FINAL PRESENTATION

BIRD SONG IDENTIFICATION

Thanyaporn Phinthuphan

8 May 2019

Outline

- introduction

- methodology

- results

- discussion

- deployment & demo

INTRODUCTION

MOTIVATION

PREVIOUS STUDIES

OBJECTIVE

BENEFITS

Introduction

Motivation

Introduction

Motivation

- large number of different species

- background noise

- multiple birds singing at the same time

- intra/inter-species variance

Introduction
Previous studies

- NIPS4B 2013 → France

- ICML 2013 → France

- MLSP 2013 → USA

- BirdCLEF 2014-2019 → USA

Bird Song Id Automatic Recognition & Reference - Birds of the British Isles 4+
Mullen & Pohland GbR
#52 in Reference
★★★★☆ 4.3, 425 Ratings
£3.99

Warblr: Identify UK bird songs
Warblr
#57 in Reference
★★☆☆☆ 2.0, 26 Ratings
£4.99

ChirpOMatic USA 4+
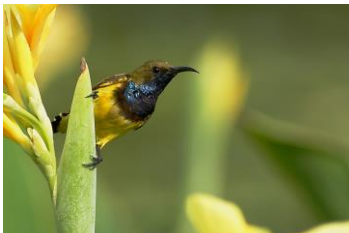Automatic Bird Song ID
Spiny Software Ltd
£3.99

What about Thai birds?

Introduction
Objective

- to develop the program for identifying bird species from their sounds

- including only 40 common birds in Thailand

Introduction
Benefits

- bird learning

- look-alike bird identification

- population/migration monitoring

METHODOLOGY

DATASET

REFERENCES

FEATURE GENERATION

NETWORK ARCHITECTURE

Methodology
Dataset

- bird song dataset - https://www.xeno-canto.org/

Methodology

Dataset

- query by science name as of March 2019

- select only class that have more than 10 files

  → 80 class / 8,746 files

- length 1 sec – 10 min

- training data - quality A/B/C with not more than 100 files/class

  → 80 class / 5,507 files (63%)

- sampling data → 11 class / 1,847 files

Methodology
References

- paper   - winning solution for NIPS4B 2013; Mario Lasseck

            - winning solution for BirdCLEF 2016; Elias Sprengel & teams

- model   - Multi-label classification with random forest

            - Multi-class classification with CNN

            - Multi-label classification with CNN

Methodology
References

Spectrogram

Median Clipping

Closing & Dilation

Median Filter & Small Objects Removed

- 1 : Multi-label classification with random forest

- referred from Lasseck (2013)

- preprocess: wav → STFT → median clipping → segment

- feature: segment correlation

- classifier: 87 random forest

Segmented Spectrogram (log)
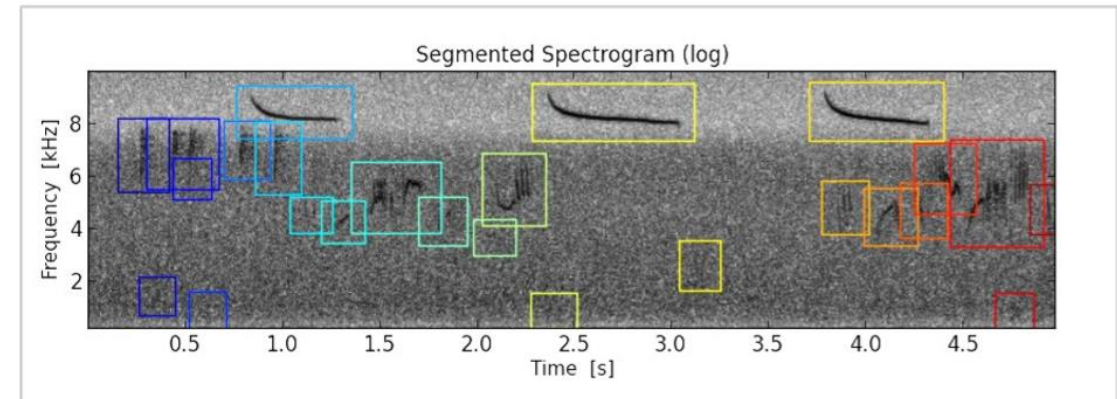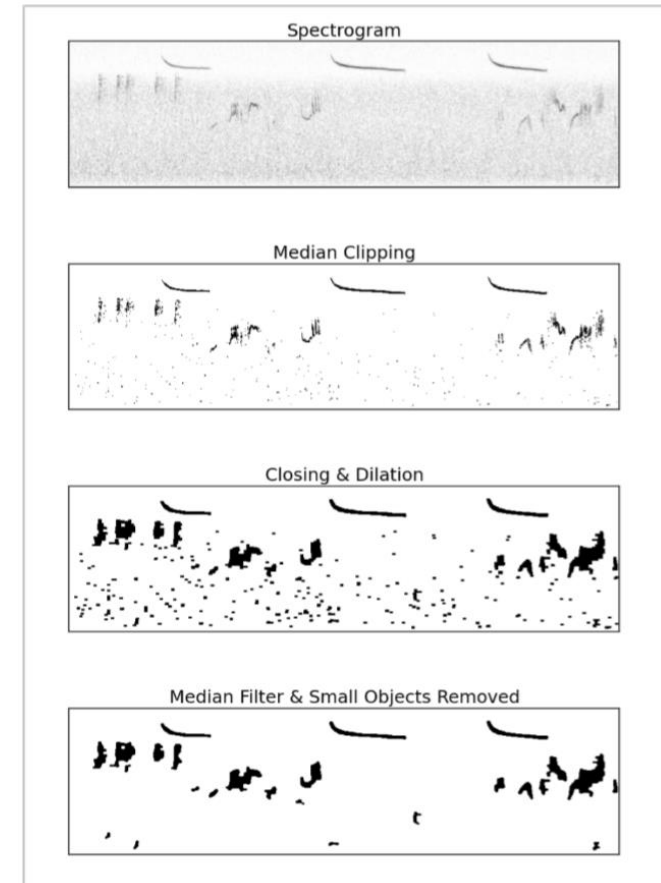
Methodology
Feature generation

- preprocessing

- STFT using hanning window → normalized

- reducing background noise with median clipping

- closing & dilation → segmentation (size/position)

