GUITAR TYPES CLASSIFICATION USING DEEP LEARNING

Valter A. Machado
University of the Cordilleras,
Baguio City, Philippines
valterandremachadodinho@gmail.com

Cedric Hortaleza
University of the Cordilleras,
Baguio City, Philippines
cedrichortaleza@gmail.com

Justice E. Kpodo
University of the Cordilleras,
Baguio City, Philippines
Justicekpodo2015@gmail.com

ABSTRACT

The Convolutional Neural Networks (CNNs) have been found very effective to learn patterns from audios and hence can be used for sound classification. In this paper, we first built a deep convolutional neural network architecture and adjust this neural network for guitar type recognition. To achieve the classification of the 3 main types of guitar, the guitars were categorized based on the sound each one of the them can produce and in order to accomplish that task Mel-Frequency Cepstral Coefficients (MFCCs) extraction technique was used. We also used data augmentation to deal with overfit issue and to improve the accuracy of the classification by splitting the audios into multiple audios using overlapping windows as there were few datasets available. The Convolutional Neural Network (CNN) along with all the techniques applied to solve certain issues encountered during this study were able to help and give a satisfactory result with overall classification accuracy of 90.2%. Convolutional Neural Network (CNN) best area of application is image classification but it can also be applied for sound, audio or speech classification tasks where it can give a decent maximum of 97% accuracy.

CCS CONCEPTS

• Artificial Intelligence • Machine Learning • Deep Learning • Digital Signal Processing

Keywords: Mel-Frequency Cepstral Coefficient (MFCC); Guitar; Deep Learning (DL); Machine Learning (ML), Fast Fourier Transform (FFT); Convolutional Neural Network (CNN).

1. INTRODUCTION

Guitar is one of the most popular musical instruments in the world, most of the people nowadays have a guitar but few of them really know the types of guitar that exist and its differences, guitars have 3 main types those are acoustic, electric and bass. Different guitars produce different sounds, the type of music would determine the type of guitar that will be suitable for it. Although musical instruments classification using Machine Learning (ML) techniques has been studied by many researchers already but it is noticeable that few research papers have dived into classification of a specific instrument.

This paper will take a closer look at how different types of guitar behave using Deep Learning (DL) in order to give people means and tools to distinguish different types of guitar, because it is known that non-musician people have difficulties to identify similar musical instrument sounds. This paper contributes to Music Information Retrieval (MIR) research field, and MIR have many applications where includes key detection, structural segmentation, music similarity measures, music transcription, music classification, playlist generation, and music recognition and other semantic analysis tasks. In this paper Machine Learning (ML) and Deep Learning (DL) techniques were explored along with Librosa library in order to achieve the expected output where includes Convolutional Neural Network (CNN) to classify the following 3 guitar types: acoustic, bass and electric.

The use of Mel Frequency Cepstral Coefficients (MFCC) was done to extract information from our data as prescribed by past work in this field.

2. RELATED WORKS

2.1 Chord Detection using Deep Learning

X. Zhou and A. Lerch (2017) utilized deep learning to learn high-level features for audio chord detection. They represented input audio into samples with sample rate of 11.056 kHz and then applied Constant Q Transform (CQT). They used Principal Component Analysis (PCA) for decorrelation, and applied Z-Score normalization. For pre-processing, time splicing followed by Convolution was done. Time splicing is a simple way to extend the current frame with the data of neighboring frames by concatenating the frames into one larger super frame. For training, a standard back propagation can be applied after pre-training to fine-tune the network in a supervised manner. The loss criterion used in this work is cross-entropy.

2.2 Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification

Salamon & Bello (2017) in their paper have proposed a deep convolutional neural network architecture for environmental sound classification. They have also proposed the use of audio data augmentation for overcoming the problem of data scarcity and explore the influence of different augmentations on the performance of the proposed CNN architecture.

2.3 Neural Networks for Musical Chords Recognition

Osmalskyj, et. al. (2012) in this paper have considered the challenging problem of music recognition and presented an effective machine learning based method using a feed-forward neural network for chord recognition. They have used the known

feature vector for automatic chord recognition called the Pitch Class Profile (PCP). Although the PCP vector only provides attributes corresponding to 12 semitone values, they have shown that it is adequate for chord recognition. Their experiments establish a twofold result: (1) the PCP is well suited for describing chords in a machine learning context, and (2) the algorithm is also capable to recognize chords played with other instruments, even unknown from the training phase.

