目录

[1. Background Introduction 2](#_Toc3465919)

[1.1 Problem Description 2](#_Toc3465920)

[1.2 Dataset Description 2](#_Toc3465921)

[2. Related Works 2](#_Toc3465922)

[2.1 Librosa 2](#_Toc3465923)

[2.2 XGBoost 2](#_Toc3465924)

[2.3 Convolutional Neural Network (CNN) 2](#_Toc3465925)

[2.4 Keras, Pytorch 3](#_Toc3465926)

[3. Data Prepocessing 3](#_Toc3465927)

[3.1 Melspectrogram 3](#_Toc3465928)

[3.2 Hdf5 3](#_Toc3465929)

[4. Training 3](#_Toc3465930)

[4.1 CNN 3](#_Toc3465931)

[4.2 XGBoost 3](#_Toc3465932)

[5. Test Result and Analysis 3](#_Toc3465933)

[6. Conclusion 3](#_Toc3465934)

# 1. Background Introduction

## 1.1 Problem Description

The automatic classification of environmental sound is a growing research field with multiple applications to largescale, content-based multimedia indexing and retrieval. In particular, the sonic analysis of urban environments is the subject of increased interest, partly enabled by multimedia sensor networks, as well as by large quantities of online multimedia content depicting urban scenes.

However, while there is a large body of research in related areas such as speech, music and bioacoustics, work on the analysis of urban acoustic environments is relatively scarce. Furthermore, when existent, it mostly focuses on the classification of auditory scene type, e.g. street, park, as opposed to the identification of sound sources in those scenes, e.g. car horn, engine idling, bird tweet.

There are primarily two major challenges with urban sound research namely

* Lack of labeled audio data. Previous work has focused on audio from carefully produced movies or television tracks from specific environment such as elevators or office spaces and on commercial or proprietary datasets. The large effort involved in manually annotating real-world data means datasets based on field recordings tend to be relatively small (e.g. the event detection dataset of the IEEE AASP Challenge consists of 24 recordings per each of 17 classes).
* Lack of common vocabulary when working on urban sounds. This means the classification of sounds into semantic groups may vary from study to study, making it hard to compare results so the objective of this notebook is to address the above two mentioned challenges.

## 1.2 Dataset Description

The dataset is called Urban Sound and contains 8732 sound excerpts (<=4s) of urban sounds from 10 classes: Air Conditioner, Car Horn, Children Playing, Dog Bark, Drilling, Engine Idling, Gun Shot, Jackhammer, Siren, Street Music.

# 2. Related Works

## 2.1 Librosa

## 2.2 XGBoost

## 2.3 Convolutional Neural Network (CNN)

## 2.4 Keras, Pytorch

# 3. Data Prepocessing

## 3.1 Melspectrogram

## 3.2 Hdf5

# 4. Training

## 4.1 CNN

## 4.2 XGBoost

# 5. Test Result and Analysis

# 6. Conclusion