Computer Engineering Program

Project

Zamplify

A Sound Recognition System for Context Awareness

CHENG Kok Hang, Sam CHEUNG Tsz Him, James
KWOK Hin Kwan, Billy SHIN Wai Ching, Martin

Advised by Professor Brian MAK Kan Wing

Abstract

Natural sound can be an important indicator for context recognition. This poster presents the design, implementation, testing and evaluation of Zamplify, a real-time natural sound recognition system. The Zamplify Android app and its complementary IoT device continuously recognize context from sound in the surroundings and provide a customizable trigger-action mechanism that performs actions when certain context is detected. The core technologies used in Zamplify include a convolutional neural network for extracting sound features and a recurrent neural network (with long short-term memory cells) for modelling sequential information of audio. We evaluated the system in a set of 9 daily-life ambient sounds and discussed its performance. The result shows that it is feasible and practical to use sound for context awareness and demonstrates the potential of a sound-based context recognition system.

What to know more?

Try it yourself Get report Contact us

What if your phone knows what you need?

What if your apps finally make recommendations that are relevant to you?

What if you can define rules to make your phone automatically do tasks that you want it to do when you are in certain context?

Zamplify makes it possible!

Overview

Project Zamplify explores the possibility of sound by letting you trigger actions when our state-of-the-art machine learning model recognizes a specific context from the sound in your surroundings. You just need to install our Android app and your phone automagically becomes an intelligent butler who hears what you need! You can also put our smart recognizer at your home or office to look after your most important place. Developers can use our Prediction API to improve user experience and provide more relevant recommendations to their users based on user context.

Why sound?

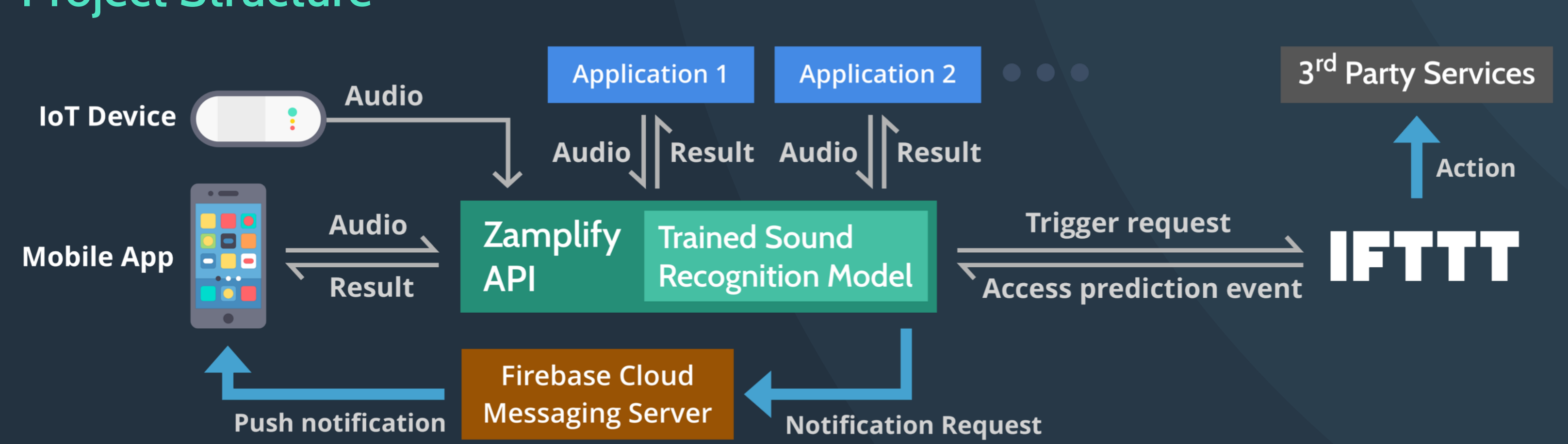
Specific signals like GPS, motions, proximity do not convey enough information about the environment. Visual signals such as image and video are hugely affected by line of sight. Speech signal includes only verbal content.

