Project Presentation

Speech Classification in Android

David Cleres – M.Sc. in Computational Science & Engineering

Monday, 13th of April 2020

Today's Menu

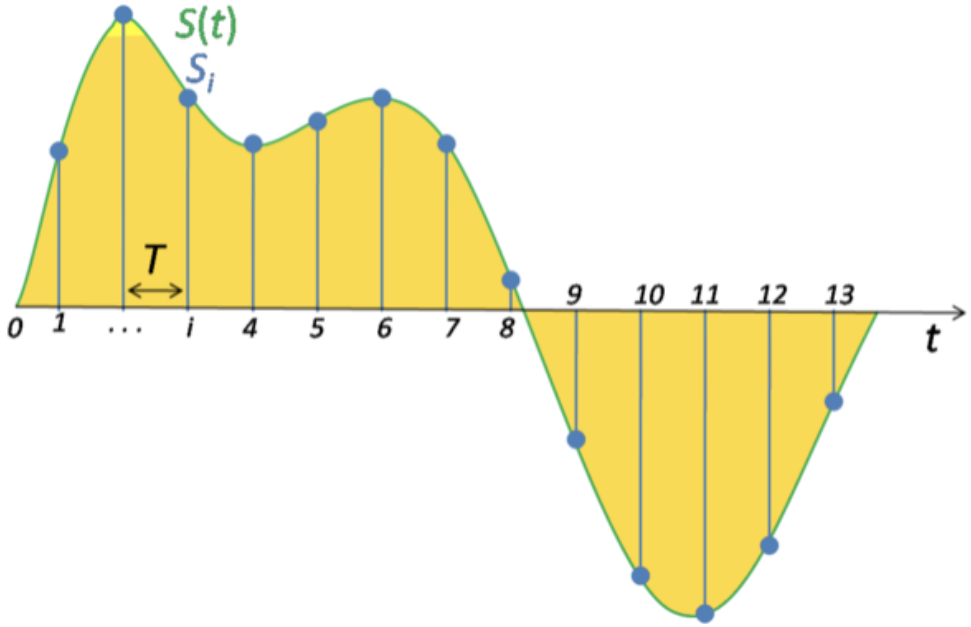
1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

Today's Menu

1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

Sound
Data

- Sound is a superimposition of sine waves
- Can be summarized in
  - amplitude (S(t))
  - frequency (1/T)
- Many transforms exist to extract information for sound data
  - FFT
  - STFT
  - MFCC
  - …