Spectrogram

Median Clipping

Closing & Dilation

Median Filter & Small Objects Removed

Segmented Spectrogram (log)
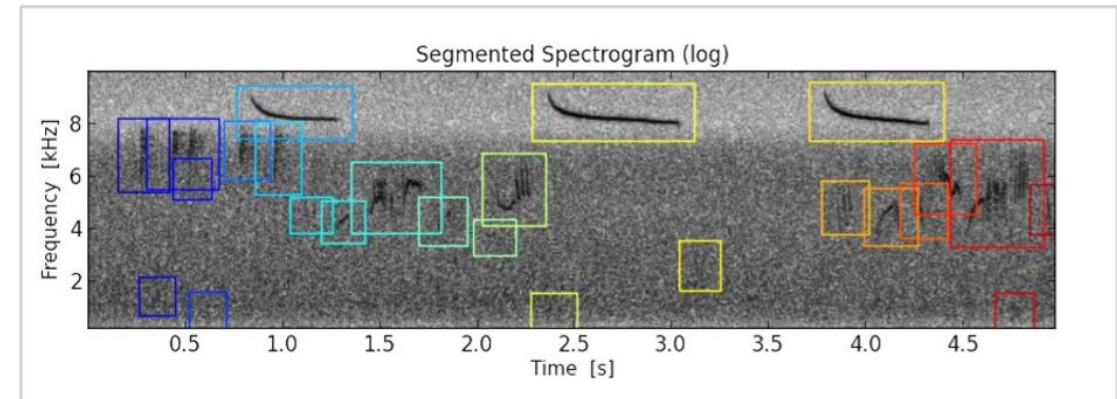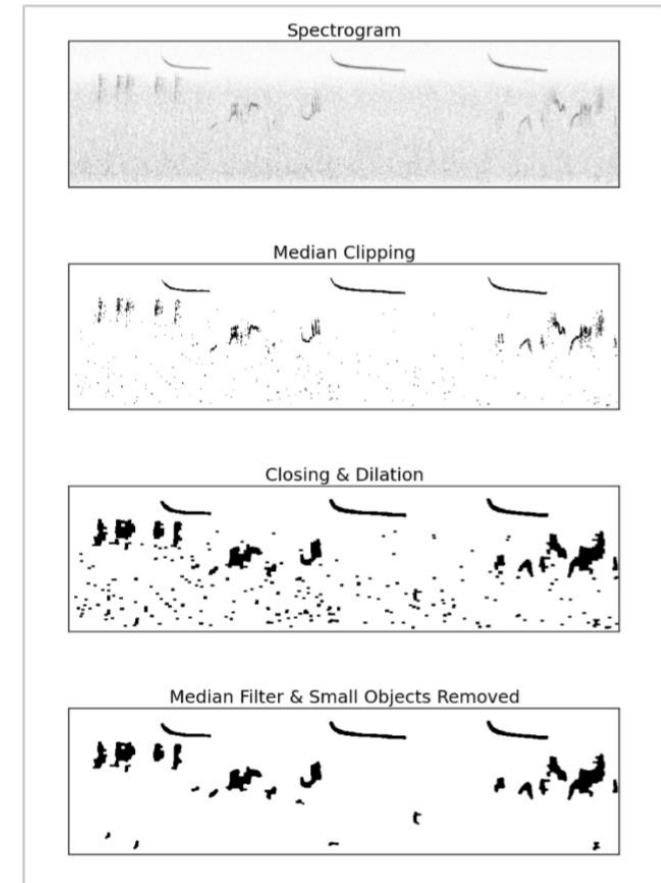
Methodology
References

- **2 : Multi-class classification with CNN**
- referred from Sprengel et al. (2016)
- preprocess: wav → STFT → median clipping
- signal/noise separation
- dividing the spectrograms into chunks
- classifier: 5 CNN + 1 dense + soft-max
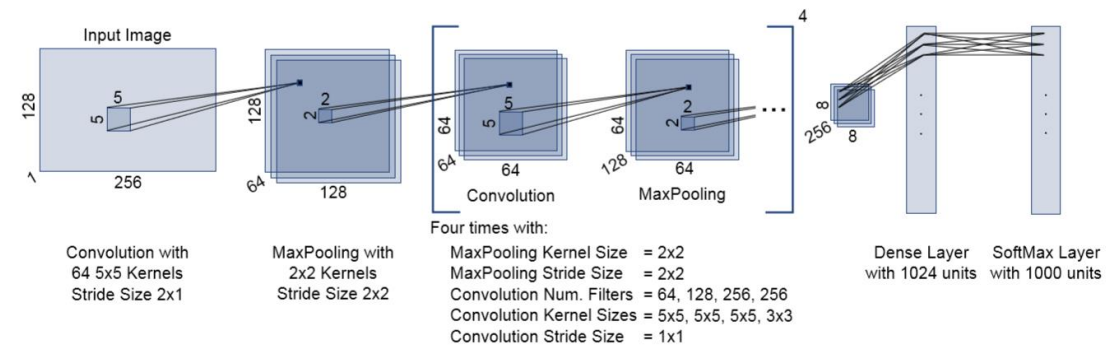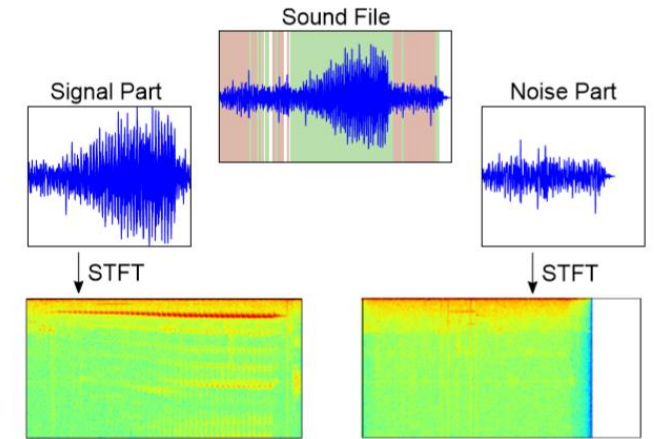
Methodology

- signal / noise separation

- STFT → normalize → smooth → median clipping

- mask vector for separate audio files

Methodology
Feature generation

- dividing the spectrograms into chunks

- fixed size input of 512 pixel (~3 sec)

- each chunk is unique sample (because separated signal / noise)