Meanwhile, the power of natural sound is often overlooked. In fact, an enormous amount of context-related information can be inferred from sound. The "noise" around us is indeed a good indicator of context.

What are included in Zamplify?

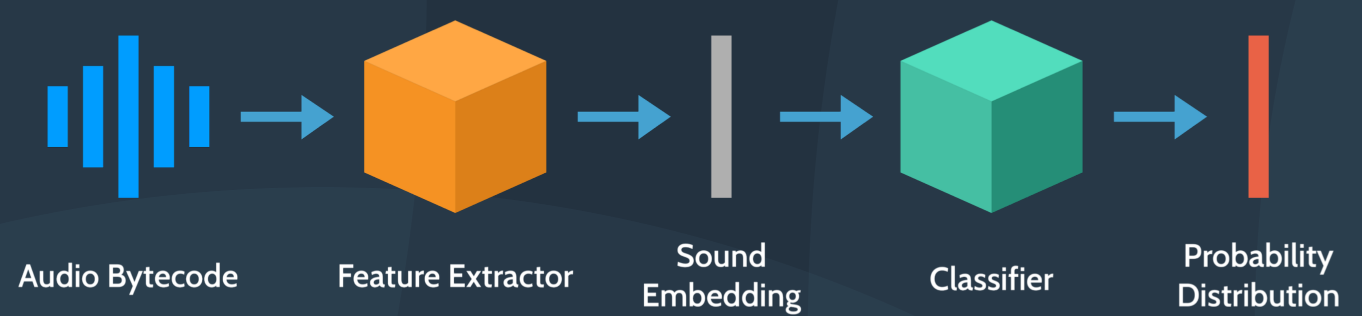
Component	Description
Sound Recognition Model	A machine learning model that recognizes context from audio
API	An online service that wraps the sound recognition model and bridges it with different devices and applications
Mobile App	An Android app that records surrounding sound on users' device, communicates with the server to provide sound recognition services, and triggers actions upon the detection of predefined context
IoT Device	Bring sound recognition service to home and businesses for security and customer insight

Project Structure



Sound Recognition Model

The sound recognition model takes raw audio bytecodes as input and gives a probability vector as output. It consists of two sub-models - a feature extractor and a classifier.



Feature Extractor

It is a partial reimplementation of SoundNet, a convolutional neural network (CNN) created by researchers at MIT. The original model was created by transferring the weights of a state-of-the-art pre-trained image recognition model to a sound recognition model, using videos as training examples. Each video is split into an image sequence and a audio clip, then fed into the transfer learning model for weight transfer. After a few trials, we used the first 18 layers of a SoundNet as the feature extractor of our model because this results in the highest accuracy.

Dataset

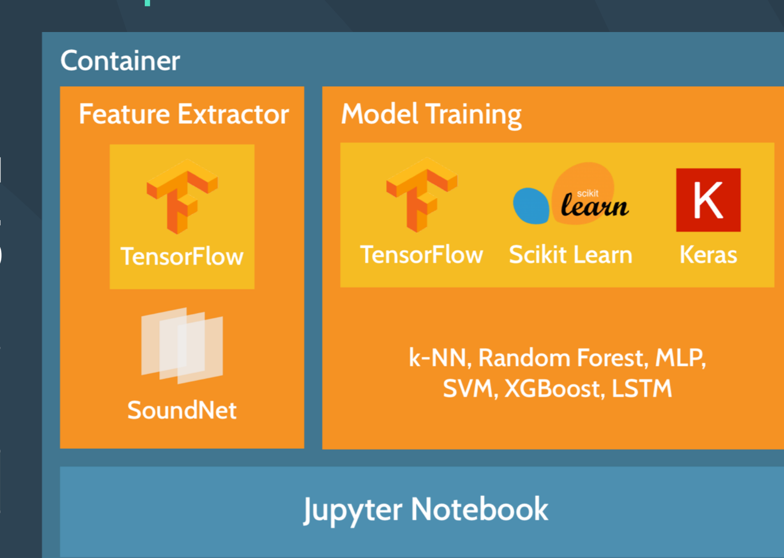
The model was trained with AudioSet, a sound clip dataset sourced from YouTube by Google. It was tested and validated with the ESC-10 dataset, in addition to AudioSet.