Today's Menu

1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

Google's AudioSet

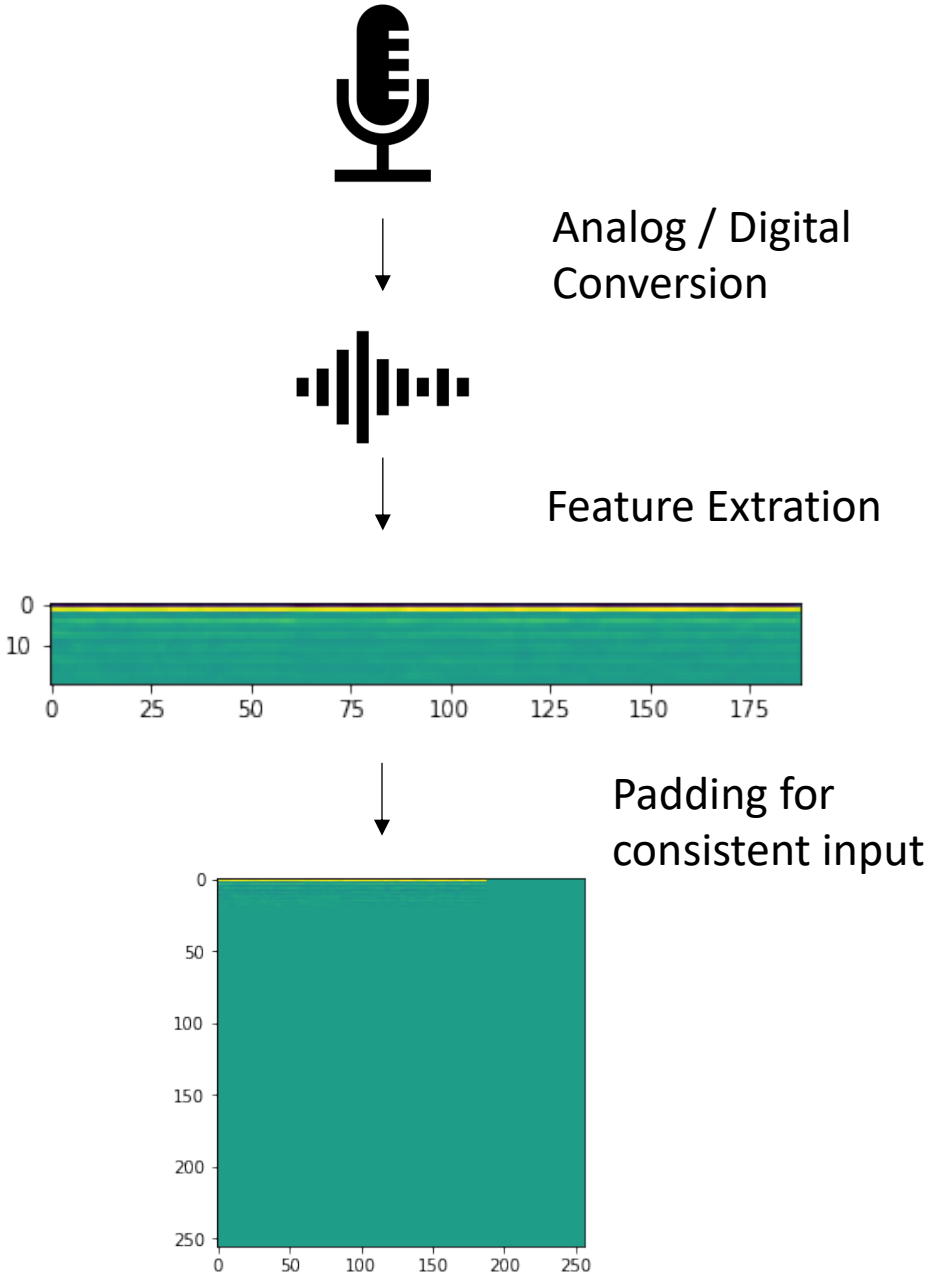
- 10 seconds recordings
- 48 kHz sampling rate
- From Youtube video
- 632 audio event classes
- Humanly labelled
- 2.1 million annotated videos
- 5.8 thousand hours of audio