- multiple prediction per file → average to final result

Methodology
Network architecture

- 5 CNN and 1 dense with max pooling and ReLU activation function

- soft-max layer → multi-class classification

- batches of 8 examples

- train : test = 90 : 10
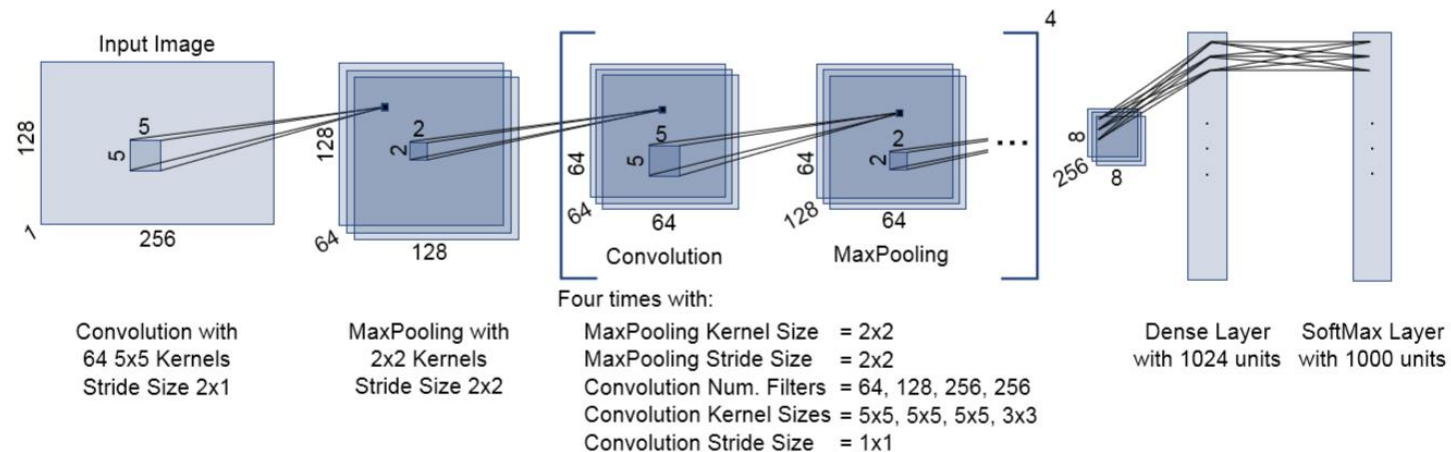
Methodology

- **3 : Multi-label classification with CNN**

- **improved from model 1 and 2**

- preprocess: wav → STFT → median clipping

- signal/noise separation

- dividing the spectrograms into chunks

- classifier: 5 CNN + 1 dense **+ sigmoid laye**

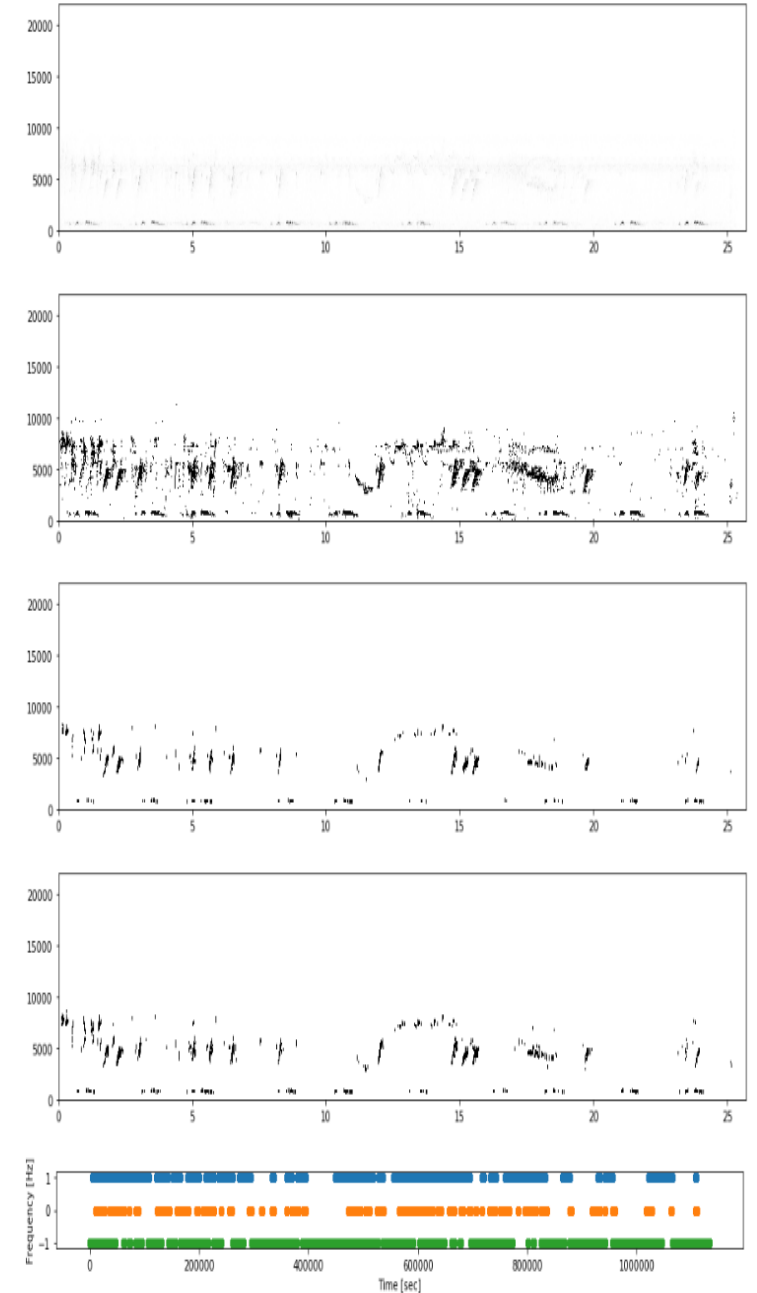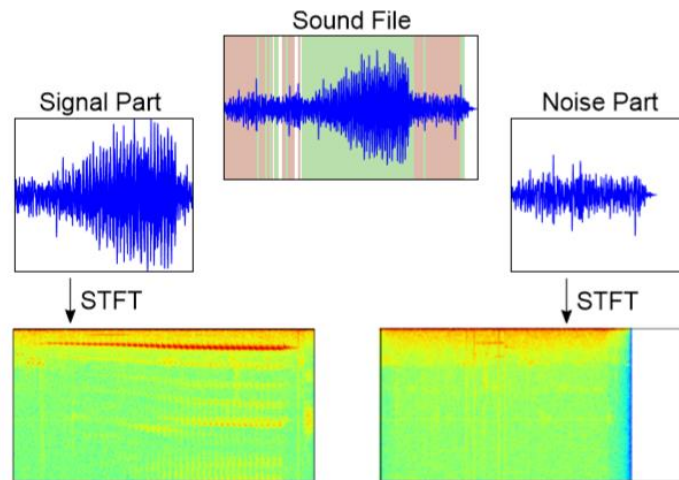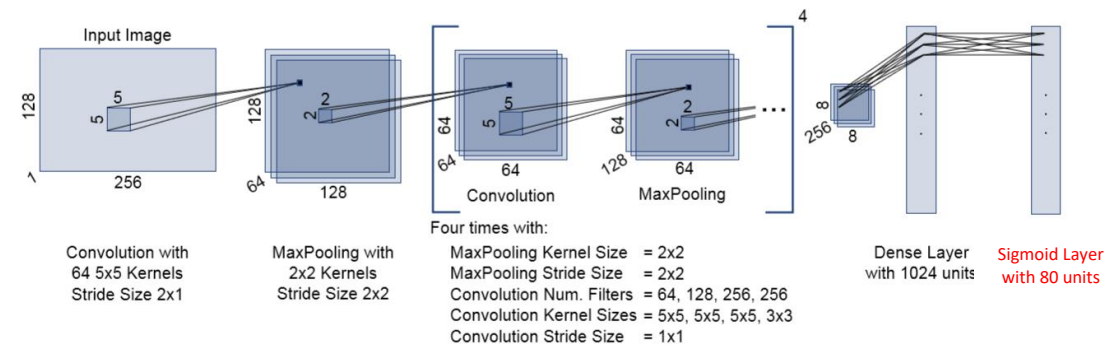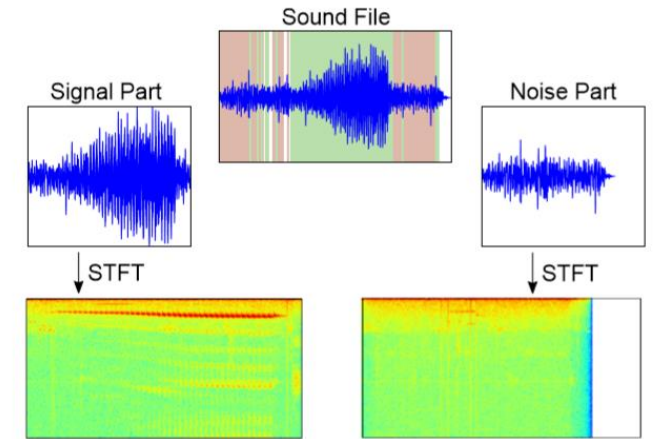# Methodology
Network architecture

