An Enhanced Musical Instrument Classification using Deep Convolutional Neural Network

S. Prabavathy, V. Rathikarani, P. Dhanalakshmi

Abstract: ***Retrieval of musical information from musical databases is a major challenging issue in a digital world. Therefore, it is necessary to develop an efficient tool for retrieving the musical information. Musical instrument classification plays a major role for retrieving the information from musical database. In order to retrieve the musical instrument efficiently, an enhanced musical instrument classification algorithm using deep Convolutional Neural Network is proposed in this paper. The proposed algorithm consists of convolutional layers interleaved with two pooling functions followed by two fully interconnected layers. There are sixteen instruments from different instrument families are taken for evaluating the performance of proposed algorithm. The experimental result shows that the proposed algorithm recognizes the instruments significantly and achieves the greater accuracy than existing algorithm.***

Keywords: ***Deep Convolutional Neural Network, Musical Instrument Classification.***

I. INTRODUCTION

Musical Instrument Classification (MIC) is one of the major critical tasks for acquiring the high level information about the musical signal. The musical notes are classified into monophonic and polyphonic. In monophonic, a single instrument is played and in polyphonic two are more instruments are played concurrently [1]. The instrument classification in monophonic notes is practically successful but polyphonic is harder to recognize the annotation of musical signal. The challenges in MIC are the variance in timbre and performance style of musical signal is combined with the perceptual similarity of some other musical instrument and superposition of multiple instruments in frequency and time [2]. Another major issue in classifying the musical signal is choosing the best feature set for given classification algorithm. Hence, Deep Learning significantly improves the method for extracting the features from the raw musical signals and provides remarkable achievement for recognizing the musical signal.

The automatic musical classification plays a major role for solving the problem of structured coding, automatic musical instrument annotation and retrieval musical instrument from database system [3]. In this paper, a new approach for recognizing the musical instrument using deep learning with integrated feature extraction algorithm is proposed. The following section reviews literature of musical instrument classification and other its related tasks. Section III describes the architecture of proposed algorithm with two different feature extraction methods. Next, Section IV describes experimental results and the evaluation metrics finally the paper is concluded in section V with future enhancement.

II. LITERATURE REVIEW

This section briefly discusses some well-known algorithm for classification the musical instrument from musical database. The proposed algorithm [4] uses the isolated notes for classification musical instrument. The Time Encoded Signal Processing produces the simple matrices for encoding the notes and it derived from complex sound wave. The encode signals are simple and light weight in computational method. The resultant matrices are given as a input of Fast Artificial Neural Network (FANN) to recognize the musical instrument with perform instrument classification with efficient results and it reduce the computational cost compared than existing system.

The classification of musical instrument using FFNN Classifier [5] combines both the features of temporal and spectral of musical instrument. It is implementing by two stages. In a first stage, extract the spectral features from musical signal that are used to recognize the instrument using various frequency estimation approaches. In a second stage, a Feed Forward Neural Network has been used for classifying the signals. The proposed work supports the Single Instrument with Single Note, Single Instrument with Multiple Note and Multiple Instrument with Multiple Note.

Indian Musical Instruments are recognized in [6] for classifying the sounds that are taken from the natural environment. The Feature extraction methods such as zero crossing, RMSE and Mel-Frequency Cepstral Coefficient are used for extracting the features. In the preprocessing stage it removes the unnecessary noise and extract only the raw sound of given samples.A novel approach for recognizing polyphonic audio signal is proposed in [7]. It uses source filter method and matrix non-negative factorization algorithm for separating the sounds.

Manuscript published on November 30, 2019.
* Correspondence Author

S. Prabavathy*, Department of Computer and Information Science, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India.
V. Rathikarani, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India.
P. Dhanalakshmi, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamilnadu, India.

It uses MFCC and GMM method for extracting the features based on the density of sound. It uses 16 different instruments that are generated randomly to evaluate the polyphonic signals. The proposed work uses GMM as a classifier and the instance of the instruments are randomized into 70% for training and 30% for testing. It classifies the 16 different instruments efficiently with complex signals.The proposed algorithm [8] uses MLP and K-Nearest Neighbors for classifying musical instrument.

The sample music is a combination of Trumpet, Percussion, Drums, Piano, Double bass, Guitar which are played simultaneously. It extracts four different features by using feature extraction algorithms. One feature in all instance are calculated using UTA algorithm. The accuracy of the proposed algorithm is compared with the existing algorithms MLP and K-Nearest Neighbors.The majority of the classification systems used so far concentrate on the timbral-spectral characteristics of the notes. Discrimination is based on features such as pitch, spectral centroid, energy ratios, spectral envelopes and mel frequency cepstral coefficients [3, 4].Temporal features, other than attack, duration and tremolo, are seldom taken into account. Classification is done using k-NN classifiers, HMM, Kohonen SOM and Neural Networks [5, 6]. A limitation of such methods is that in real instruments the spectral features of the sound are never constant. Even when the same note is being played, the spectral components change. One has to take into consideration many timbral components and the way they can vary, which is often rather random, in order to develop a robust classification system.