Today's Menu

1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

Pre-Processing

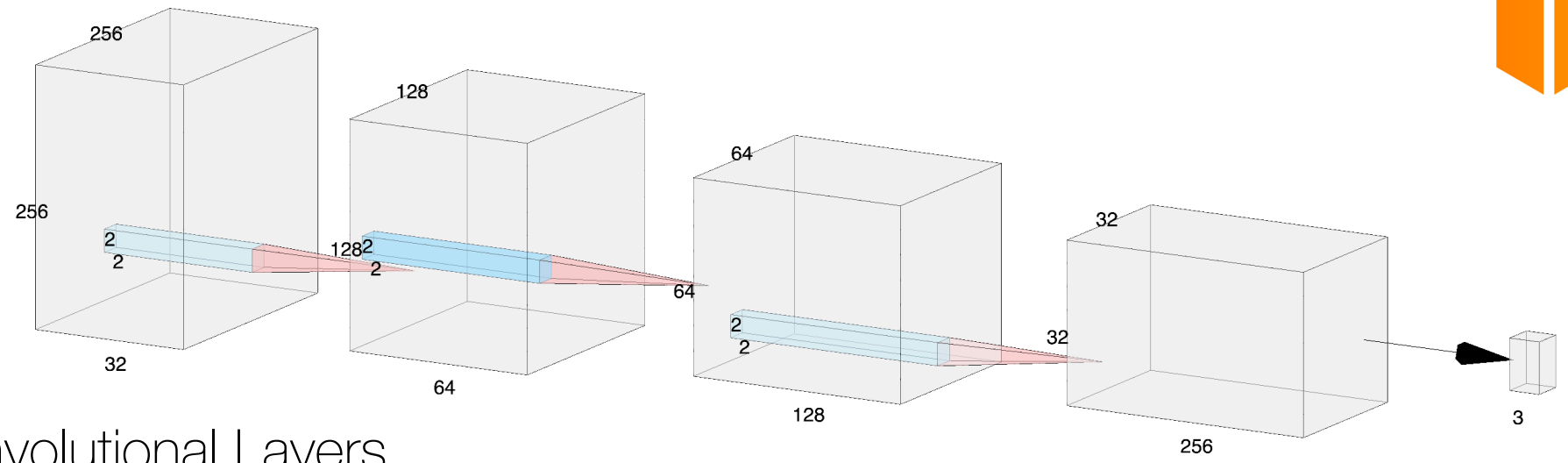
- Convolutional Neural Networks are state of the art also for audio data
- Sound can be shown as images through spectrograms
- **Mel Frequency Cepstral Coefficient (MFCC)** provides good training data



Today's Menu

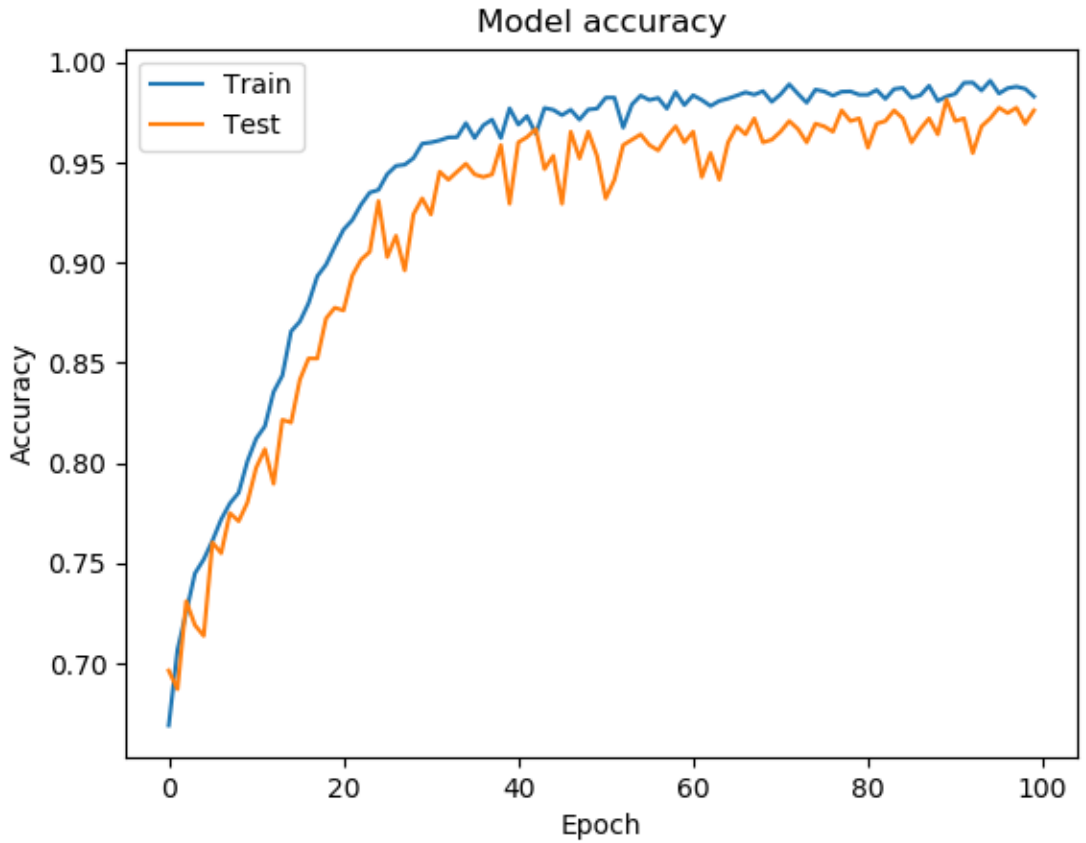
1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

