Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

R&D Project Proposal

ENVIRONMENTAL SOUND CLASSIFICATION USING DEEP LEARNING

Manoj Kolpe Lingappa

Supervised by

Prof. Dr Paul G. Plöger
Dr. Anastassia Küstenmacher

May 2020

1 Introduction

Autonomous robots is a field of artificial intelligence which deals with designing of robots that can perform a task without any intervention from external sources. Autonomous robots will have a huge impact on our lives at home, industries and public places. There is a need for these robots to understand the surrounding environment to exhibit intelligence behaviour. One of the ways robots perceive the surroundings is through sound. The mechanical control technology of robots has grown in considerable rate in recent years. However their ability to perceive the surrounding environment especially through auditory scenes are still in the nascent stage. Acoustic scene classification supplement the image based classification in many ways such as microphone is omni-directional in nature compared with the limited camera view angle and audio signal require low computation resource and lower bandwidth. A robot fitted with a microphone can listen and interact with humans at any angle by just analyzing the sound signal from the source and it can enhance the application domain of behavioral and assistive autonomous robots. A number of researchers are working on the intelligent sound recognition(ISR) system to provide the ability to understand the real surrounding environment to the robot. The goal of environmental sound classification systems is to analyze human auditory awareness characteristics and embedding such percept ability in autonomous robots.

2 Problem Statement

Intelligent sound recognition(ISR) system is a technology for understanding sound events that exist in the real environment and integrating such perceptual ability in machines or robots. Sound classification becomes more challenging and difficult as noise levels in audio signals get increased. Traditional methods showed lower performance in the presence of ambient noise. The aim of the R&D project is to analyze the environmental sound and precisely classify the class of a detected sound.

3 Related Work

Intelligent sound recognition(ISR) is a field that addresses the need for finding the auditory scenes in the real surrounding environment. The procedure is based on studying the characteristics of the human hearing system and incorporating such abilities in machines or robots. Environmental sound classification is a basic building block of the ISR.

With the development of deep learning techniques, it is found that convolutional neural network(CNN) based models perform well in classifying the environmental sound and extensively applied for the auditory scene detection in recent years. [3] first applied CNN in ESC task and performance is analyzed. In this approach the developed model consists of 2 layer CNN with max pooling and 2 fully connected layers. Neural networks are trained with log-mel spectrograms which is an auditory feature. The classification accuracy is 5.6% higher than the traditional method. [2] [5] proposed a model that uses mel-spectrograms as features to train the CNN model. [4][1] propose a CNN model to extract the features from raw waveform and SVM is used for classification. The result shows that the model outperformed the CNN architecture trained by MFCC. However the accuracy is only 70.74%.

In order to reduce the feature dimension in traditional CNN pooling layers were applied to in CNN models which is a major drawback for auditory tasks since dimension reduction leads to loss of information and performance of the neural network is reduced. From these previous work we can find that raw waveform is directly used or single features are used to train the ESC model yielded poor accuracies.

4 Project Plan

The following sections explains work packages, milestones and project schedule.

4.1 Work Packages

WP1 Literature Search

This section aims to extensively search for reference to papers that are related

to environmental sound classification.

T1.1 Literature review
In this task collection of literature related to environmental sound classification is done and conceptual understanding of the environmental sound classification is achieved.

WP2 Data aggregation and preprocessing
This section explains the data collection and data preprocessing.

T2.1 Data collection
In this section data is collected from multiple sources and the nature of the data is examined and analyzed using visualization tools or statistical methods. Analysis is carried out to ensure data is diverse, unbiased and abundant in nature.

T2.2 Data preprocessing
Preprocessing of data is carried out based on the input requirement of the model. The preprocessing step converts the raw sourced data into a format that enables successful training of the model.

WP3 Model implementation
This section explains the development and implementation of the model.

T3.1 Evaluation of state of the art model
In this task state of the art model is implemented.

T3.2 Modification of the architecture
In this section state of the art model is modified.

WP4 Evaluation
This package aims to evaluate the results.

T4.1 Evaluation of the modified model
In this task, evaluation is conducted based on the evaluation metrics. Modified model in WP3 is trained with the dataset collected from WP2.

T4.2 Results reporting
In this task, the output of evaluation is reported.

WP5 Project Report
This work package involves writing the project report. It is done in parallel with all previous work packages.

T5.1 Documentation of literature reviewed
In this task, a detailed analysis of the state of the art is done and all the findings are documented in the project report.

T5.2 Documentation of state of the art architecture
In this task, the implementation of state of the art is documented.

T5.3 Documentation of results
In this task, the evaluation is conducted on proposed architecture and on different datasets and the results are documented.

