Reading Summaries

Reading 1: Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks

url:https://www.researchgate.net/profile/Georgios_Kouroupetroglou/publication/221238999_Musical_Instrument_Recognition_and_Classification_Using_Time_Encoded_Signal_Processing_and_Fast_Artificial_Neural_Networks/links/0fcfd505729f306d99000000/Musical-Instrument-Recognition-and-Classification-Using-Time-Encoded-Signal-Processing-and-Fast-Artificial-Neural-Networks.pdf

Musical instrument recognition systems are traditionally use frequency domain analysis and shape analysis to extract a set of relevant audio features. Often times the models utilize k-NN classifiers, HMM, Kohonen SOM and Neural Networks. In paper describes an instrument recognition system that analyzes isolated notes using a Time Encoded Signal Processing method to produce matrices from complex sound waveforms. These matrices are used for instrument note encoding and recognition by being processed in a Fast Artificial Neural Network (FANN).

The dataset used in the project described uses both synthesized and recorded instruments. The synthesized samples come from 5 different synthesizers, and is used for training, while real instrument recordings where obtained from the from Iowa University for testing. These samples where created and collected based on 19 distinct types of instruments.

Time Encoded Signal Processing and Recognition (TESPAR) Coding is a method used to digitally code speech waveforms. The process uses infinite clipping to preserve the real zeros of the waveform from the waveform itself and not at the sampling frequency. In the paper, the sample music waveforms are presented to the software using TESPAR coder in Matlab. Since this was originally a method used for speech processing, which is usually between 300Hz to 3 kHz, the bandwidth of the signal encoding is extended to 100Hz to 5.5 kHz for musical representation. The coding process results to a symbol stream which can be converted into a fixed-dimension matrix called S-Matrix which is then used for classification purpose.

The fixed dimensions of the TESPAR coding make the classification task using Artificial Neural Networks (ANN) a good fit. The paper utilizes FANN which is a library that can be used to implements a multilayer feedforward ANN. The paper highlights that the advantages of this implementation includes faster training and testing, compared to similar libraries on systems without a floating point processor while maintaining similar performance to other libraries on systems with a floating point processor. The implementation is split into a two step process: train and test. The training process is optimized through the Backpropagation algorithm which will minimize the MSE for all the training data.

Overall, the method proved to provide relatively high recognition rates with notes produced from synthesizers and notes from real-instrument recordings with results varying based on which synthesizer the audio samples came from. The performance ranged from 32-99% depending on instrument and synthesizer. This result is significant taking into account that random selection would have resulted in an accuracy of 5.26%.

Reading 2: A Single Predominant Instrument Recognition of Polyphonic Music Using CNN-based Timbre Analysis

Url:https://www.researchgate.net/publication/328073709_A_Single_Predominant_Instrument_Recognition_of_Polyphonic_Music_Using_CNN-based_Timbre_Analysis/fulltext/5cc310f9a6fdcc1d49b20897/A-Single-Predominant-Instrument-Recognition-of-Polyphonic-Music-Using-CNN-based-Timbre-Analysis.pdf

This article tackles the challenge of classifying musical instrument from polyphonic music. While spectrogram analysis, especially when used in audio processing, uses Short Time Fourier Transform (STFT) and Mel Frequency Cepstral Coefficient (MFCC), this project uses the addition of Hilbert Spectral Analysis (HSA) in its preprocessing processes. Using a modified convolutional neural networks (CNN) the article states that the model has had and 3% performance improvement in individual instrument recognition using the IRMAS dataset compared to state-of-the-art techniques.

The proposed approach of this article used the timbre of the musical signal in order to recognize the predominant instrument. The breakdown of the method proceeds as follows: audio signal preprocessing, image preprocessing, network learning. Input audio signal is a stereo signal which is converted to mono signal using root-mean-square (RMS). CNN is used to automatically analyzes the image and finds the relevant features. Then, Hilbert Spectrum Analysis – Intrinsic Mode Functions HSA-IMF is applied according to the designated time window to generate the image. The HSA-IMF is commonly used in EEG analysis and can measure both instantaneous amplitude and phase. A CNN receives the image as input after histogram equalization is applied to pre-processed image to reduce the image size and achieve better performance in image classification step.

The CNN used is called LeNet. During the feature extraction process of the CNN, the convolutional layer extracts input features using convolution filters and activation functions which process filter outputs to non-linear values. Afterwards, a sub-sampling process leaves the necessary features in the pooling layer. These steps reduce the size of data and computing resources which can preventing overfitting.

By training and testing the system on the IRMAS dataset. The proposed approach was found to be 80% accurate, which is a 3% improvement on the 77% accuracy provided by traditional approaches using STFT and MFCC

The paper summarizes that audio signal processing is difficult to learn through NN because it uses the input spectrum which has contains both time and frequency information. Therefore, the paper suggests that, in order to analyze the frequency easily, it is possible to obtain higher performance by applying the HSA-IMF method.