3. METHODS

3.1 Dataset Preparation

Guitar types audio dataset were collected from MUSICRADAR website. The dataset consists of 3 classes (acoustic, electric and bass) with 200 audio samples for each class and all the samples are supplied as 24-bit WAV files. In the dataset the guitar types are represented by an identifier ranging respectively from 0 to 2.

Overall the dataset is composed of 600 samples and it has been divided into two sets. Training Set which consists of 99.8% of the whole dataset and Test Set which consists of 0.2%. The preprocessed dataset was saved into a NumPy array file of dimensions 11400 X 128 for further processing rather than the actual audio dataset.

Due to overfitting issue Early Stopping, Dropout and Data Augmentation techniques were used on preprocessing phase. The data was augmented by splitting the audios into multiple audios using overlapping windows.

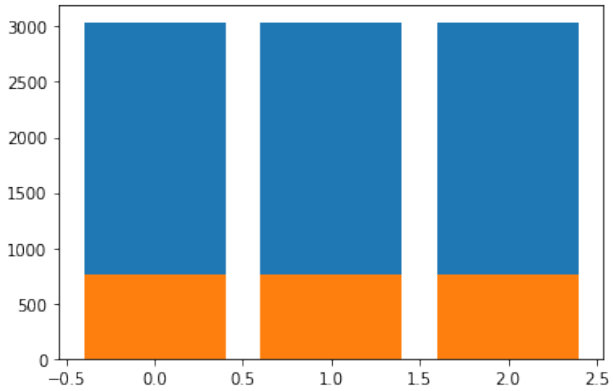

Figure 1. Representation of the guitar types (0-2) with its amount of train and validation samples.

3.2 Feature Extraction

Mel-Frequency Cepstral Coefficients were extracted from each audio file. Audio samples were processed in other to get an array of features and labels (the actual features but ranging from 0-2 where each number represents a class), every feature inside of the array is composed of 128 filter banks.

Feature Extraction Process:

- The audio were analyzed over short analysis window.
- For each short analysis window a spectrum is obtained using FFT.
- Spectrum is passed through Mel-Filters to obtain MelSpectrum.

These Mel-filters are non-uniformly spaced on the frequency axis, more filters in the low frequency regions and less number of filters in high frequency regions (similar to human ear).

The following figures (figure 2-6) are visualization of one of the audio sample from the electric guitar dataset where waveform, power spectrum, spectrogram, mffcs and mel filterbank techniques were applied for preprocessing task.

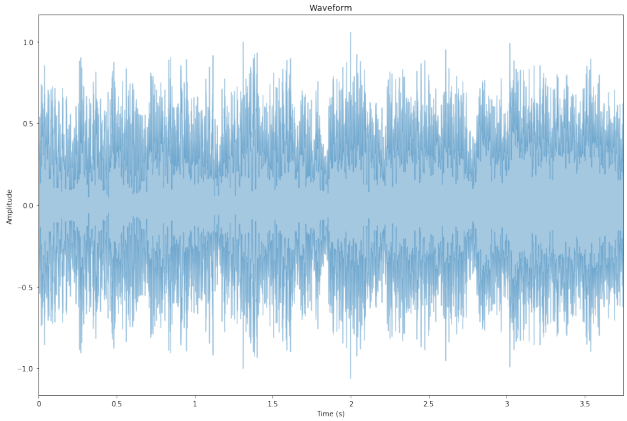

Figure 2. Waveform.

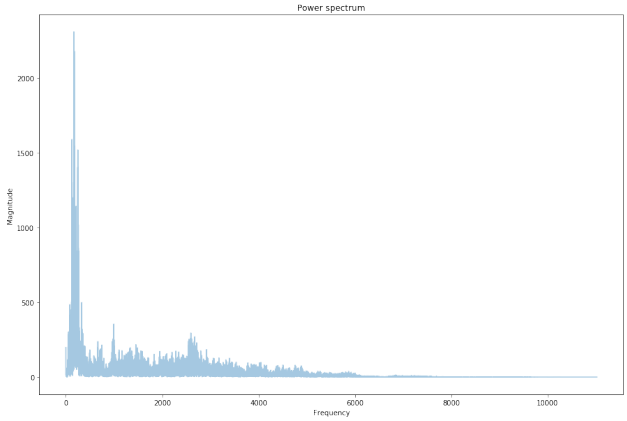

Figure 3. Power spectrum.