- sigmoid layer → multi-label classification

- loss function → binary cross entropy
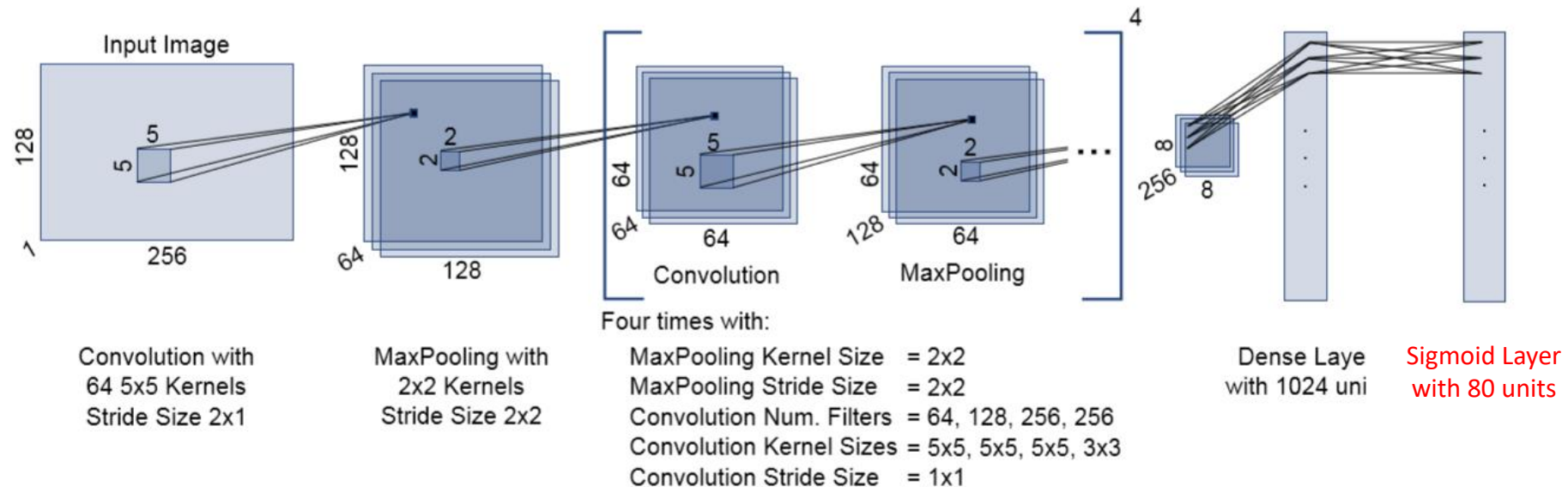
RESULTS

MODEL EVALUATION
METHODS

ACCURACY &
CONFUSION MATRIX
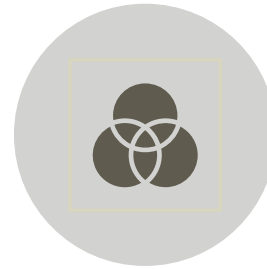
FINAL MODEL
PERFORMANCE

ERROR ANALYSIS

Results
Model evaluation methods

- accuracy on chunks

- accuracy on files
  - average mode with confidential values
  - only one true

- confusion matrix

- other metrics
  - average precision, AUC
  - f1, precision, recall, hamming loss
  - coverage, ranking AP, ranking loss

