Recognize and Extract Special Sound Signals using Neural Networks

Jinghuai Tang

SID 520378067

Github ID: JinghuaiTang https://github.com/JinghuaiTang/elec5305-project-520378067

7/9/2025

Overview

The theme of this project is to use neural networks to recognize and extract special sound signals. Our plan is to build and train an appropriate model so that the system can automatically distinguish target sounds from background noise and extract useful features. This will not only help us gain a deeper understanding of sound signal processing and deep learning methods, but also lay the groundwork for future applications such as speech recognition, medical sound detection, or environmental monitoring. In this project, we will focus on implementing a basic model and verifying its performance, combining with complex structures if the simple model reaches a high qualify.

Background and Motivation

When studying sound signals, we found that traditional approaches often struggle to effectively recognize and extract special sounds in complex environments. For example, in noisy scenarios, target sounds are frequently masked, which reduces the accuracy of analysis. With the development of deep learning and neural networks, stronger automatic feature learning capabilities have emerged, providing new opportunities to address this challenge.[1] The motivation of this project is to explore these methods in order to improve the processing of special sound signals. In addition, I have already studied topics related to neural networks and machine learning, which gives me the necessary foundation and confidence to apply what I have learned in class to a real project. As students, we also aim to further build up our experience through this project and prepare ourselves for more advanced research and applications in the future.

Methodology

In this project, we adopt a neural network approach to recognize and extract special sound signals. The methodology is divided into four main stages: data preparation, model design, training process, and performance evaluation.[2] Within these stages, we also specify the tools and platforms to be used, the signal processing techniques applied, and the sources of data.

First, during the **data preparation** stage, we will obtain datasets from publicly available sound databases that contain both speech and noise samples. If necessary, we may also record specific sound samples to enrich the dataset. Preprocessing will include noise reduction, normalization, and framing to ensure the neural network can effectively learn the features. To improve robustness, data augmentation will be applied, such as adding background noise or performing time stretching.[3]

In **model design**, we focus on three aspects:

1. **Input feature representation**: Sound signals will be converted into spectrograms (e.g., through STFT) or Mel-Frequency Cepstral Coefficients (MFCCs), which provide rich time-frequency features for neural network learning.
2. **Signal processing techniques**: In addition to neural network methods, we will incorporate traditional signal processing techniques, such as filtering or spectral subtraction, as baseline comparisons.
3. **Tools and platforms**: Python and its deep learning frameworks (e.g., PyTorch) will be used for model implementation, while common signal processing libraries (e.g., Librosa) will assist in preprocessing.

During the **training process**, we will use the cross-entropy loss function and optimization algorithms such as Adam for parameter updates. Mini-batch training will be adopted, and validation performance will be monitored to adjust hyperparameters dynamically.[4]

Finally, in **performance evaluation**, the model will be tested on a separate dataset, with accuracy, precision, recall, and F1-score as evaluation metrics. We will also assess robustness under different noise environments to validate practical feasibility.

Overall, the proposed methodology integrates appropriate data sources, a combination of traditional and modern signal processing techniques, and widely used deep learning tools to form a complete workflow. This process not only enables us to achieve the project objectives of recognizing and extracting special sound signals but also helps us gain practical experience in combining sound signal processing with artificial intelligence.

Expected Outcomes

The expected outcomes of this project can be summarized in three main aspects. First, we aim to successfully train a neural network model capable of recognizing and extracting special sound signals. The model should demonstrate the ability to distinguish target sounds from background noise in complex environments and achieve satisfactory performance in metrics such as accuracy, precision, recall, and F1-score.

Second, we expect to gain valuable hands-on engineering experience throughout the project. In particular, we will become familiar with widely used deep learning platforms (e.g., Python and PyTorch) and signal processing tools (e.g., Librosa), while also practicing the integration of traditional signal processing methods with modern neural networks. This will not only help us complete the current task but also prepare us for tackling more advanced research projects in the future.[5]

Finally, we hope to establish a small-scale experimental workflow for sound signal recognition and extraction, covering the full process of data preparation, model construction, training, and evaluation. Such an outcome may serve as a reference for similar student projects and demonstrate the potential of artificial intelligence methods in sound signal processing.

References

[1] V. Boddapati, A. Petef, J. Rasmusson, and L. Lundberg, "Classifying environmental sounds using image recognition networks," *Procedia computer science*, vol. 112, pp. 2048–2056, 2017.

[2] A. d. Garcez, K. Broda, and D. M. Gabbay, "Symbolic knowledge extraction from trained neural networks: A sound approach," *Artificial Intelligence*, vol. 125, no. 1-2, pp. 155–207, 2001.

[3] O. Gencoglu, T. Virtanen, and H. Huttunen, "Recognition of acoustic events using deep neural networks," in *2014 22nd European signal processing conference (EUSIPCO)*. IEEE, 2014, pp. 506–510.

[4] Y.-K. Lin, M.-C. Su, and Y.-Z. Hsieh, "The application and improvement of deep neural networks in environmental sound recognition," *Applied Sciences*, vol. 10, no. 17, p. 5965, 2020.

[5] H. Zhang, I. McLoughlin, and Y. Song, "Robust sound event recognition using convolutional neural networks," in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2015, pp. 559–563.