IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 9, June 2022

Acoustic Event Detection Using Machine Learning

Vaibhavi U N

Student, Department of Computer of Application
Jawaharlal Nehru New College of Engineering, Shimoga, Karnataka, India
sangeethams642000@gmail.com

Abstract: *Nowadays Audio event detection is playing an important vital role in research area it has become the main part of machine learning which plays an important role in everyday life it consists of audio tagging, classified music, emotional speech, audio sounds. Convolutional neural networks are proposed and applied on sound event detection complications. This system detects sound events such has Laughter, crying sounds of humans, Singing of Birds, Firing, speaking sounds, speech, blast and boom sounds even including animals and birds' sounds were also detected it can also include news broadcasting, each and every situation were included. Sometimes sounds might overlap at that time it becomes hard to detect the overlapped sound events so such problems can be solved by using CNN models.*

Keywords: Sound Event Recognition, Machine Learning, Convolutional Neural Networks

I. INTRODUCTION

Audio event detections can be applied on real world applications as we are surrounded by many numbers of sound events which provides highest information about what's going on around us in that current scenario and also about currently where are we. Recently the research interest in audio sound event recognitions has been increased to a greater extent some of the applications like automatic surveillance, machine hearings, and auditory scenes were depended fully on automatic detection of events performed through computers. Recently there are a lot of improvements have been brought under machine learning and also for human hearings by computer representation the difficult task is to recognize a sound from a vast noisy environment where a lot of sounds are surrounded around it includes more than 500 noises in a video can be classified using the system here the sound can be analysed automatically and tagging is also done automatically in all other applications there might be existence of specific sound events which defines clearly about an audio scene in a single file various different kinds of sound events can be detected here a multi class description used for both audio or video files class for example, one person may want to tag a holiday recording as being made just before the storm while playing with the kids and dog at the beach. While the beach may be inferred as a setting from the sound annotations, they are at different levels of annotation he can tag only playing with dog in that recording by ignoring all the other noises in that surrounding area of beach.

II. LITERATURE SURVEY

Chendong et al., [1], The sound event Detection is to recognize the environmental surrounding sound here there will be many sound events which is happening currently around us adaptive modules detect sound events.

Zhaoetal.,[2],Sound Event Early Detection is most needed task to detect the acoustic events earlier but earlier detection sometimes provides false results so Polyphonic Evidential Neural Network model is used.

Mnasrietal.,[3],The main focus is on anomalous event detections, so the main task is to detect anomalous sound events rather than concentrating on known events here we use machine learning techniques and cnn.

Me saros et al ., [4], The main approach is to detect sound event task based on supervised learning.

Suruthhi, V. S et al.,[5], The SELD which considers multichannel audio also for multiple sound classes can be predicted by CRNN.

Y. R. Pandeya et al., [6], Designs and implements an automatic sed tool for cattle sounds detection this proposed tool is tested for rare sound by using region based convolutional neural network models.

DOI: 10.48175/IJARSCT-5330

III. PROPOSED METHODOLOGY

The overall system architecture is designed by using python and the software used here is python IDLE.
Proposed system must follow the steps which are mentioned below

- Data Balancing
- Data Augmentation
- Algorithm

3.1 Data Balancing

The total number of audio clips accessible for training varies from one sound class to another sound class. There might be a large distribution of different audio clips which belongs to different sound classes training data acts as input to a PANN in mini-batches during the time of training without data balancing, audio clips will be uniformly sampled from the audio set therefore, some sound classes which is having more training clips like “Speech” should get sampled during training. In an severe case, all data in a mini-batch might belong to the same sound class. This causes PANN to overfit to the sound classes with more training clips, and underfits sound classes with a smaller number of training clips. To solve such problems, balanced sampling is designed that trains PANNs. audio clips are approximately equally sampled from all sound classes to form a minibatch the term “approximately” is used because an audio clip may contain more than one tag.

3.2 Data augmentation

Data augmentation is mainly used to ignore a system from overfitting some of the sound classes in audio set contains a lesser number (e.g., hundreds) of training clips which reduces the performance of PANNs so mix-up and Spec Augment will be applied to augment data during training.

