NIET Journal of Engineering & Technology

Volume 10, Issue 1, 2019 ISSN: 2229-5828 (Print)

STUDY OF SOUND CLASSIFICATION USING DEEP LEARNING

Shalbin Benny , Arvind Kumar Tripathi , Rishabh Malik , Naman Mittal

Department of Computer science and engineering, Noida Institute of Engineering and Technology,
Greater, Noida 201306, India.

Abstract: ***Sound plays a crucial part in every element of human life. Sound is a crucial component in the development of automated systems in a variety of domains, from personal security to essential monitoring. There are a few systems on the market now, but their efficiency is a worry for their use in real-world circumstances. Image classification and feature classification are the same as sound classification, just like other classification algorithms like machine learning. We also construct CNN architecture here. Deep learning architectures' learning capabilities can be leveraged to construct sound categorization systems that overcome the inefficiency of standard methods. The goal of this paper is to use deep learning networks to classify environmental sounds based on the spectrograms that are created. The convolutional neural network (CNN) was trained using spectrogram images of environmental noises. For investigation, this paper used one dataset: Urbansound8K.There are 8732 sound clips (<=4s) of urban noises from ten classes in the collection. On this dataset, system was trained, and the accuracy acquired during training and testing was 98 percent and 91.92 percent respectively. The proposed approach for sound classification using spectrogram images of sounds can be efficiently employed to construct sound classification and recognition systems. Which can be used to distinguish audio evidence during crime investigation, remove noises and other useless sounds from music recording, or to classify different animals sounds in the forest.***

Keywords- Convolutional Neural Network, Deep Learning, Mel-Frequency Cepstral Coefficients, Spectrograms, Urbansound8K.

1. Introduction

Automatic sound recognition research has gained traction in recent years, with applications in domains as diverse as multimedia [1], bioacoustics monitoring [2], intrusion detection in wildlife regions [3], audio surveillance [4], and environmental noises [5]. This can be used to:

- Distinguish audio from audio evidence during crime investigation.
- Sound Production companies can use this project to remove noises and other useless sounds from their recording.
- Musicians can use this project to remove unwanted audio from live concerts.

The challenge of sound recognition is divided into three stages: signal pre-processing, extraction of specific features, and classification. Signal pre-processing separates the input signal into segments that are then utilized to extract related features. Feature extraction is a technique for reducing the quantity of data and representing it as feature vectors. Various classifiers such as decision trees, random forest, and k nearest neighbor were used to classify crossing rate, pitch, and frame features used in speech recognition applications. The terms Spectrogram image features (SIF), Stabilized auditory image (SAI), and Linear prediction coefficients (LPC) are all used to describe the features of a spectrogram image. In recent years, SIF has begun to generate sound waves, resulting in more accurate results in noisy environments. High-pressure and low-pressure zones move through a medium to create sound waves. Every distinguishable sound has a distinct pattern formed by such high- and low-pressure zones. Wavelength, frequency, wave speed, and time periods are some of the features of these waves [6]. These characteristics are used to categorize the sounds in the same way that humans do. A spectrogram, as shown in Fig. 1, is a visual representation of the frequency spectrum of a sound wave. In simple terms, it is a snapshot of the sound wave's frequency spectrum [7]. Because the sound signal's generated spectrogram is rare, noise intensity is found in the bottom part and strong components are discovered in the upper region. The spectrogram images created can be combined with a variety of machine learning classifiers. This paper will show how to use Deep Learning algorithms to classify environmental noises, with a particular emphasis on the identification of specific urban sounds. We want to be able to detect if an audio sample of a

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

few seconds duration in a computer readable format (such as a.wav file) contains one of the target urban noises with a matching Classification Accuracy score.

2. Methodology

For data classification, we follow a set of steps-

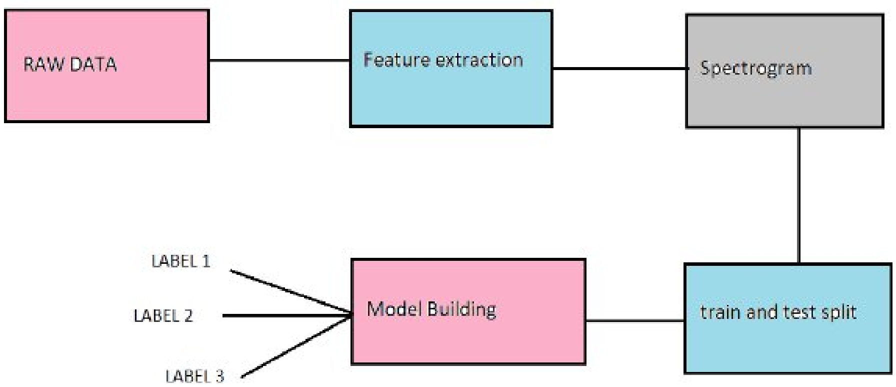
1) Using the Librosa library to load sound data.

