes as an input the spectrogram/sonogram of music frames and analyzes the image using a Convolutional Neural Network (CNN) plus a Recurrent Neural Network (RNN). The output of the system is a vector of predicted genres providing maximum accuracy. This system will be useful to predict or analyze the mood or character based on music preferences which can help to cure depression, anxiety and stress.

Index terms: Transfer learning, Multi-Framing,CRNN

I.INTRODUCTION

Music plays a very important role in people's lives. Music brings like-minded people together and is theglue that holds communities together. Communities can be recognized by the type of songs that they compose, or even listen to. Different communities and groups listen to different kinds of music. One main feature that separates one kind of music from another is the genre of the music.

A genre is defined as a category of artistic composition characterized by similarities in form, style or subject matter. Two songs with the same genre usually have more similarities than two songs belonging to different genres. The classification based on genre can be done by extracting information that can be retrieved from the raw data.

The application of convolutional neural networks and convolutional recurrent neural networks for the task of music genre classification is discussed. We focus primarily on computational and data budget constraints where we cannot afford to train with large datasets. Transfer learning

techniques are applied to make the model adapt to the task of genre classification. Different strategies for fine-tuning, initializations and optimizers will be discussed to see how to obtain the model that fits better in the music genre classification. Moreover, we introduce a multi framing approach with an average stage in order to analyze in detail almost the entire song. It is used to generate more samples during the model training phase and at test time to achieve better accuracy . Finally, we evaluate its performance both in a handmade dataset and in the GTZAN dataset, used in a lot of works, in order to compare the performance of our approach with the state of the art.

II. LITERATURE REVIEW

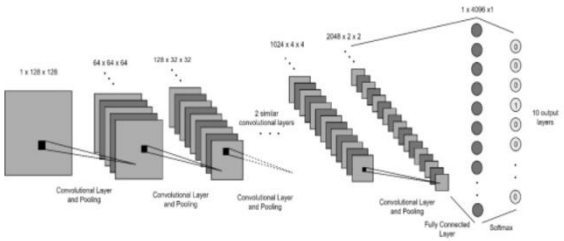
Music genre classification utilizing neural networks (NNs) has achieved some limited success in recent years. Differences in song libraries, machine learning techniques, input formats, and types of NNs implemented have all had varying levels of success. This article reviews some of the machine learning techniques utilized in this area. It also presents research work on music genre classification. The research uses images of spectrograms generated from time slices of songs as the input into an NN to classify the songs into their respective musical genres. With respect to musical genre classification,the research implementation involved taking songs, converting them into short-time segments and representing the time segments by their respective spectrogram images. Each of these spectrograms was labeled by music genre and then used as inputs into a CNN. The current implementation has only been trained using grayscale spectrogram images. Future work may include additional features embedded into a colored spectrum.Convolutional Neural Network algorithm is implemented and maximum accuracy is obtained.
With respect to musical genre classification,the research implementation involved taking songs, converting them into short-time segments and representing the time segments by their respective spectrogram images. Each of these spectrograms was labeled by music genre and then used as inputs into a CNN.

A complex task in the field is the grouping of music tracks according to their criteria. Additionally a critical aspect of the identification and aggregation of music in related genres is the recommended method for song and album generator. Unconsciously, music training reflects the user's moment. It is an important undertaking over the respective field to describe and analyse these moments. We explore the effect of machine learning techniques on the implementation of models of predictive genre classification aimed at capturing distinctions between genres. The analysis of genres as aggregated bodies of musical publications such a notion includes analogous inductive models as per arbitrary and local parameters. This process can thus be modeled as an example driven process for learning.Using machine learning techniques,on the implementation of models of predictive genre classification aimed at capturing distinction between genres. Discrete Cosine Transform (DCT) is required to discard noise. CNN proved to have the highest accuracy.

III. SYSTEM ANALYSIS

1) EXISTING SYSTEM

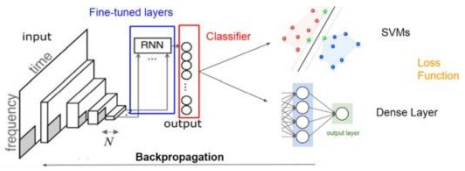
In the Existing System Convolutional neural networks (CNNs) have been actively used for various music classification tasks such as music tagging, genre classification, and user-item latent feature prediction for recommendation. CNNs assume features that are in different levels of hierarchy and can be extracted by convolutional kernels. The hierarchical features are learned to achieve a given task during supervised training. For example, learned features from a CNN that is trained for genre classification exhibit low-level features (e.g., onset) to high-level features (e.g., percussive instrument patterns



CNN ARCHITECTURE

2) PROPOSED SYSTEM

CNNs combined with recurrent neural networks (RNNs) are often used to model sequential data such as audio signals or word sequences. This hybrid model is called a convolutional recurrent neural network (CRNN). A CRNN can be described as a modified CNN by replacing the last convolutional layers with a RNN. CNNs and RNNs play the roles of feature extractor and temporal summariser, respectively.