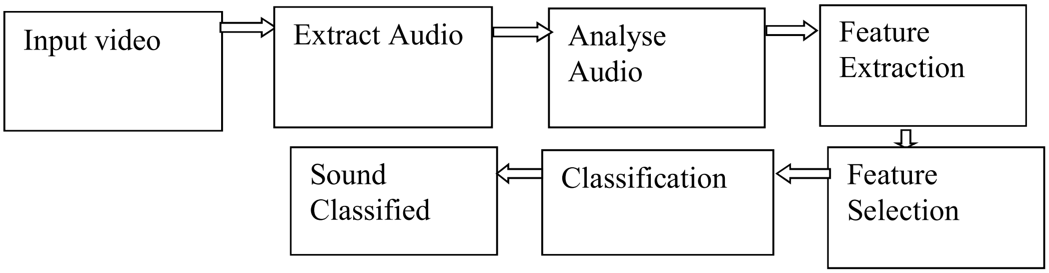


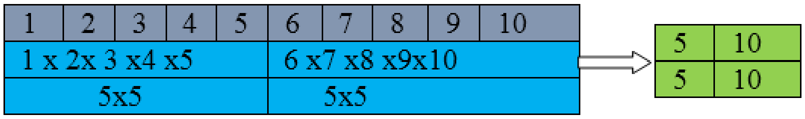
Figure: Flow of Proposed Methodology

The overall working of the sound classification is shown in the above figure.

The first block is video input block, where the unknown audio is given to the system for recognition of video is given as input the system extracts audio from video to perform analysis. once the audio gets extracted than it performs audio analysis where the system extracts the features audio by using machine learning the audio signal after enhancement is characterized into meaningful units called segments. relevant features are extracted and classified into various categories based on the information extracted sound event classification framework that compares auditory image front end features with spectrogramimage-based front-end features, using support vector machine and deep neural network classifierthe audio analysis is performed, next it performs classification based on the trained model, where it classifies the audio as Speech, Shout, Sneeze, Explosion Gunshot and gunfire.

Maxpooling

In the below example



1	2	3	4	5	6	7	8	9	10
1 x 2x 3 x4 x5					6 x7 x8 x9x10				
5x5					5x5				

5	10
5	10

Firstly 1-10 matrix will be considered in that when we apply Maxpooling to 10x10 matrix is considered in that the highest number will be taken to the another matrix.in the same way a lot of inputs of videos will be given in the audios will come from another path original audio will be converted into images as graphs by using librosa library and Mel frequency highest features from the input will be extracted and plotted as time and frequency as x and y axis.

Convolutional

From the below example

Filter

1	2	3					
4	5	6					
7	8	9					

1	2	3
4	5	6
7	8	9

When the whole matrix gets overlapped filter length of 11 is applied by using cnn then it is reduced to 5 means it will skip 5 blocks this means it takes large number of input so it also reduces the memory space than 3 convolutional layers are there and maxpooling and reshaping is also done later prediction will be conducted.

Algorithm

Wave gram-Log Mel model

#Basic steps of Algorithm#

1. System use time domain.
2. Time and frequency is implemented in our project.
3. Log Mel model is similar to wave gram but it is trained with neural network.

#Built in functions are used#

Step 1: Filter length of 11 and stride of 5 to reduce memory.

Step 2: 3 cnn layers are taken.

Step 3: Each is followed by max pooling stride 4(which is also called as down sampling)

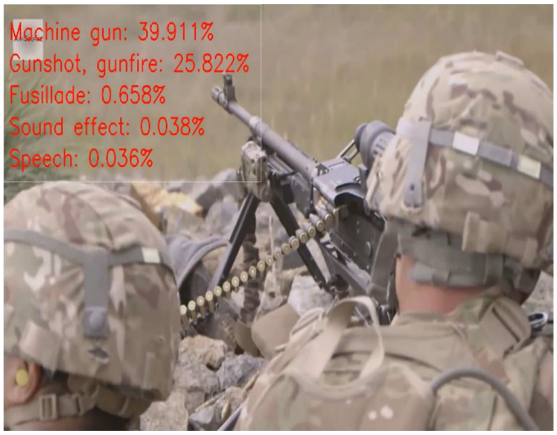
Step 4: Down sampling a 32 kHz audio recording to 32, 000/5/4/4/4 = 100 frames of features generated per second.