Results
Accuracy

- model testing with 11 bird class

- model selection with 80 bird class

Results
Accuracy

Accuracy on 11 class

	0.9174	0.8442	0.8052
	model 1	model 2	model 3

feature generate long time   longer time training

Results
Accuracy

Accuracy on 11 vs 80 class

	11 class	80 class
model 1	0.9174	0
model 2	0.8442	0.2404
model 3	0.8052	0.4322
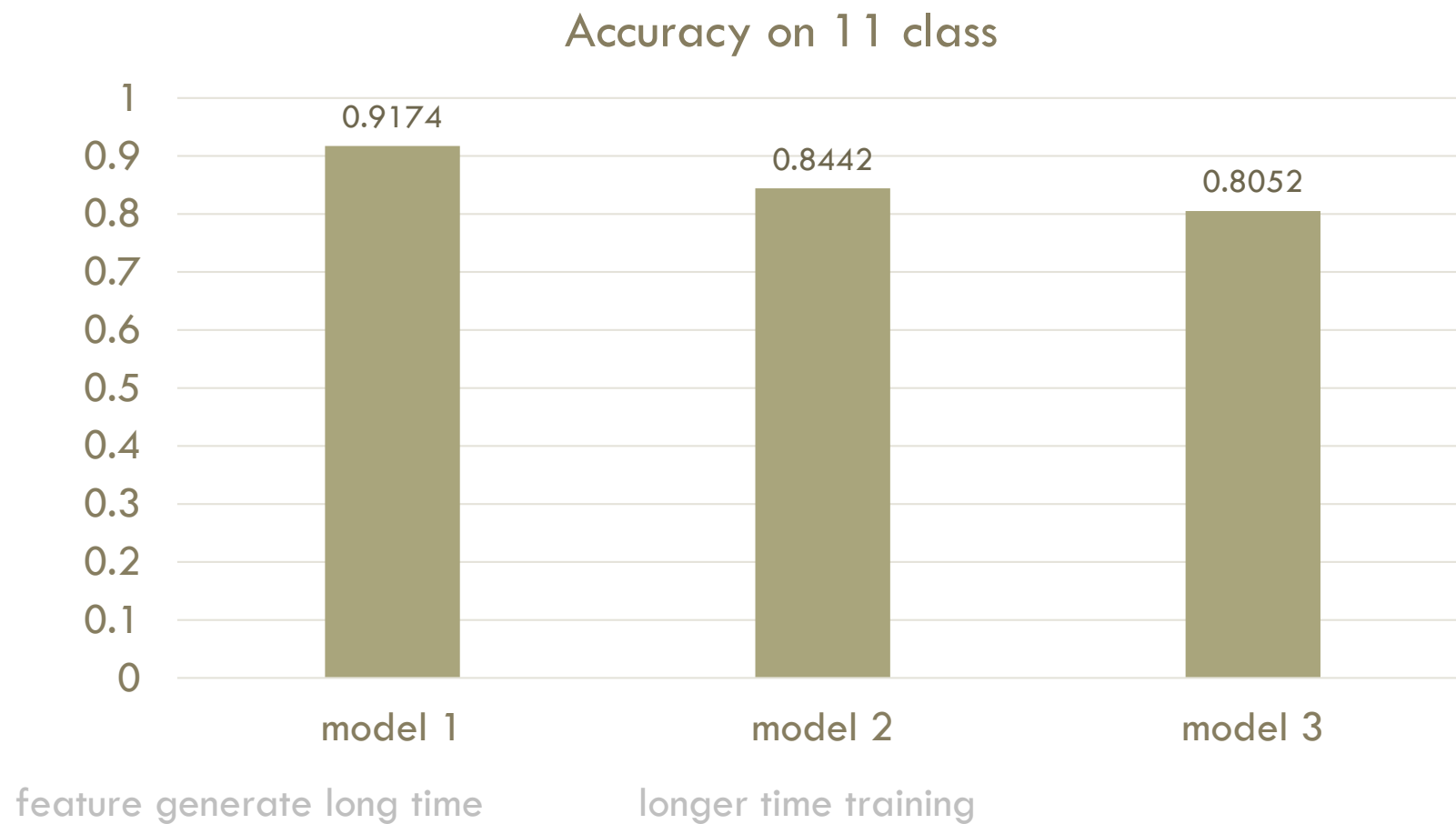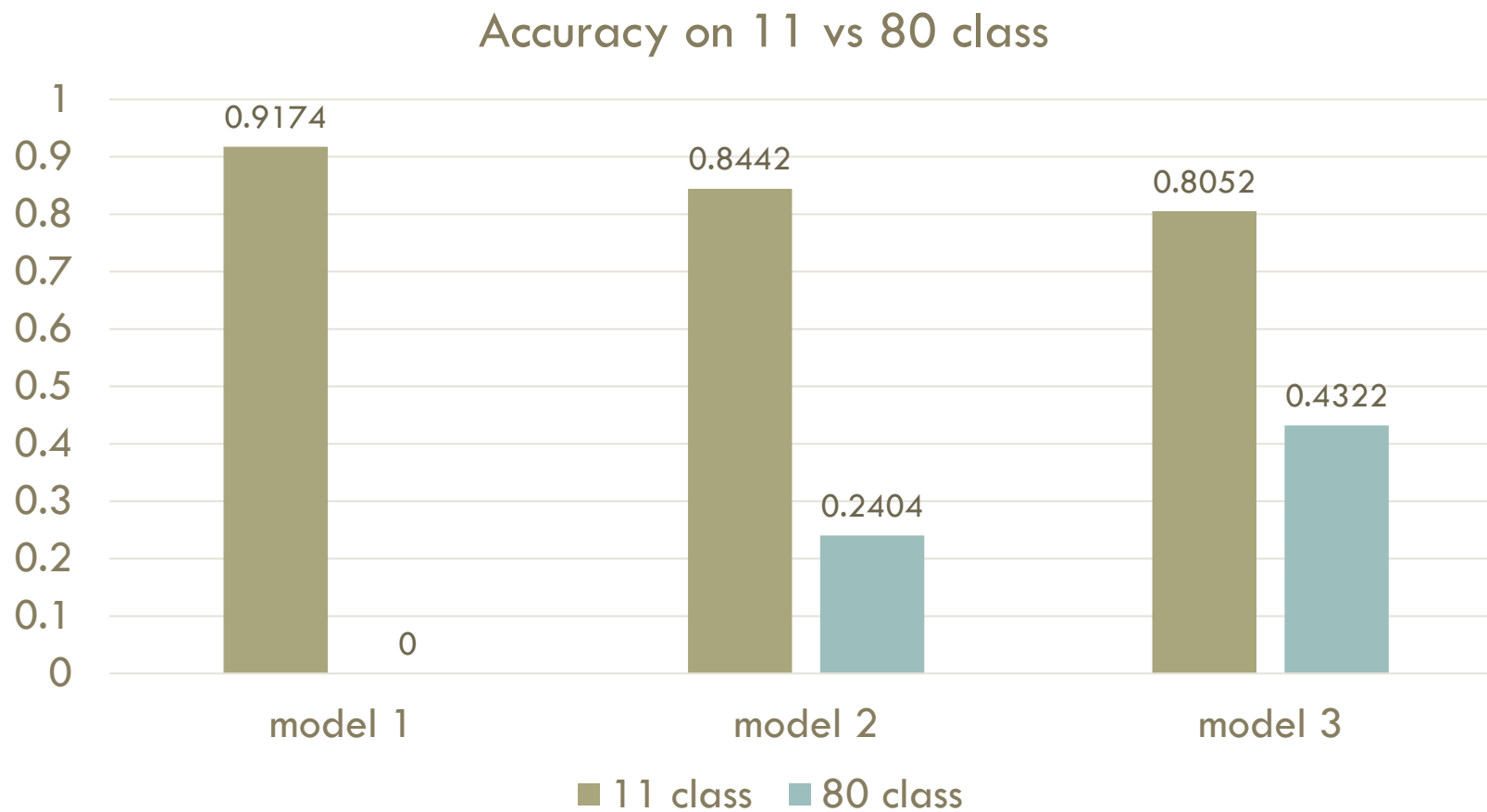
Results