2) Sound data is converted into numerical vector spectrograms.

3) Constructing a deep neural network.

4) Predicting how sound data will be labelled.

FLOW CHART FOR AUDIO CLASSIFICATION



Overview of the audio file: The sound clips are.wav files, which are digital audio files. Sound waves are sampled at discrete intervals, which is known as the sampling rate, and then digitized (typically 44.1kHz for CD quality audio meaning samples are taken 44,100 times per second). Each sample is the amplitude of the wave at a specific time period, with the bit depth determining how precise the sample will be, as well as the signal's dynamic range (typically 16bit which means a sample can range from 65,536 amplitude values).

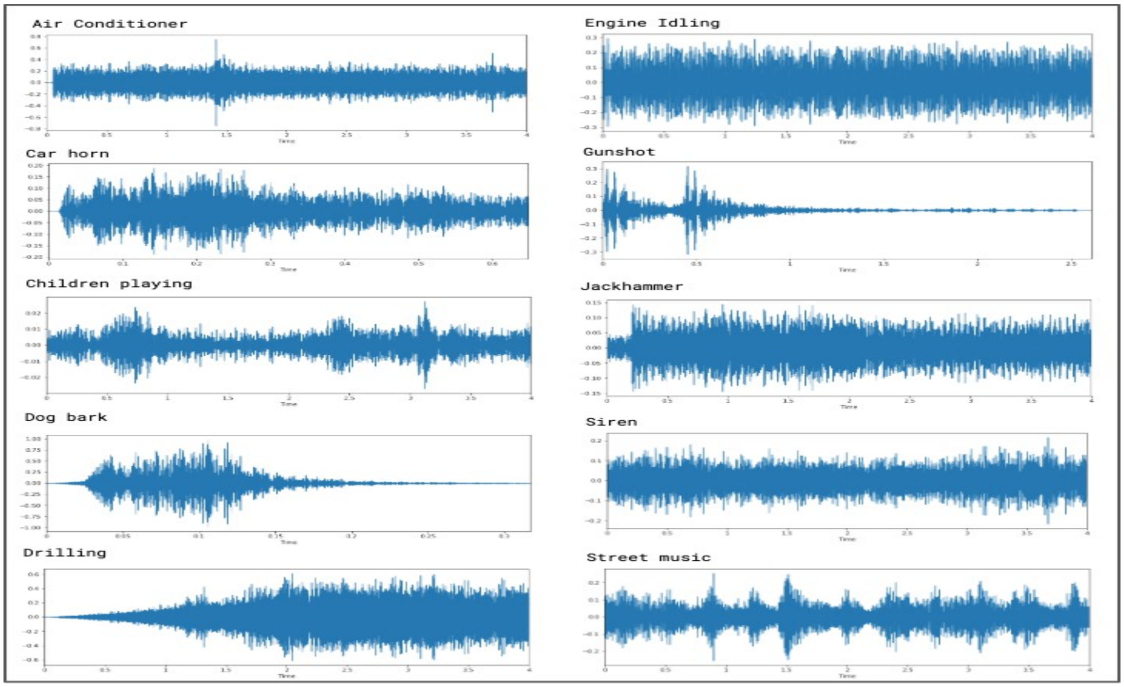
Exploratory Data: A visual assessment below reveals that visualizing the differences between some of the classes is difficult. The waveforms for recurrent noises such as air conditioner, drilling, engine idle, and jackhammer, in particular, have a similar pattern. Following that, we'll take a closer look at each of the audio files' attributes, such as the number of audio channels, sampling rate, and bit-depth.

Audio channels**:** The majority of the samples contain two audio channels (so they're stereo), but a couple just have one (mono).

Rate of sampling**:** There is a large variation of sample rates employed across all samples, which is a source of worry (ranging from 96kHz to 8kHz).

Bit-depth: There are a variety of bit-depths available (ranging from 4bit to 32bit).