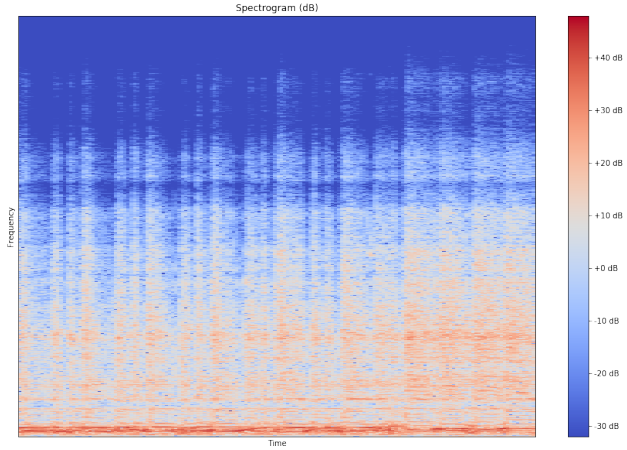

Figure 4. Spectrogram (dB).

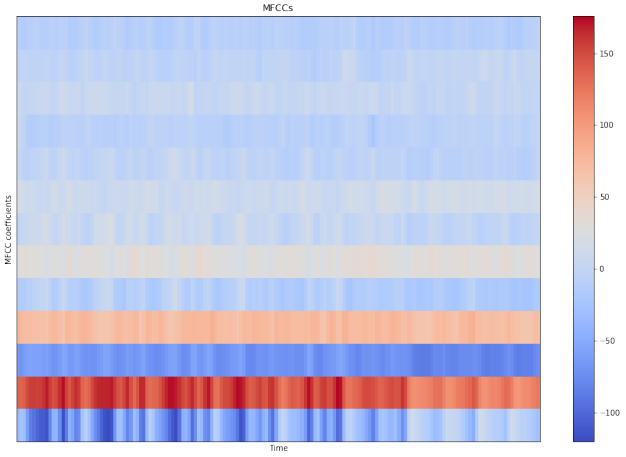

Figure 5. MFCCs.

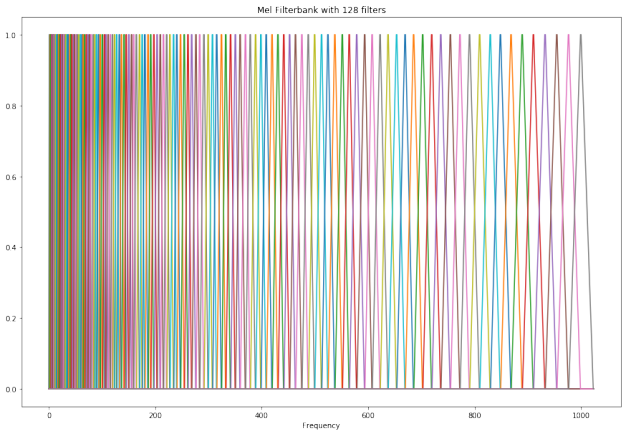

Figure 6. Mel filterbank.

3.3 Training and Validation

The model used for this research consists of a feed forward CNN. The final architecture of the model is a 4 layer CNN with 1 MaxPooling operation, a GlobalAveragePooling to supply the features to the Softmax function layer that is the output layer of the model which is fully connected with 3 filters (numbers of classes), and a Dropout layer. The number of filters are 64, 64, 128, and 128 in increasing order from the shallow to the deep layers of our network. And the MaxPooling node have a kernel size of 3. Model summary shown in figure 7.