Confusion matrix

Model 2

Model 3

Results
Final model performance

- average precision, f1-score, precision, recall

    80 class → 40 class with model 3
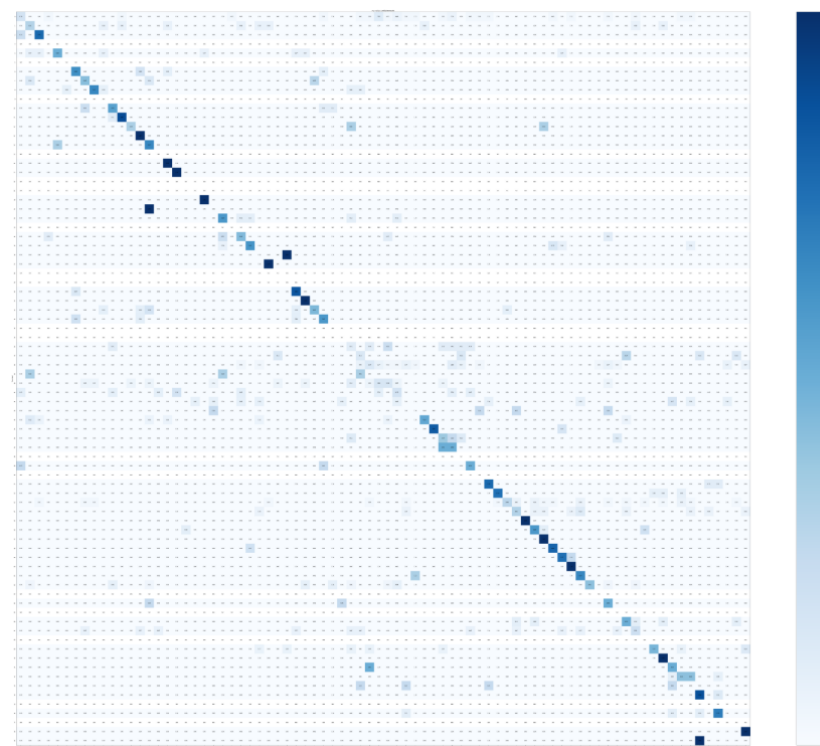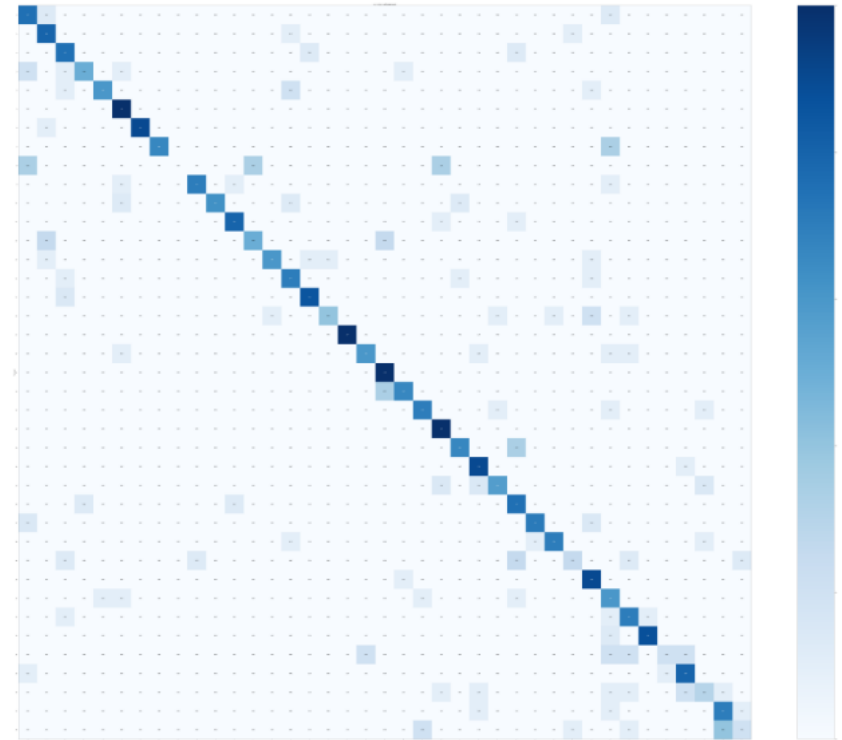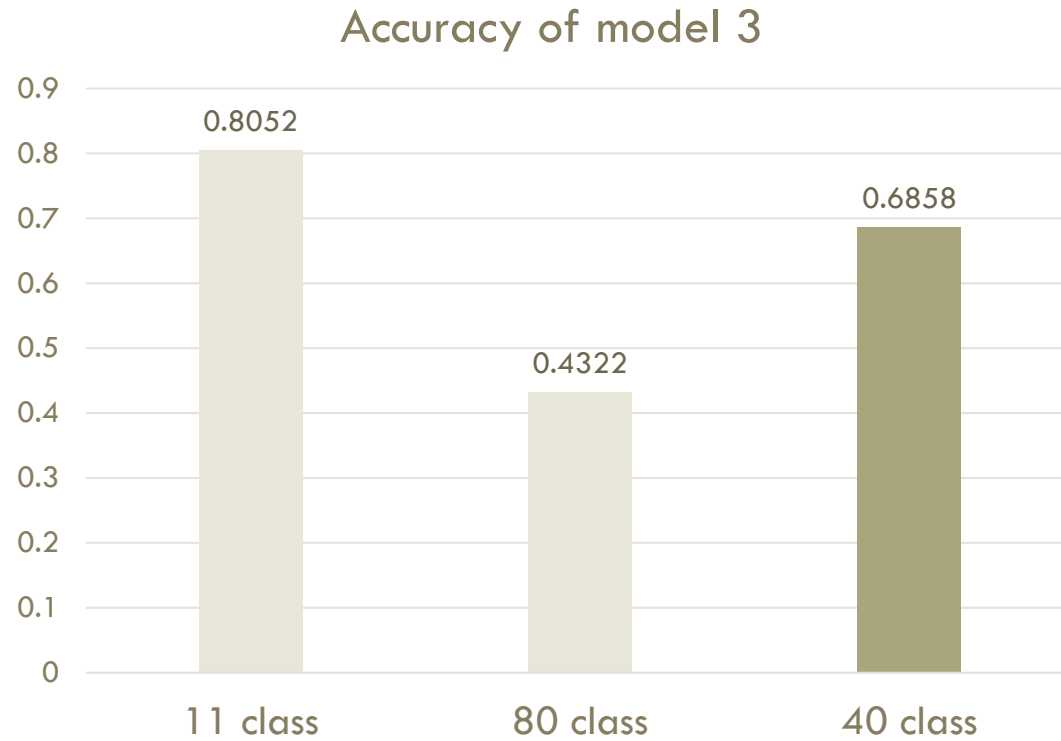
Results
Final model performance