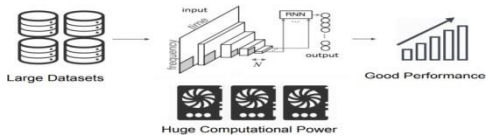


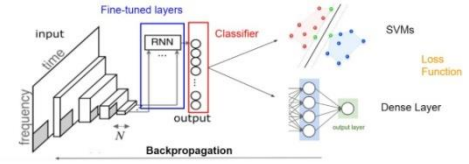
CRNN ARCHITECTURE

Transfer Learning (TL)

Transfer learning is the idea of overcoming the isolated learning paradigm and utilizing knowledge acquired from one task to solve related ones. The two most common practices that will be used are:

- Using the network as feature extractor
- Fine-tuning the network.





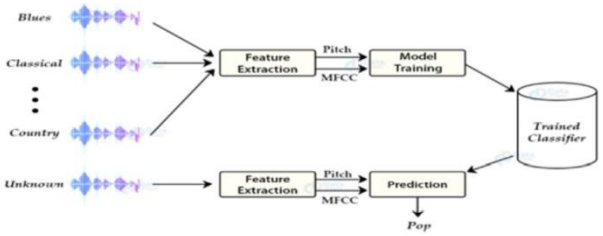
Change the classifier and also fine-tune the weights of the network

Multi-framing

Multi Framing strategy allows to extract more than one frame (or mel-spectrogram image) per song. For each song the first and last N seconds are discarded and then divided the rest into frames of equal time-length t. The final parameters are stated in the experiments. This approach has two advantages: At training time: This is a very useful tool to provide data augmentation. More data can begenerated for training the neural network. At test time: We can average or perform a KNN with the scores of every frame to infer the genre tag for the complete song with more confidence

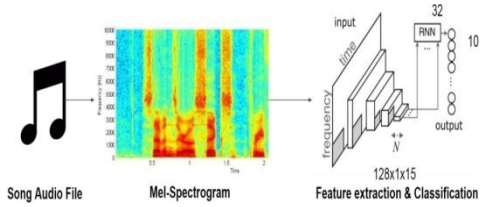
DATA FLOW DIAGRAM:

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. It differs from the flowchart as it shows the data flow instead of the control flow of the program. The DFD is designed to show how a system is divided into smaller portions and to highlight the flow of data between those parts. It concerns things like where the data will come from and go to as well as where it will be stored.The graphical depiction identifies each source of data and how it interacts with other data sources to reach a common output. This type of diagram helps business development and design teams visualize how data is processed and identify or improve certain aspects

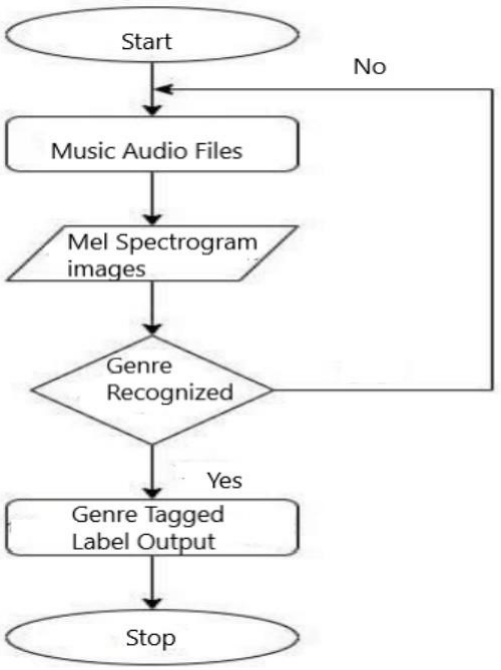


ARCHITECTURE DIAGRAM

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. The architecture diagram of a system in which the principal parts of functions are represented by blocks connected by the lines that show the relationships of the blocks. The architecture diagramis typically used for a higher level, less detailed description aimed more at understanding the overall concepts and less at understanding the details of implementation. It shows the relationship between different components of a system. System architecture can comprise system components, the externally visible properties of those components, the relationships between them. A system's architecture can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system.