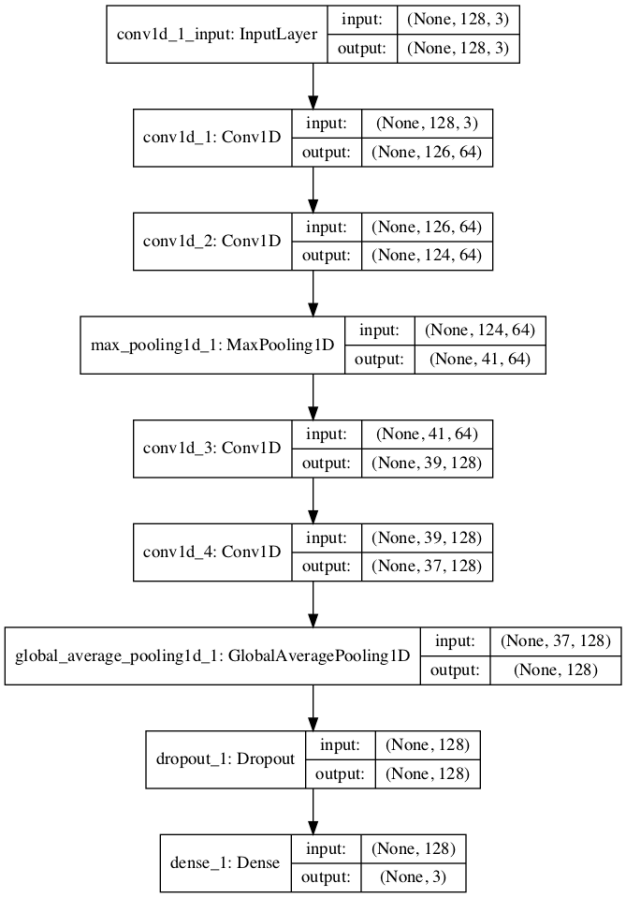

Figure 7. Model architecture.

The model's architecture has 87,363 trainable parameters as shown in figure 8. A Softmax activation function was implemented in the output layer. The model was compiled with a Categorical Crossentropy loss function in order to handle the multi-class classification task.

Layer (type)	Output Shape	Param #
conv1d_41 (Conv1D)	(None, 126, 64)	640
conv1d_42 (Conv1D)	(None, 124, 64)	12352
max_pooling1d_11 (MaxPooling	(None, 41, 64)	0
conv1d_43 (Conv1D)	(None, 39, 128)	24704
conv1d_44 (Conv1D)	(None, 37, 128)	49280
global_average_pooling1d_11	(None, 128)	0
dropout_11 (Dropout)	(None, 128)	0
dense_11 (Dense)	(None, 3)	387

Total params: 87,363
Trainable params: 87,363
Non-trainable params: 0

Figure 8. The total number of trainable parameters.

4. RESULTS AND DISCUSSION

9,120 samples were trained, with batches of size 128, 60 epochs. Figure 9 presents the loss over the epochs for training and validation. An early stop was forced resulting on a total of 60 epochs after which was confirmed the model's trend to slowly start to overfit the training data. To verify that the model was not having overfit the evaluation metrics was calculated for the training data and compared them with the results for the validation dataset.

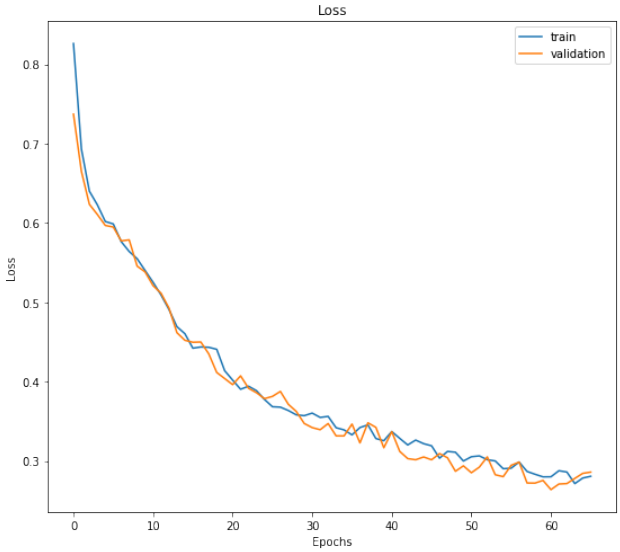

Figure 9. Training and Validation Loss over the Epochs.