#Processed in Tensor flow#

Step 5: Select T Frames and C Channels apply and process the channels and frames as TXFXC/F.

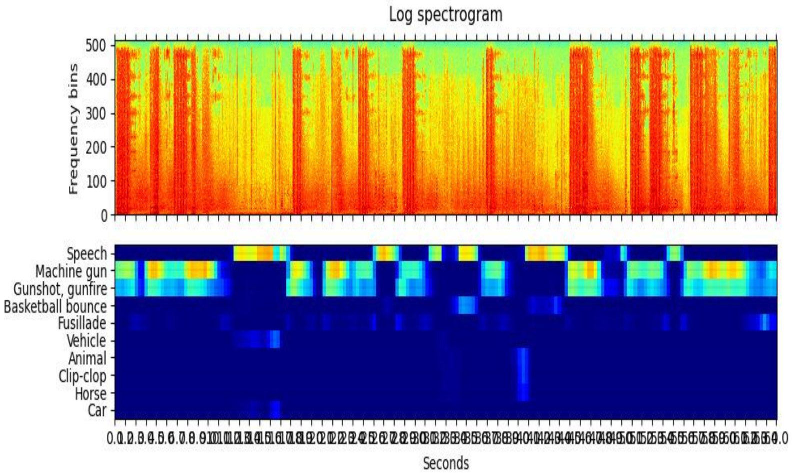
IV. RESULTS

From the below figure current scenarios of highest sound events are detected in that top 5 with their percentage accuracy will be automatically tagged in texts namely gunshot, gunfire, speech, machine gun, fusillade, Artillery fire and the % represents the higest sound events in the particular given situation.

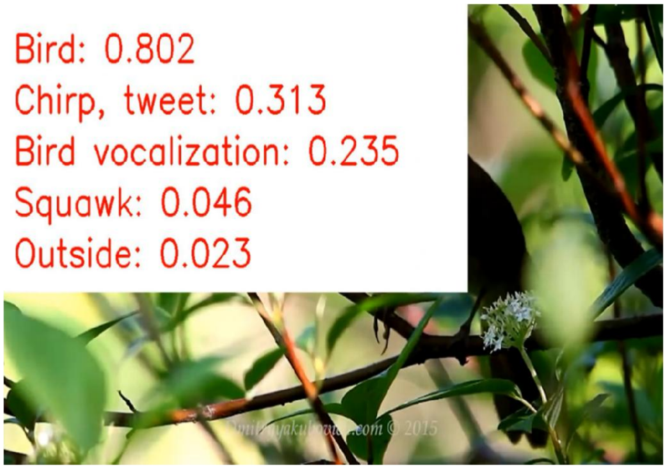


Machine gun: 39.911%
Gunshot, gunfire: 25.822%
Fusillade: 0.658%
Sound effect: 0.038%
Speech: 0.036%

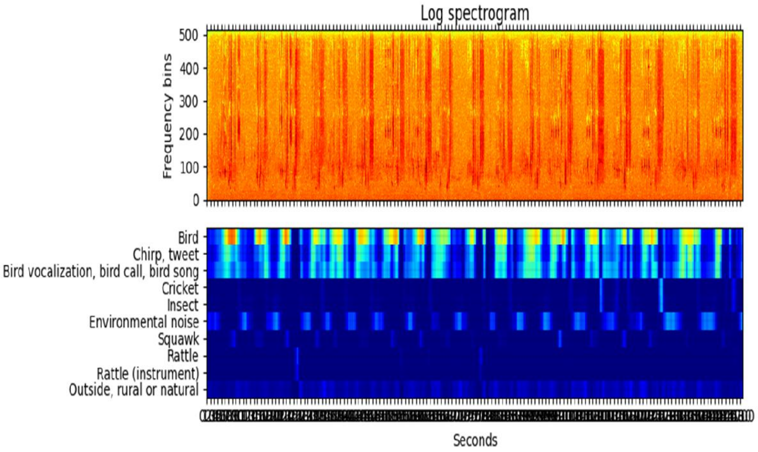
In the above graph the librosa library converts audio sound to graphs by using Mel spectrogram which can be plotted as time of seconds as x axis and frequency as y axis.



