Visual Sound Assistant

Ke Xia, Song Yang, Xin Yang and Tong Wu
{kx38, sy540, xy213, tw445}@scarletmail.rutgers.edu
Advisor: Prof. Yingying Chen

Goal

- **Goals ---- Helping people with hearing loss**
  - Aiming to help people suffering from hearing loss, we developed an Android application that can collect sound information.
  - Our App extracts the information hidden in the background sounds and convert it into text. If something dangerous is happening, our App will notify users on time.

Motivations and Objectives

- **Motivations**
  - Apply **deep learning** and **mobile technology** in hearing assistance area.
  - Various sound recognition.
  - Help people with hearing issues to have the environmental sound sensitivity.
- **Technical Objectives**
  - An **Android application** to identify sound classification.
  - Collect **sound information around** like barking, shooting, alarming, etc...
  - Send **real-time notifications** to users.

Research Challenges

- Sound signal processing algorithms.
- Neural network training with limited datasets.
- Noises can interfere the accuracy.
- Duration of the audio sample also can affect the accuracy.
- Manually customize server. Updating and synchronizing the neural network model between clients and server.
- User experience issues on Android.

Acknowledgement

We would like to thank Prof. Yingying Chen for her kind support and guidance through the whole program.

Methodology

Sound Signal Processing

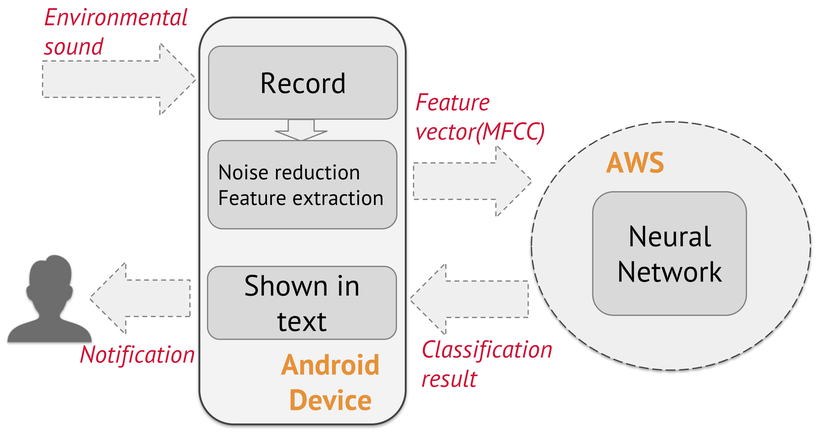
- MFCC(Mel-Frequency Cepstral Coefficients) based on LibROSA[1] extracts 40 features per audio.

Neural Network Structure

- A fully connected neural network built on TensorFlow.
- Classify up to 10 categories using UrbanSound8k[2] dataset.

Android Development

- Multi-thread audio recorder and background service ensure zero omission.
- Adjustable notification priority to avoid annoyance.
- Interactions with the server can effectively collect cases and achieve more robust model in the long term.

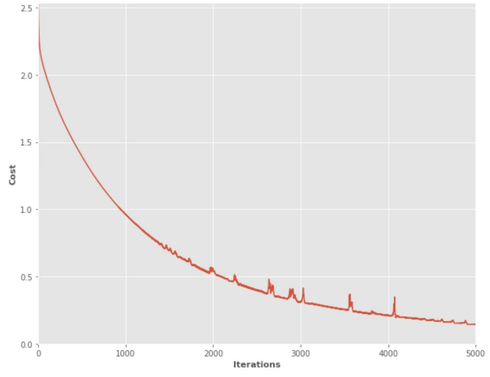


Server Implementation

- Serve a Tensorflow Model online.
- The server is based on Amazon Web Service(AWS).
- The server receives feature vector and sends classifying results back to Android device.

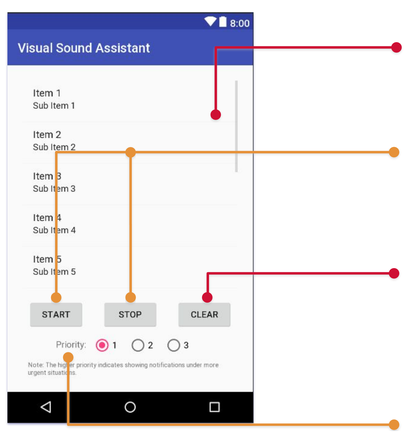
Results

- Extract **40** MFCC features.
- **10** Classification categories.
- Accuracy over **96%** on UrbanSound8k.
- Real-time response **within one second.**



F-Score: 0.852

f-score of neural network



Classification results are shown in the textbox

"Start" & "Stop" Buttons to start/stop sound recognition

"Clear" Button to clear the textbox above

User can choose different notification priorities

Future Development

- Keep optimizing the neural network using users feature vectors. Hot-update features. More layers to enhance the recognition accuracy. Move software to a smaller Linux device.

References

[1] LibROSA: https://github.com/librosa/librosa
[2] UrbanSound8k: https://serv.cusp.nyu.edu/projects/urbansounddataset/urbansound8k.html
[3] Piczak, Karol J. "Environmental sound classification with convolutional neural networks." Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on. IEEE, 2015.