Confusion matrix was computed for the test dataset, which is shown in Table 1. It can be observed that the biggest errors are between the classes "Bass" and "Electric". The CNN has achieved an overall classification accuracy of 90.2% on the dataset. Confusion matrix summary shown in table 1.

Actual \ Predicted	Acoustic	Bass	Electric
Acoustic	647	24	31
Bass	78	727	45
Electric	35	9	684

Table 1. Confusion Matrix for the 3 Classes.

Precision, recall and metrics F1-score were computed for the datasets. Table 2 shows the metrics for the test dataset. The CNN model attained an average of 90% for precision, recall, and F1-score.

	Acoustic	Bass	Electric	Total (%)
Precision	0.92	0.86	0.94	90.6
Recall	0.85	0.96	0.90	90.3
F1-score	0.89	0.90	0.92	90.3

Table 2. Precision, Recall and F1-score for the 3 classes and Total Data.

5. CONCLUSION

In this paper we presented a study on multi-class machine learning classification for the problem of guitar types classification. A deep convolutional neural network architecture was developed in which guitar types dataset in combination with a set of audio data augmentations extracted from the existing dataset produces the desired prediction for the guitar types classifier. The CNN model was able to achieve a precision of 90.6%, a recall of 90.3% and a F1-score of 90.3% for the multi-class classification task. We came to a conclusion that data augmentation helped to fight the overfit issue and make a great improvement in the classifier accuracy as well.

In the end of our study a solution on the guitar types classification problem was created. However, it could be further extended out in a few ways. For instance, within these 3 main types of guitar there are also different variations, especially on electric and bass type of guitar, that variation within every single guitar type affects somehow the prediction of our CNN model and we believe we can aim to fill out those gaps on a future study.

6. ACKNOWLEDGMENTS

The researchers would like to thank Rey Benjamin M. Baquirin, for giving us the opportunity to explore a part of A.I world with this project study.

7. REFERENCES

[1] S. Essid, G. Richard and B. David, "Musical instrument recognition by pairwise classification strategies", Audio Speech and Language Processing IEEE Transactions on, vol. 14, July 2006.

[2] A. Eronen and A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features", Acoustics Speech and Signal Processing 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on, vol. 2, pp. II753-11756, 2000.

[3] K. D. Martin and Y. E. Kim, "Musical instrument identification: A pattern recognition approach", The Journal of the Acoustical Society of America, vol. 104.

[4] Lewis Guignard & Greg Kehoe (2015), Learning Instrument Identification.

[5] Araju Nepal & Manasi Kattel & Ayush Kumar Shah & Deepesh Shrestha (2020), Guitar Chord Recognition.

[6] J. Osmalskyj, J-J. Embrechts, S. Piérard, M. Van Droogenbroeck (2012), Neural Networks for Musical Chords Recognition.

[7] J. C. Brown, "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features", The Journal of the Acoustical Society of America, vol. 105, no. 3, pp. 1933-1941, 1999.

[8] D. Johnson and G. Tzanetakis, "Guitar model recognition from single instrument audio recordings," 2015 IEEE

Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), Victoria, BC, 2015.

[9] J. Marques and P. J. Moreno, "A study of musical instrument classification using gaussian mixture models and support vector machines", *Tech. Rep.*, 1999

[10] Kailash Patil, Mounya Elhilali. (2015) Biomimetic spectro-temporal features for music instrument recognition in isolated notes and solo phrases. EURASIP Journal on Audio, Speech, and Music Processing 2015:1.

[11] Farbod Hosseyndoust Foomany, Karthikeyan Umapathy. (2013) Classification of music instruments using wavelet-based time-scale features. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)

[12] M. Erdal Özbek, Nalan Özkurt, F. Acar Savacı. (2012) Wavelet ridges for musical instrument classification. Journal of Intelligent Information Systems.

[13] Jayme Barbedo. 2011. Instrument Recognition. Music Data Mining, pages 95-134.

[14] S. Essid, G. Richard, B. David. (2006) Musical instrument recognition by pairwise classification strategies. IEEE Transactions on Audio, Speech and Language Processing.

[15] M. Erdal Ozbek, F. Acar Savaci. (2007) Music Instrument Classification Using Generalized Gaussian Density Modeling. 2007 IEEE 15th Signal Processing and Communications Applications.