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

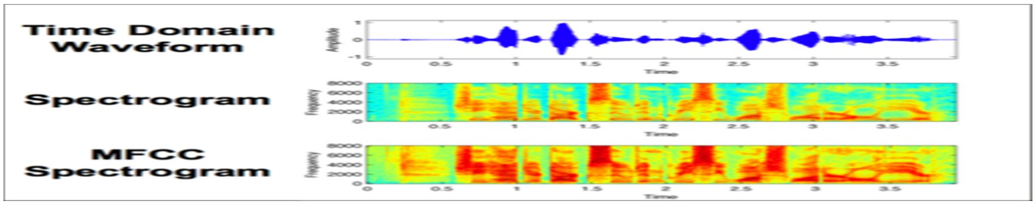


Pre-processing the data: We discovered the following audio features that require preprocessing in the previous step to maintain uniformity throughout the entire dataset:

- Sample rate
- Bit-depth
- Audio Channels

Brian McFee's Python program [8] Librosa for music and audio processing allows us to load audio into our notebook as a NumPy array for analysis and editing. We'll be able to utilise Librosa's load() method for a lot of the preprocessing, which transforms the sampling rate to 22.05 KHz, normalizes the data so the bit-depth values range between -1 and 1, and flattens the audio channels into mono by default.

Features Extraction: After that, we'll extract the features we'll need to train our model. To do so, we'll generate a visual representation of each of the audio samples, which will allow us to find characteristics for classification using approaches similar to those used to accurately classify photographs. Spectrograms are a handy tool for visualizing a sound's spectrum of frequencies and how they change over time. Mel-Frequency Cepstral Coefficients, a comparable method, will be used (MFCC). A spectrogram utilizes a linear spaced frequency scale (each frequency bin is separated an equal number of Hertz apart), whereas an MFCC uses a quasi-logarithmic spaced frequency scale, which is more analogous to how the human auditory system analyses sounds. The graphic below compares three alternative visual representations of a sound wave, the first of which compares amplitude over time in the temporal domain. Then there's a spectrogram, which shows how the energy in different frequency bands changes over time, and finally there's an MFCC, which looks a lot like a spectrogram but has more identifiable detail.

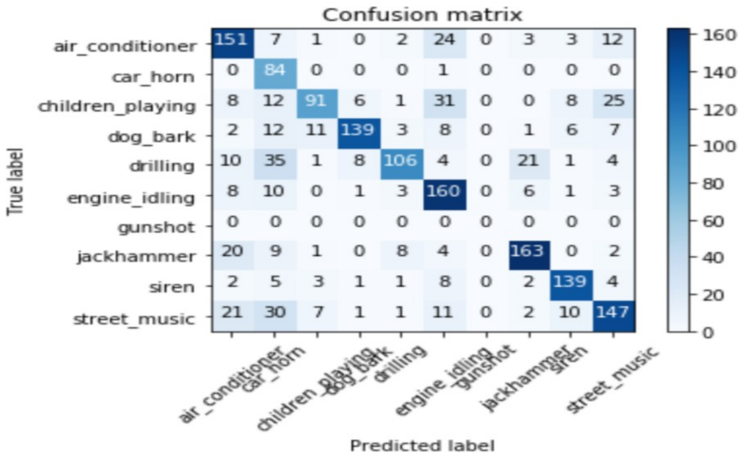


Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Creating a model: The next phase will be to use these data sets to develop and train a Deep Neural Network, which will then be used to generate predictions. We'll utilize a Convolutional Neural Network in this case (CNN) [9]. Due to their feature extraction and classification portions, CNNs often make effective classifiers and do particularly well with image classification problems. This, we believe, will be highly effective at discovering patterns inside MFCCs, much as they are at finding patterns within photos. We'll employ a sequential model, starting with a simple model architecture of four Conv2D convolution layers and ending with a dense layer as our final output layer. Our output layer will have ten nodes (num labels), which corresponds to the number of classifications that can be made.

3. Result and Discussion

Our trained model had a 98.19 percent Training accuracy and a 91.92 percent Testing accuracy. The model's performance is excellent, and it has generalized well, predicting well when tested against new audio data.