Accuracy of model 3

11 class	80 class	40 class
0.8052	0.4322	0.6858

Results

Final model performance

metric	values
macro-AUC	0.9399
micro-AUC	0.9402
average precision	0.7242
coverage	3.4924
ranking AP	0.7770
ranking loss	0.0656

metric	values
hamming loss	0.3142
f1	0.6565
precision	0.6865
recall	0.6652
accuracy	0.6858

Results
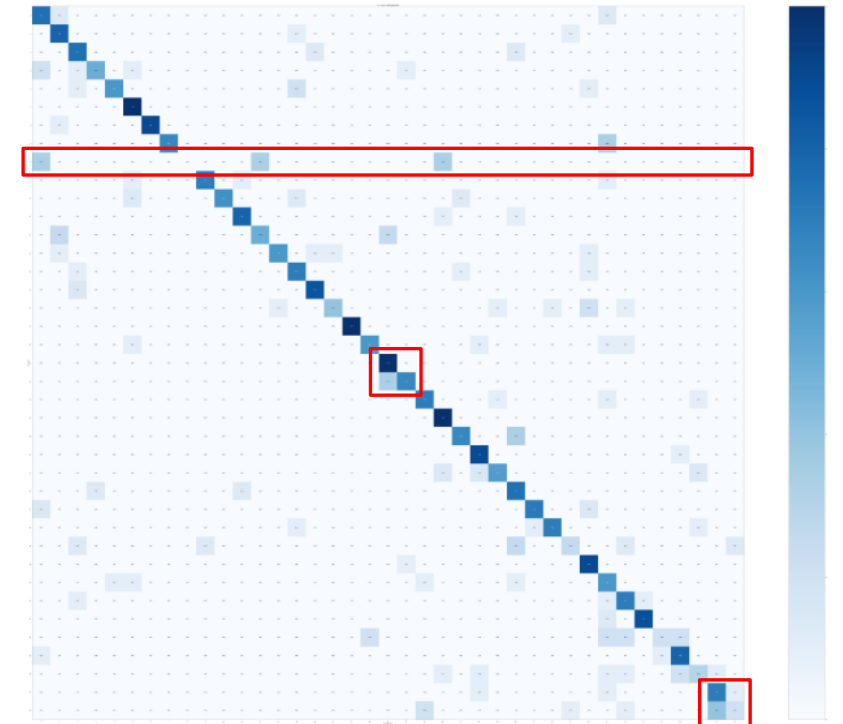Final model performance

- if show all class that has confidential value > 5%

  → show ~4 class on average
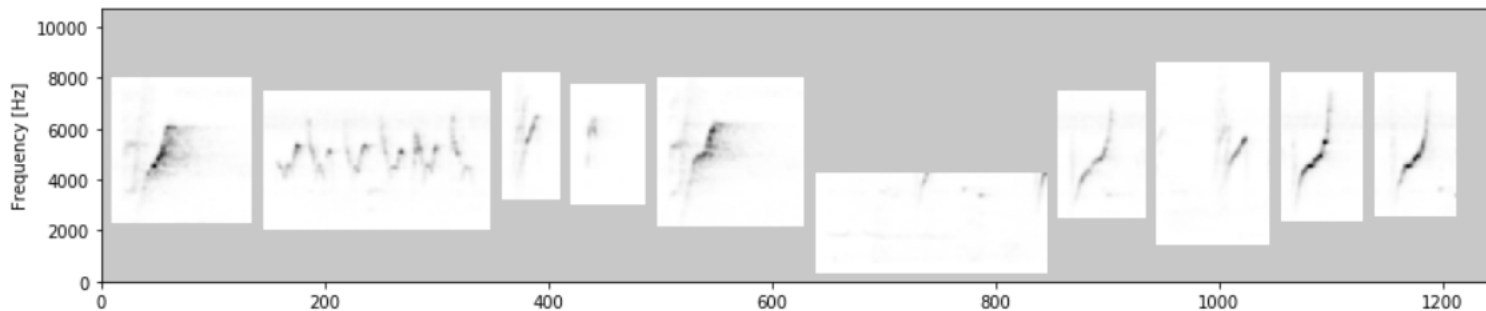
  → accuracy 87%

Results
Error analysis

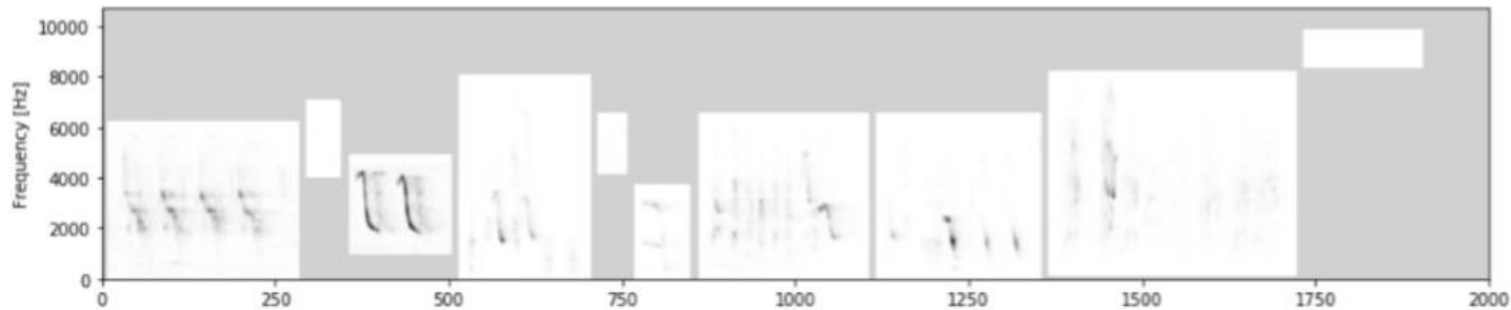- for 40 class, only 1 class นกแอ่นบ้าน has f1-score = 0

