Sound-Profound

Apoorv Mehrotra, MS in Software Engineering, San Jose State University

Bhavana Bangalore Sathyaprakash, MS in Software Engineering, San Jose State University

Namratha Bhat, MS in Software Engineering, San Jose State University

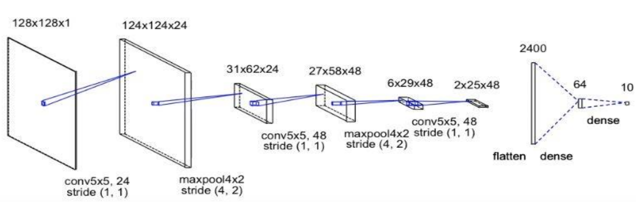
Pranjay Sagar, MS in Software Engineering, San Jose State University

***Abstract*--- The project on Sound Profound is identifying the urban sounds, labeling them to the identified category. One major use case implemented is contacting 911 in case of identifying a gun-shot sound.**

Keywords--- Deep CNN, Mel Spectrogram, Keras, Machine Learning, Python, Flask, Librosa, Angular JS, AWS cloud service, IBM Cloud

I. Introduction

The idea presented in the IEEE paper "Environmental sound recognition with time-frequency audio features" to solve the problem of environmental sound classification has been taken into consideration for the implementation of this project. The model built can be appended with the traffic signals and streetlamp posts. The basic idea is to classify and identify the maximum possible urban sounds. The environmental sound usually is difficult to detect since on most of occasions the sounds are overlapped by noises which are in turn also one of the sounds of interest, e.g., the sound of dogs barking overlapped by revving engine. Deep convolutional neural networks used in this project, facilitated with an advantage of capturing energy modulation patterns across time and frequency when applied to spectrogram-like inputs, which is an important trait for distinguishing between different noise like sounds such as jack-hammer and engine-idling.



II. Project Focus and Process Flow

A model termed as "Sound-Identifier" has been developed using Deep CNN. It can take sound wavelength files as input. It can then classify and predict the category of the sound observed in the urban areas. One of the major use cases that have been focused on is the detection of a gunshot sound. It takes immediate action on the prediction of the mentioned sound and sends a text message and an email to 911 emergency services for urgent assistance. Primarily, it has been devised to focus on the identification of day to day urban sounds by comparing them with the patterns identified for various categories. Due care has been taken to train the model on the deformed sound so that it can predict a realistic sound. With the initial training, it acquired an efficiency of 92%. For further enhancement in the acquired efficiency, the augmentation technique has been adopted. The available dataset has been augmented across two factors – Time Stretching and Pitch-Shifting. It rose the prediction efficiency to 95%. For usage consumption, the implementation of Machine learning has been carried out.

III. Technologies stack implemented

A. Deep Convolutional Neural Networks

It comprises of 3 convolutional layers interleaved with 2 pooling operations, followed by 2 fully connected (dense) layers. The input comprises a time-frequency graph from the Mel-spectrogram. Following it, Librosa is used to extract the log-scaled spectrograms with 128 components spanning over the audible frequency range, a window size of 23 ms, and a hop size of the same tenure.

B. Machine Learning- Keras Model

Training of the model has been done using Machine Learning principles. Tensorflow's Keras API has been used to load the data, clean it. After the initial training has been done, data augmentation has been implemented to expand the dataset to four-folds using time-shift and pitch-frequency attributes. This expanded dataset further improvised the efficiency of the model.

C. Python

Python has been used as the complete back-end coding language. The utilities available in Python have been made the best use of. Librosa being one of them has been used for audio analysis. Pandas is used for data analysis and data-shaping. The numpy library supports the handling of the convolutional matrix.

D. Librosa

The complete audio analysis of the wavelength files has been instrumented through Librosa. Mel-spectrogram, one of its feature utility has been used to analyze the audio files and generate a time-frequency plot for the input files.

E. AWS Cloud services

A container in AWS has been created using a docker file in the AWS instance. The entire business logic and the model have been mounted on it. The benefit of using a docker is its portability to any system.

F. IBM cloud

The front-end has been dockerised. A Kubernetes service cluster was initialized through the services offered by the IBM cloud. The docker file was placed into the cluster generalizing its availability.

G. Angular JS and AJAX based User-Interface

The web page has been designed using HTML and CSS. Soundfile input is handled utilizing Angular JS. Communication with Flask middleware is via AJAX-XMLHttpRequest.

H. Flask

It is a middleware API that connects the front-end hosted on the IBM cloud to the back-end files hosted on the AWS cloud. It handles the RESTful services for the communication passage between the input and output.

IV. Use cases

1. It can help avoid burglary in public places like malls, street corners using a gun.

2. It can save the lives of people who are shot by timely information to medical services.

3. It can help identify a spot if there is any unusual sound heard.

V. Future Enhancements

In the second stage, this model can be clubbed with edge computing. It could help detect the location of the sound traced. The message can then be conveyed about the location also for the sound captured. Also, on receiving the alert signals, a video-capturing device can be activated, which could forward the sound tenure video to the concerned authorities.

VI. Conclusion

To conclude, the designed Sound Identifier has been a successful implementation of the IEEE paper. It identifies the urban sound accurately even when mixed with other noises. As, an extension to the paper implementation, gunshot sound has been given a step ahead treatment by dropping a text and an email to 911 for immediate attention and help to the victims.

VII. Acknowledgment

A special mention to Professor Rakesh Ranjan for his continuous guidance and support throughout the lifecycle of this project. His encouragement to use the trending technologies has helped us gain a lot of exposure to the upcoming technologies and its possible use cases.

VIII. Project Repository

https://github.com/SJSUSpring2020-CMPE272/Sound-Profound

IX. References

- S. Chu, S. Narayanan, and C.-C. Kuo, “Environmental sound recognition with time-frequency audio features,” IEEE Trans. on Audio, Speech, and Language Processing, vol. 17, no. 6, pp.1142–1158, Aug. 2009.

- V. Bisot, R. Serizel, S. Essid, and G. Richard, “Acoustic scene classification with matrix factorization for unsupervised feature learning,” in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, Mar. 2016, pp. 6445–6449

- https://i.pinimg.com/474x/21/53/5d/21535d4840db763f83a3d732e3f3a497--wallpaper-backgrounds-desktop-wallpapers.jpg

- https://cloud.ibm.com/docs/containers?topic=containers-cs_cluster_tutorial

- https://loopback.io/doc/en/lb4/deploying_to_ibm_cloud_kubernetes.html#obtain-the-nodeport-of-your-service

- https://towardsdatascience.com/simple-way-to-deploy-machine-learning-models-to-cloud-fd58b771fdcf

- Dataset download link https://urbansounddataset.weebly.com/urbansound8k.html

- Video link https://www.vyond.com/