Confusion matrix

	air_conditioner	car_horn	children_playing	dog_bark	drilling	engine_idling	gunshot	jackhammer	siren	street_music
air_conditioner	151	7	1	0	2	24	0	3	3	12
car_horn	0	84	0	0	0	1	0	0	0	0
children_playing	8	12	91	6	1	31	0	0	8	25
dog_bark	2	12	11	139	3	8	0	1	6	7
drilling	10	35	1	8	106	4	0	21	1	4
engine_idling	8	10	0	1	3	160	0	6	1	3
gunshot	0	0	0	0	0	0	0	0	0	0
jackhammer	20	9	1	0	8	4	0	163	0	2
siren	2	5	3	1	1	8	0	2	139	4
street_music	21	30	7	1	1	11	0	2	10	147

4. Conclusion

The purpose of this work was to assess the utility of CNN architecture to classify sound data using sound spectrum spectrograms. Convolutional Neural Networks are commonly used to solve picture categorization issues. This research demonstrates how deep neural architectures may be used to classify sounds. In comparison to direct sound classification, this strategy using CNN for sound classification using spectrograms reduced the amount of trainable parameters. In comparison to other existing methods, we achieved a classification accuracy success rate of 91 percent in the experimental testing utilizing CNN. This approach provides promising results for the creation of sound categorization systems in key regions, according to the results of the trial.

References

[1] E. Wold, T. Blum, D. Keislar, and J. Wheaten, ‘‘Content-based classification, search, and retrieval of audio,’’ IEEE Multimedia, vol. 3, no. 3, pp. 27–36, Jun. 1996.

[2] F. Weninger and B. Schuller, ‘‘Audio recognition in the wild: Static and dynamic classification on a real-world database of animal vocalizations,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2011, pp. 337–340.

[3] M. V. Ghiurcau, C. Rusu, R. C. Bilcu, and J. Astola, ‘‘Audio based solutions for detecting intruders in wild areas,’’ Signal Process., vol. 92, no. 3, pp. 829–840, 2012.

[4] A. Rabaoui, M. Davy, S. Rossignol, and N. Ellouze, ‘‘Using one-class SVMs and wavelets for audio surveillance,’’ IEEE Trans. Inf. Forensics Security, vol. 3, no. 4, pp. 763–775, Dec. 2008.

[5] S. Chu, S. Narayanan, and C.-C. J. Kuo, ‘‘Environmental sound recognition with time–frequency audio features,’’ IEEE Trans. Audio, Speech, Language Process., vol. 17, no. 6, pp. 1142–1158, Aug. 2009.

[6] (2017). Sound Classification. [Online]. Available: http://www.paroc.com/knowhow/sound/sound-classification

[7] R. A. Altes, ‘‘Detection, estimation, and classification with spectrograms,’’ J. Acoust. Soc. Amer., vol. 67, no. 4, pp. 1232–1246, 1980.

[8] R. A. Altes, ‘‘Detection, estimation, and classification with spectrograms,’’ J. Acoust. Soc. Amer., vol. 67, no. 4, pp. 1232–1246, 1980.

[9] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[10] Pradip Kumar Yadava, Surya Deo Choudhary; 'Design & Performance Analysis of CPW Fed Square slot Antenna and Horizontal H-Shaped Stub Antenna',Volume No.4,Issue No.2,2016,PP.057-061,ISSN :2229-5828

[11] Deepak Kumar, Harendra Singhal, Somesh Kumar; 'Simulation of Steady State and Dynamic Response of Multi -effect Evaporators in Paper Industry',Volume No.4,Issue No.2,2016,PP.062-070,ISSN :2229-5828

[12] Rahul Yadav, Sanjay Gairola; 'Zigzag Connected Autotran-sformer Based 12,24 &36-Pulse Rectifiers',Volume No.4,Issue No.2,2016,PP.071-076,ISSN :2229-5828

[13] P.Kumar,R.K.Rajuvanshi,R.sharma, P.Yadav& P.Chaturvedi; 'Energy Auditing and Management :A Case Study to Improve Energy Efficiency and Setting Benchmarking',Volume No.1,Issue No.1,2012,PP.001-007,ISSN :2229-5828

[14] Gupta, Rajeev Prasad; 'Applicatin of High Voltage Pulse Electric Field in Food Industries',Volume No.1,Issue No.1,2012,PP.008-011,ISSN :2229-5828

[15] H Rathaur, N.K.Singh, & S.K.Tripathi; 'Power System Stability Enhancement with STATCOM Power Oscillation Damping Controller',Volume No.1,Issue No.1,2012,PP.012-018,ISSN :2229-5828

[16] Rao, L.Navinkumr.; 'Dynamic Performance of a Small Rating Photovoltic Module',Volume No.1,Issue No.1,2012,PP.019-022,ISSN :2229-5828