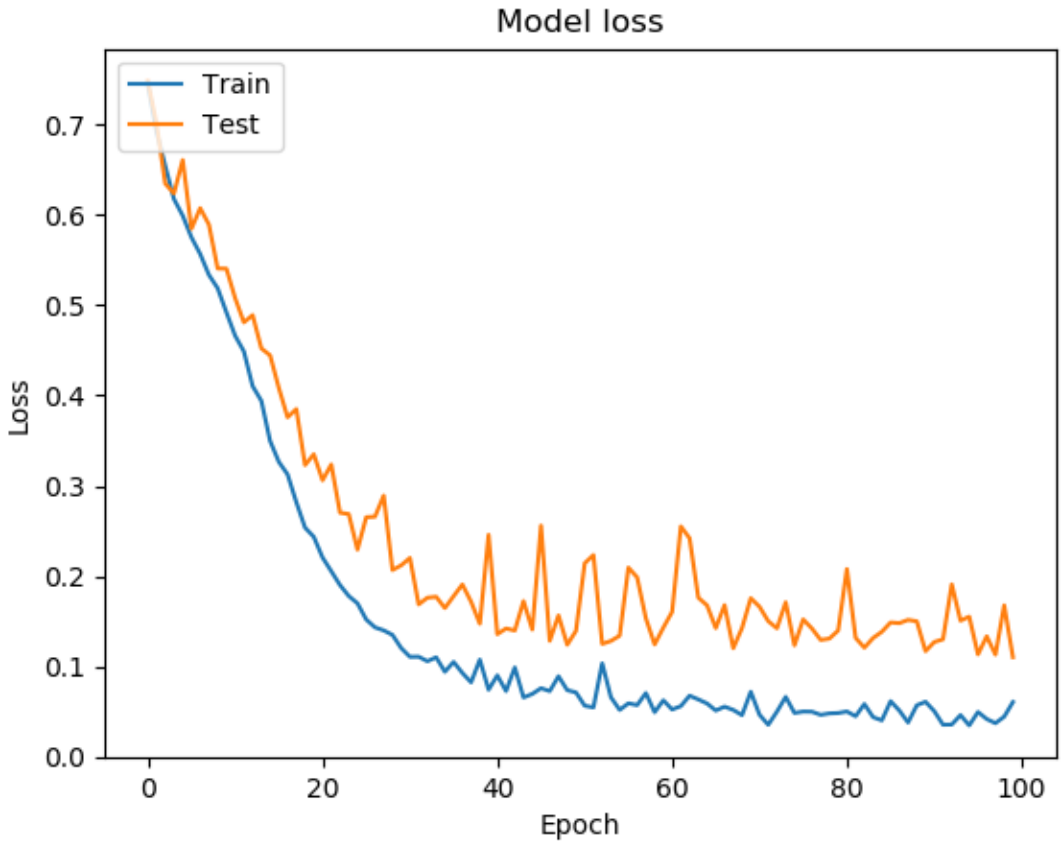
Model
Details



- 5 layers
  - 4 Convolutional Layers
  - 1 Fully Connected layer
- Max pooling and Dropout in the first four layers
- **Total params:** 173,411 **Epochs**: 100 **Optimizer**: Adam
- **TFlight file**: 697 KB **Training time:** 7 hours 35 minutes