	# of class	# of samples	Audio Length
ESC-10	10	400	5 sec
AudioSet	632	2,084,320	10 sec

Classifier

With the output of feature extractor, we tested various classifiers, including SVM, XGBoost, Random Forests, k-NN, MLP and LSTM. We found that LSTM gives the best performance because it can capture sequential information, in addition to spatial information. Such novel approach of applying LSTM to CNN output successfully achieves an accuracy of about 80% on 7 to 9 sound types.

Setup



API

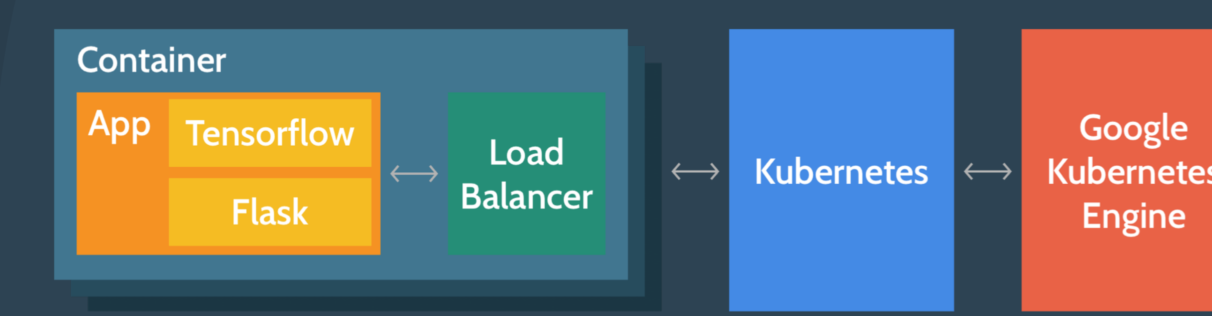
Zamplify API is in fact a collection of RESTful APIs that manages different aspects of the service. The official Zamplify app and IoT device use this API to manage users, recognize context and integrate with third-party services. Apart from our official app and device, other developers can also access this API to add context awareness capability to their applications with ease.

API Functions

API	Description
Prediction API	Convert uploaded audio files into bytecodes, pass it to the sound recognition model, and return a context probability distribution in JSON.
IFTTT Integration API	Convert the prediction result into a format readable from IFTTT, and request IFTTT to read the result.
Device and Push Notification API	Pair up IoT devices with user accounts, and push recognition result of IoT devices to mobile app through Firebase Cloud Messaging API.
User Management API	Authenticate users with their Facebook account and manage the association between Zamplify accounts and IFTTT accounts/IoT devices.

Architecture

The API uses Flask, a Python REST framework, to provide endpoints. User data are stored in MongoDB and the associated Prediction API wraps our sound recognition with additional functionalities. The entire backend system adopts containerization to abstract the runtime environment from different hardware infrastructure. A Docker image packs the code into a container to further simplify the deployment. To make the API service scalable, Kubernetes is used to manage and scale Docker containers depending on network traffic.



Android App

Our Android app allows users to manage their context-action triggers, while at the same time serves as a demonstration of what our API can do. In addition, it allows users to view their recognition history and take photos using sound as trigger.

IoT Device

The IoT device that we built is an Internet-connected, always-on sound recognizer that can be used in homes and offices. It demonstrates the power of sound recognition in the context of smart home or smart office.

The device was developed using Raspberry Pi Zero W microcontroller and a ReSpeaker Mic Hat microphone array. A case was 3D-printed to protect the electronic components and to improve the aesthetics. The device constantly records environmental sound and triggers actions specified by its users. Users can use the Zamplify mobile app to configure their trigger-action pair for sounds collected from the IoT devices.