III. PROPOSED ALGORITHM

The proposed algorithm consists of number of layers that are arranged in a stacked manner in a deep network architecture. The layers are: Input layer, 3 convolutional layers, 3 pooling layers, 2 fully connected layers and finally an output layer. The spectrogram is given as input with the size of 64x80. The kernel size of the input layer is 5x5. It passes through the max pooling layer with 3x3 kernel size. The layers in proposed algorithm are defined as follows:

Layer 1: 80 filters with receptive field of (5x5) followed by max pooling and rectilinear activation function, h(x)=max(x,0);

Layer 2: 120 filters with receptive field of (5x5) followed by max pooling and rectilinear activation function

Layer 3: 120 filters with receptive field of (5x5) with no pooling

Layer 4: 80 Hidden layers with 256 units

Layer 5: Output layer with 18 units followed by softmax activation function

The architecture of proposed algorithm is shown in figure 1. In the proposed algorithm the receptive field is considered in 5x5 block. Each hidden unit instead of connecting to all the inputs from previous layer is limited to processing only a tiny part of the whole input space called its receptive field. The operation of Max pooling is:

- The dimensionality reduction can be achieved through pooling layers. It merge the adjacent cells of feature map
- The common pooling operation is choosing the max or mean of input cells
- This process is called downsampling

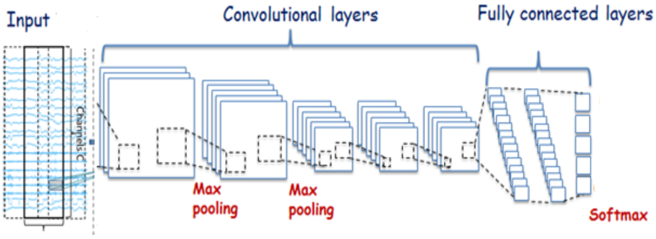


Figure 1. Architecture of proposed algorithm

The number of estimated value is bigger than the number of training at most of the time. Thus the result is poor out-of-sample generalization In order to handle this problem dropout learning is introduced in each training iteration, every hidden unit is randomly removed with predefined probability and the training procedure is continued normally. The music signals are preprocessed before fed into the feature extraction block. Normalize the music by scaling the sampled music data file should fall within the range of -1 to 1.

Feature Selection

Feature extraction utilizes the short-time processing approach, where the analysis is done periodically in short-time segments referred to as analysis frames, to capture the signal in quasi-stationary state [9]. In frame blocking the audio signal is sliced into fixed length analysis frames, shifted with a fixed time step. Typical analysis frame sizes are between 20 and 60 ms, and the frame shift is typically selected so that the consecutive frames are overlapping at least 50%. The analysis frames are smoothed with a windowing function to avoid abrupt changes at the frame boundaries that can cause distortions in the spectrum. The windowed frame is then transformed into spectrum for further feature extraction.

Mel-band energies and MFCCs provide a compact and smooth representation of the local spectrum, but neglect temporal changes in the spectrum over time, which are also required for the classification of instruments [10]. Temporal information can be included by using delta features, which represent the local slope of the extracted feature values within a predefined time window. Another way to capture the temporal aspect of the features is to stack feature vectors of neighboring frames into a new feature vector. Therefore it combined with the generative classifiers include Gaussian mixture models (GMM). In GMM, the aim is to model the joint distribution $p(x; y)$ for each class separately and then use Bayes' rule to find maximum posterior $p(y/x)$, i.e., from which class a given input x is most likely to be generated [11].

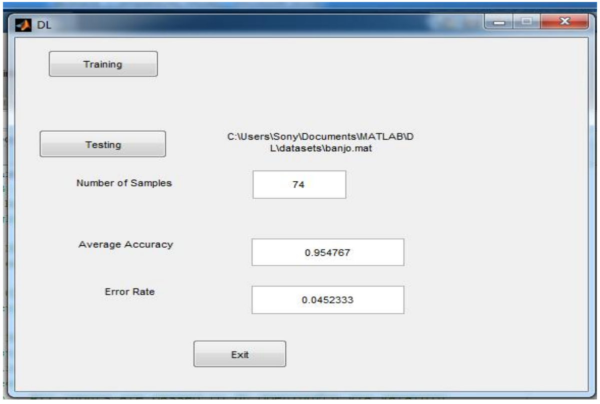
IV. RESULTS AND DISCUSSION

The signals are sampled at 48.2 KHZ. Down mixed to mono and normalized by mean squared energy. Each track is chunked to 1s-2s long snippets.

Each snippet is transformed into spectrogram In training process, leaky rectified linear units is used as loss function and Stochastic gradient descent optimizer is used to optimize the loss function.

Learning rate is initialized to 0.0001. Momentum is 0.8 and each input is trained with 500 epochs. In testing process, dropout is used with the probability of 0.5 for layer 1 and 2 and 0.75 for layer 3. The proposed algorithm is implemented by using MATLAB R2014a.