Model training
Performance

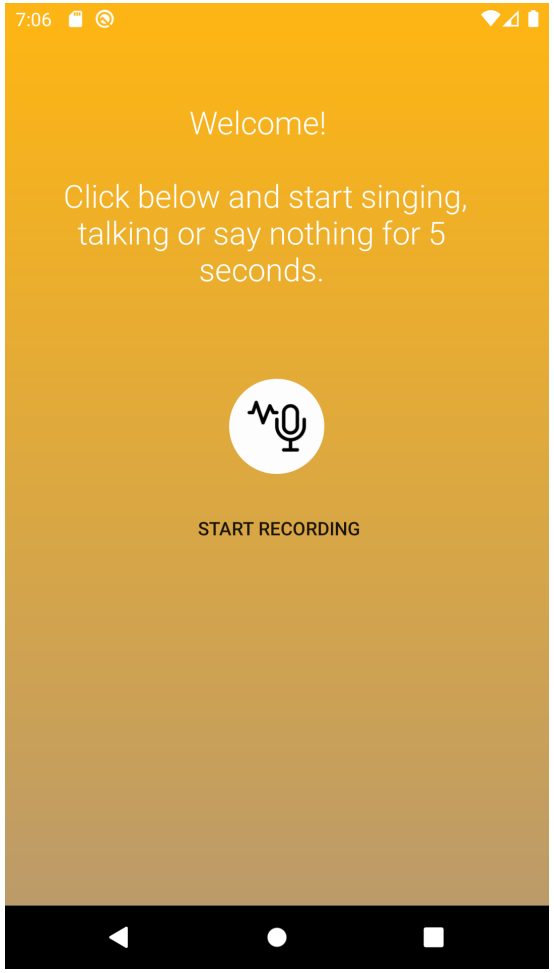


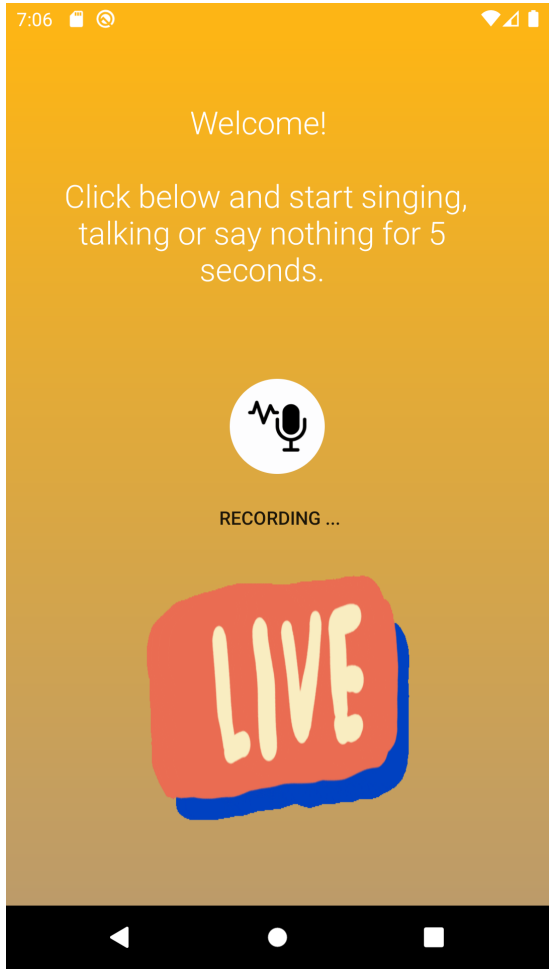


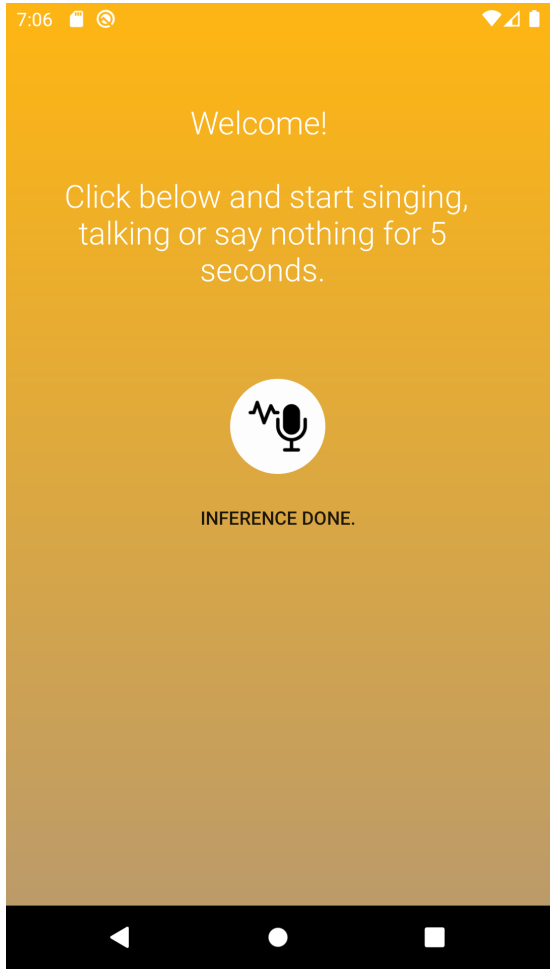
Today's Menu

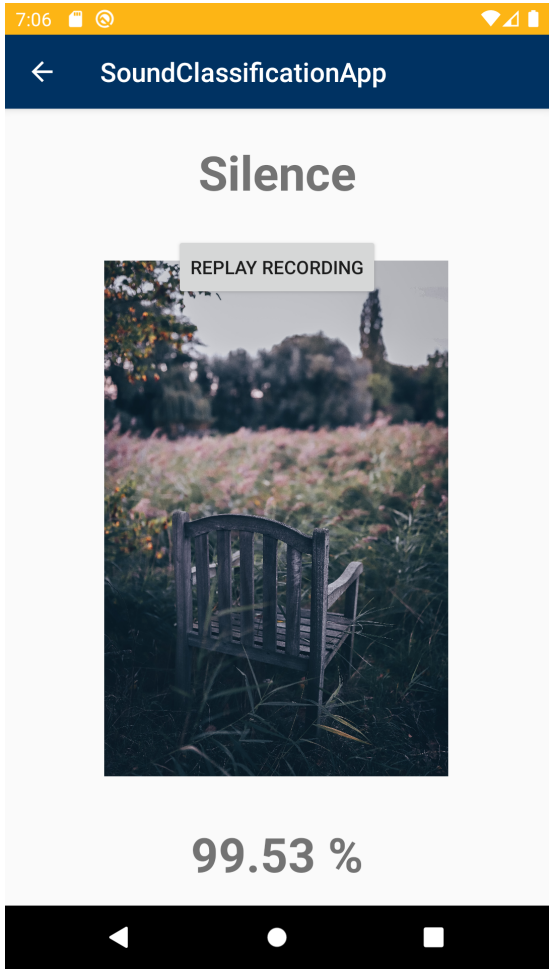
1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

App Screenshots









Today's Menu

1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

Challenge
Conclusions

- Data from AudioSet Dataset (~9000 train recordings of 2 seconds)
- Robust Model with over 95% accuracy on training and validation set
- Lightweight model for mobile devices
- Functional App
- Predictions shown directly on screen
- Good generalization of the model to predict phone microphone input

Today's Menu

1. Introduction
2. Dataset
3. Pre-processing
4. Models
5. Android App
6. Conclusions
7. Outlook & Next Steps

Outlook & Ideas

- Use data augmentation to train on more data
  - Add noise / Noise Injection
  - Accelerate / Slow down recording
  - Play with amplitude
  - Shifting time
- Download more data from other datasets
  - Examples
    - LibriSpeech
    - VoxCeleb
    - Common Voice

ETH zürich

Thanks

Any questions?