4.2 Milestones

M1 Literature search

M2 Data collection and preprocessing

M3 Implementation of state of the art model

M4 Modification of state of the art and evaluation of modified model

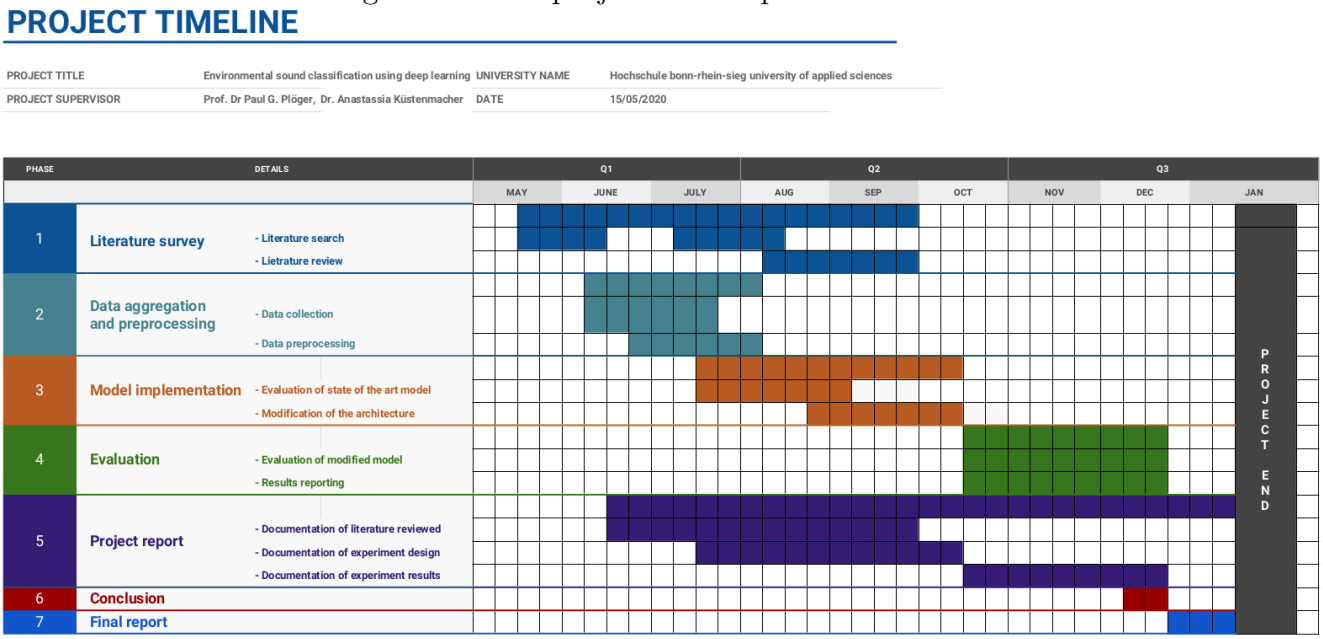
M5 Report submission

4.3 Methodologies

Literature related to state of the art environmental sound classification architecture is studied in detail. Three dataset concerning environmental sound is collected from multiple sources. Dataset is splitted into training and testing dataset. In the preprocessing step acoustic features from the data is extracted. State of the art architectures are taken and trained and tested with the three dataset. After testing the model accuracy, F1 score and confusion matrix is calculated. This model result is taken as baseline. Now state of the art model is modified by making sure that the model takes care of both temporal and spectral charcteristics of data. Modified architecture is trained and tested with the three datasets. Accuracy, F1 score and confusion matrix of the modified architecture is noted and compared with the state of the art model. Modified model is evaluated with different dataset and results are compared with state of the art model.

4.4 Project Schedule

Figure 1: R&D project roadmap



4.5 Deliverables

The following describes the deliverables in terms of minimum viable, expected and desired.

Minimum Viable

- Analysis of state of the art environmental sound classification techniques.
- Collection of appropriate environmental sound dataset for training and testing.
- Implementation of state of the art.
- Submission of report.

Expected

- Modification of state of the art.
- Evaluation of modified architecture with gathered datasets.

Desired

- Evaluation of modified model with different dataset.
- Comparison of modified model accuracy with different dataset and previously developed model accuracy.

References

[1] Sumair Aziz, Muhammad Awais, Tallha Akram, Umar Khan, Musaed Alhussein, and Khursheed Aurangzeb. Automatic scene recognition through acoustic classification for behavioral robotics. *Electronics*, 8(5):483, 2019.

[2] Matthias Meyer, Lukas Cavigelli, and Lothar Thiele. Efficient convolutional neural network for audio event detection. *arXiv preprint arXiv:1709.09888*, 2017.

[3] Karol J Piczak. Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2015. Number of citation - 411.

[4] Jordi Pons Puig and Xavier Serra. Randomly weighted cnns for (music) audio classification. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019 May 12-17; Brighton, United Kingdom. New Jersey: Institute of Electrical and Electronics Engineers; 2019.* Institute of Electrical and Electronics Engineers (IEEE), 2018.

[5] Yu Su, Ke Zhang, Jingyu Wang, and Kurosh Madani. Environment sound classification using a two-stream cnn based on decision-level fusion. *Sensors*, 19 (7):1733, 2019.