WORKFLOW DIAGRAM



IV. METHODOLOGY

A Convolutional Recurrent Neural Network (CRNN) is trained to recognize music genres through transfer learning (TL) with maximum accuracy. It uses a 2-layer RNN with gated recurrent units (GRU) to summarize temporal patterns on the top of two- dimensional 4-layer CNNs. The underlying assumption in this model is that the temporal pattern can be aggregated better with RNNs then CNNs, while relying on CNNs on the input side for local feature extraction.

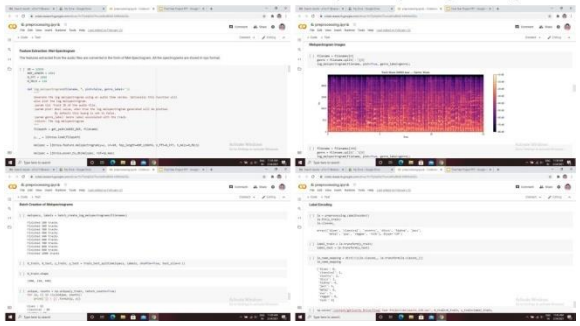
i.Data Collection:

Data Collection Data collection is the process of gathering and measuring information from countless different sources. GTZAN genre collection dataset consists of 1000 audio files each having 30 seconds duration. There are 10 tracks (10 music genres) each containing 100 audio tracks. Each track is in .wav format.

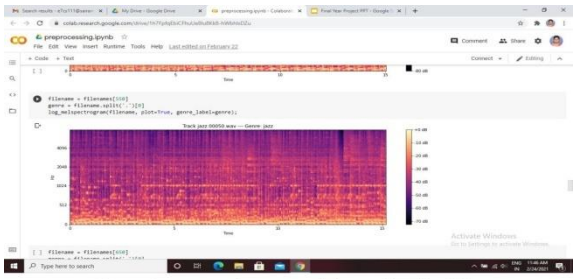
ii. Data Preprocessing

Data Preprocessing is the step in which the data gets transformed, or encoded, to bring it to such a state that now the machine can easily parse it. The features considered for music genre prediction are spectral centroid, zero crossing rate, chroma frequencies, spectral roll off, RMSE, spectral Bandwidth, Mel Frequency Cepstral Coefficient (MFCC).

The audio files obtained are converted into respective spectrogram images.



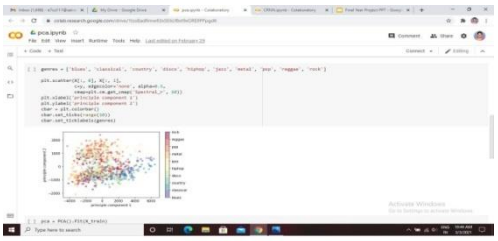
Data Preprocessing - Spectrogram image of Audio Files



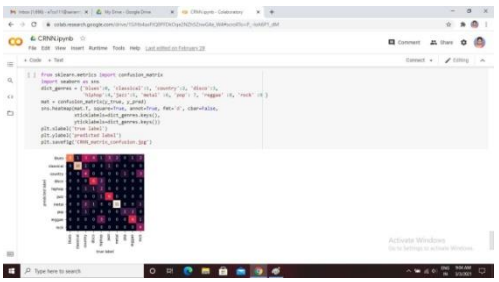
Spectrogram image of Jazz Audio File

iii. Feature Extraction

The features considered for music genre prediction are spectral centroid, zero crossing rate, chroma frequencies, spectral roll off, RMSE, spectral Bandwidth, Mel Frequency Cepstral Coefficient (MFCC)