There are 16 instruments with different number of samples are considered for experiment. The samples are trained and tested successfully. The figure 2 shows the testing phase of the proposed algorithm.



C:\Users\Sony\Documents\MATLAB\D L\datasets\banjo.mat

Number of Samples: 74

Average Accuracy: 0.954767

Error Rate: 0.0452333

Figure 2. Testing phase of proposed algorithm

The classification accuracy is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives. The average accuracy of each instrument is calculated based on the result of confusion matrix. The figure 3 describes the average accuracy of each samples. The accuracy of most of the musical instrument is more than 95%.

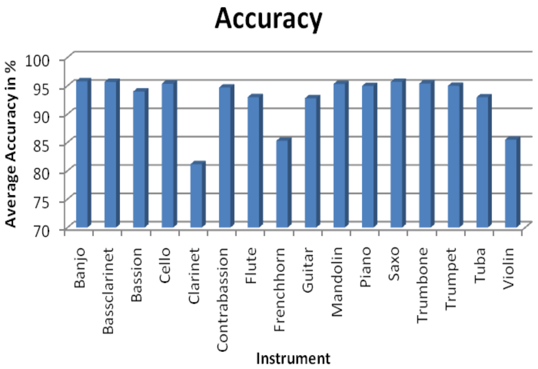


Figure 3. Classification accuracy for different instrument

V. CONCLUSION

This paper presents an enhanced deep convolutional neural network for recognizing musical instrument from large musical database. MFCC feature extraction algorithm is used to extract the features from the sample musical instruments. The extracted features are trained for predicting the instrument exactly. The proposed algorithm classifies the instrument according to the extracted features and classified correctly. The result shows that the proposed algorithm classifies the musical instrument with the accuracy of 97.5%. The overall performance shows that the proposed algorithm achieves the better classification accuracy rate than existing algorithms.

REFERENCES

1. P. Herrera, G. Peeters and S. Dubnov, "Automatic Classification of Musical Instrument Sounds", Journal of New Music Research, Vol. 32, 2003.
2. M. J. Newton and L. S. Smith, "A neurally inspired musical instrument classification system based upon the sound onset", The Journal of Acoustic Society of America, 131(6):4785-98. doi: 10.1121/1.4707535, June 2012.
3. S. Muhury, G. Neogi, P. Debnath and J. Ghosh Dastidar, "Design of a voice-based system by recognizing speech using MFCC", Computational Science and Engineering Proceedings of the International Conference on Computational Science and Engineering (ICCSE2016), CRC Press , pp 77–80, DOI: 10.1201/9781315375021-16, October 2016.
4. Giorgos Mazarakis, Panagiotis Tzevelekos, and Georgios Kouroupetroglou, "Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks ",SETN 2006, pp. 246 – 255, 2006.
5. V. S. Shelar, D. G. Bhalke, "Musical Instrument Recognition and Transcription using Neural Network", International Journal of Computer Applications (0975 – 8887) Proceedings on Emerging Trends in Electronics and Telecommunication Engineering (NCET 2013) 31-36
6. Sankaye, S.R. , Tandon U.S., "Indian Musical Instrument Recognition based MFCC Feature Set", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 01-07
7. Toni Heittola, Anssi Klapuri and Tuomas Virtanen, "Musical Instrument Recognition in Polyphonic Audio Using source filter model for sound separation", International Society for Music Information Retrieval, 2009
8. Akram Azarloo, Fardad Farokhi, "Automatic Musical Instrument Recognition Using K-NN and MLP Neural Networks", Fourth International Conference on Computational Intelligence, Communication Systems and Networks, 2012
9. C. Weihs, U. Ligges, F. Morchen, and D. Mullensiefen, "Classification in music research," Adv. Data Anal. Classification., vol. 1, no. 3, pp. 255–291, 2007.
10. Xie, C., Cao, X., & He, L. "Algorithm of Abnormal Audio Recognition Based on Improved MFCC", International Workshop on Information and Electronics Engineering (IWIEE), China: Elsevier. pp. 731-737, 2009.
11. S. Essid, G. Richard, and B. David,] "Musical instrument recognition by pairwise classification strategies," IEEE Trans. Audio, Speech, Lang.Process., vol. 14, no. 4, pp. 1401–1412, 2006.

AUTHORS PROFILE

S. Prabavathy, M.C.A, Ph.D (CS). She is currently pursuing Ph.D in Annamalai University, Chidambaram, Tamilnadu, India. Her research interests are machine learning and deep learning.

V. Rathikarani, B.E., M.E., Ph.D (CSE). Her interests are Computer Networks, Pattern Classification, Medical Image Processing. She is Assistant Professor in Annamalai University, Chidambaram, Tamil Nadu, India. She has 13 years' experience in teaching.

P. Dhanalakshmi, B.E., M.Tech., M.B.A., Ph.D. Her interests are Audio and Speech Signal Processing, Pattern Recognition, Data Mining and Software Engineering. She is Professor in Annamalai University, Chidambaram, Tamil Nadu, India. She has 21 years' experience in teaching.