From the below figure current scenarios of highest sound events are detected in that top 5 with their percentage accuracy will be automatically tagged in texts here bird, Environmental noise ,outside.

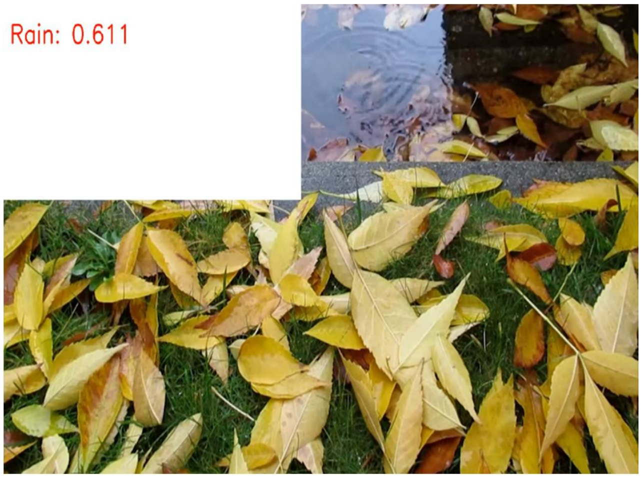


In the below graph the librosa library converts audio sound to graphs by using Mel spectrogram which can be plotted as time of seconds as x axis and frequency as y axis.



From the below figure current scenarios of highest sound events are detected in that 1 with their percentage accuracy will be automatically tagged in texts namely rain.

DOI: 10.48175/IJARSCT-5330



A single audio event can also be detected as the above diagram.

V. CONCLUSION

Over the past of years, sound events have been classified has a many variety of applications, including surveillance systems, audio content analysis, sports analysis and smart homes. Sound events were recognized automatically by computers has become an important concept of developing lot of applications such as automated surveillance, machine hearing and auditory scene understanding. Earlier, traditional methods used manual processing of parameters from sound signals which were time consuming and costly. In this proposed system a sound is classified utilizing the power and intelligence of machine learning technology.

REFERENCES

[1]. Zhao, Chendong & Wang, Jianzhong & Li, Leilai & Qu, Xiao yang & Xiao, Jing. (2022). Adaptive Few-Shot Learning Algorithm for Rare SoundEventDetection.10.48550/arXiv.2205.11738.

[2]. Zhao, Xujiang, Xuchao Zhang, Wei Cheng, Wenchao Yu, Yuncong Chen, Haifeng Chen, and Feng Chen." SEED: Sound Event Early Detection via Evidential Uncertainty." In ICASSP 2022- 2022 IEEE International Conferenceon Acoustics, Speech and Signal Processing(ICASSP),pp.3618-3622. IEEE,2022.

[3]. Mnasri, Zied&Rovetta, Stefano & Masulli, Francesco &Cabri, Alberto. (2022). Anomalous sound event etection: A survey of machine learning based methods and applications.

[4]. Me saros, Annamaria, Heittola, T., Virtanen, T., &Plumbley, M. D. (2021)"Sound event detection: A tutorial." IEEE Signal Processing Magazine 38.5(2021):67-83.

[5]. Suruthhi, V.S, V.S and Smitha V and Gini J,Rolant and Ramachandran Convolutional recurrent neural network(CRNN),gated recurrent unit(GRU),long short-term memory(LSTM),sound event localization and detection(SED).

[6]. Y. R. Pandeya, B. Bhattarai and J. Lee, "Sound Event Detection in Cowshed using Synthetic Data and Convolutional Neural Network," 2020 International Conference on Information and Communication Technology Convergence(ICTC), 2020,pp.273-276,doi: 10.1109/ICTC49870.2020.9289545.