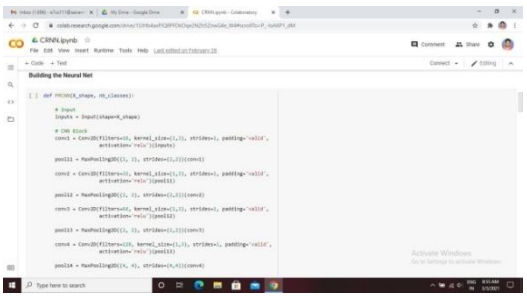


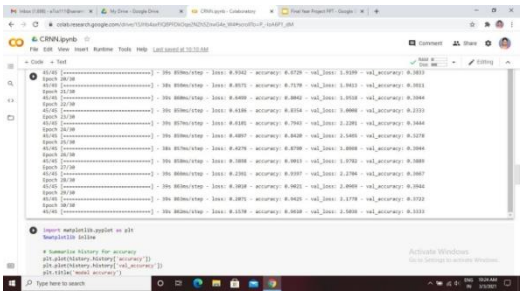
Principle Component Analysis



iv.Model Training

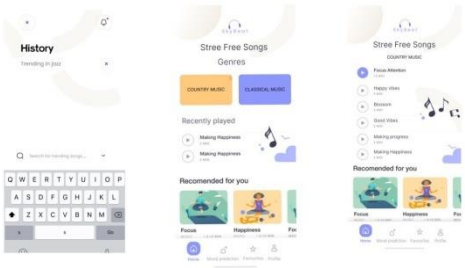
Training a model involves learning accurate values for all the weights and the bias from labeled examples. A Convolutional Recurrent Neural Network (CRNN) is trained to recognize music genres through transfer learning (TL) with maximum accuracy.





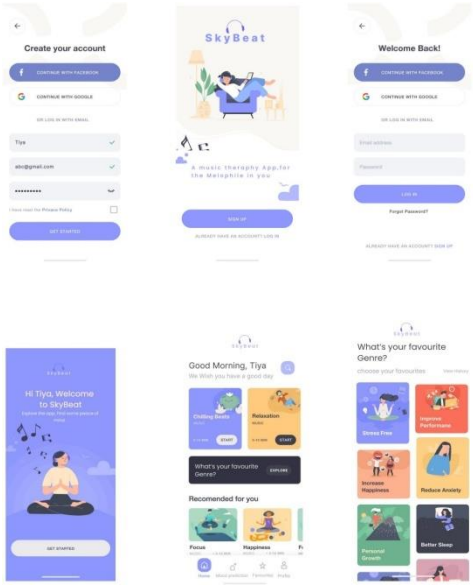
v. Testing and optimization

Testing is the process of adjusting hyperparameters in order to minimize the cost function by using the optimization technique to provide accurate predictions. Handmade datasets obtained through multi-framing are used to test the trained model and to obtain the desired results.



UI SCREENS

We have designed the UI screens for an app SKYBEAT.The app lets you browse songs on different genres.Based on the genre the user is listening to, the mood of the person can be predicted.This system will be useful to predict or analyze the mood or character based on music preferences.



V.RESULT

The model takes as an input the spectrogram/sonogram of music frames and analyzes the image using a Convolutional Neural Network (CNN) plus a Recurrent Neural Network (RNN). The output of the system is a vector of predicted genres providing maximum accuracy. This system will be useful to predict or analyze the mood or character based on music preferences which can help to cure depression, anxiety and stress. This can further be extended to produce personalized recommendations and organize music libraries.

VI.CONCLUSION

We explore the application of CNN and CRNN for the task of music genre classification focusing in the case of a low computational and data budget. The results have shown that this kind of network needs large quantities of data to be trained. In the scenario of having a small dataset and a taskto perform, transfer learning can be used to fine-tune models that have been trained on large datasets and for other different purposes. We have shown that our multiframe approach with an average stage improves the single-frame song model. In the experiments, a homemade dataset compounded by songs longer than our frame duration has been used. These songs belong to 10 different genres and the experiments have revealed that the average stage achieves better results in 9 of these 10 genres and a higher total accuracy. Therefore, using the average stage we are able to remove the non-representative frames dependency.

REFERENCES

[1] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," IEEE ransactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[2] S. Gollapudi, Practical Machine Learning. Birmingham, U.K.: Packt, 2016

[3] Keunwoo Choi, George Fazekas, and Mark Sandler, "Explaining deep convolutional neural networks on music classification," arXiv preprint arXiv:1607.02444, 2016

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning.(Report)," Nature, vol. 521, no. 7553, p. 436, May 2015, 2015.

[5] Matthew D Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," in European Conference on Computer Vision. Springer, 2014, pp. 818–833.

[6] O. Mogren, "C-RNN-GAN: Continuous recurrent neural networks with adversarial training," CoRR, vol. abs/1611.09904, Nov. 2016. Accessed: Jan. 12, 2019. Available:https://arxiv.org/abs/1611.09904

[7] M. Lopes, F. Gouyon, A. L. Koerich, and L. E. S. Oliveira, "Selection of training instances for music genre classification," in Proc. 20th Int. Conf. Pattern Recognit., Aug.
2010, pp. 4569–4572.