  (only 33 wav file in dataset)

- some common mistake
  - นกตีทอง ➔ นกกะรางหัวขวาน
  - นกกระจอกบ้าน ➔ นกกระจอกใหญ่, นกนางแอ่นบ้าน
  - นกปรอดเหลืองหน้านวล ➔ นกปรอดเหลืองหัวจุก
  - นกกินปลีอกเหลือง ➔ นกกินปลีดำม่วง

Results
Error analysis

- **the low-performance class e.g.** นกเอี้ยงสาริกา, นกกินปลือกเหลือง

Results
Error analysis

- Multi-label classification with random forest

  - feature generation take long time and not scalable

- Multi-label classification with CNN

  - sensitive to class imbalance

    $\rightarrow$ only 8/80 class after 1000 epoch (acc. 24%)

  - longer time training

DISCUSSION

LIMITATIONS

FUTURE WORK
SUGGESTIONS

Discussion
Limitations

- restricted species due to conservation concerns

- time and resource constraints

- microphone quality & noise handling

Discussion

Future work suggestions

- data augmentation → noise handling + class imbalance

- subclass/call-song for some species

- add other similar sounds e.g. squirrel frog cricket

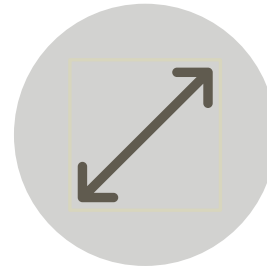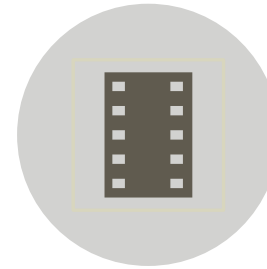- recording noise handling

- bird song detection
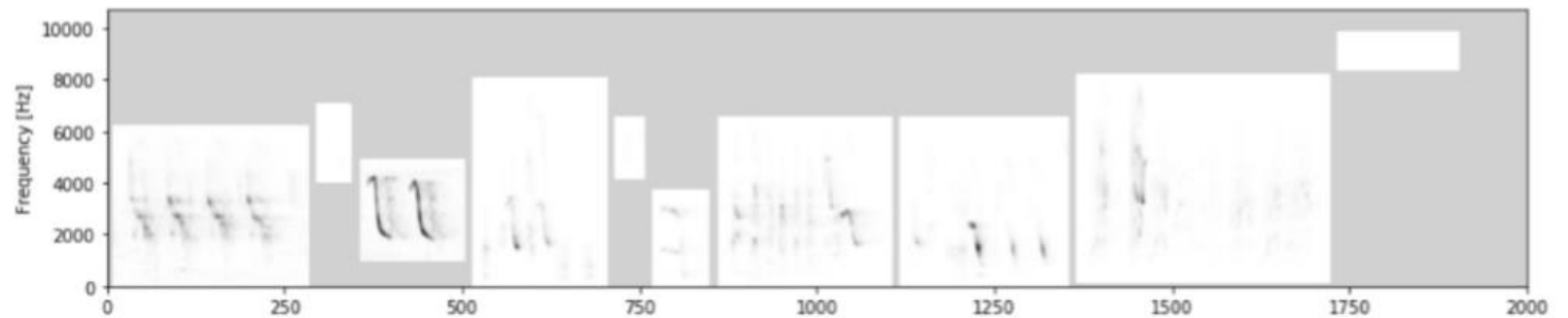
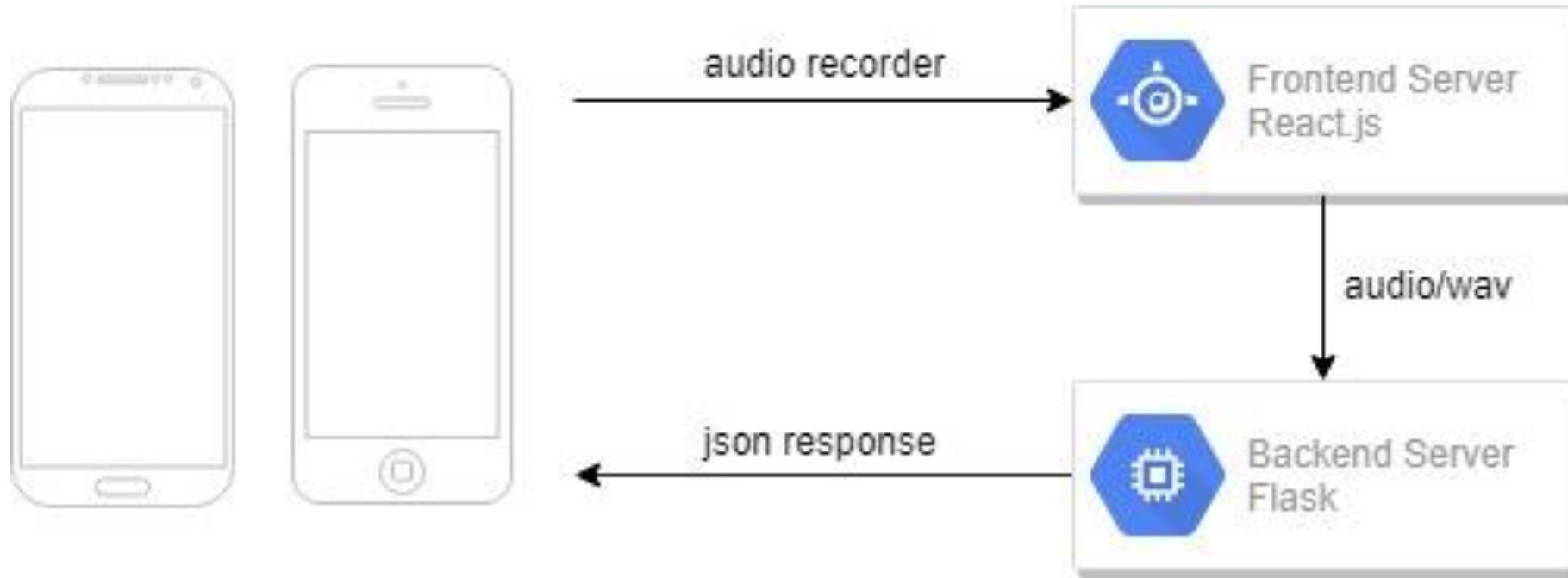# DEPLOYMENT & DEMO

DEPLOYMENT DIAGRAM

DEPLOYMENT TOOLS

SAMPLE UI

LIVE DEMO
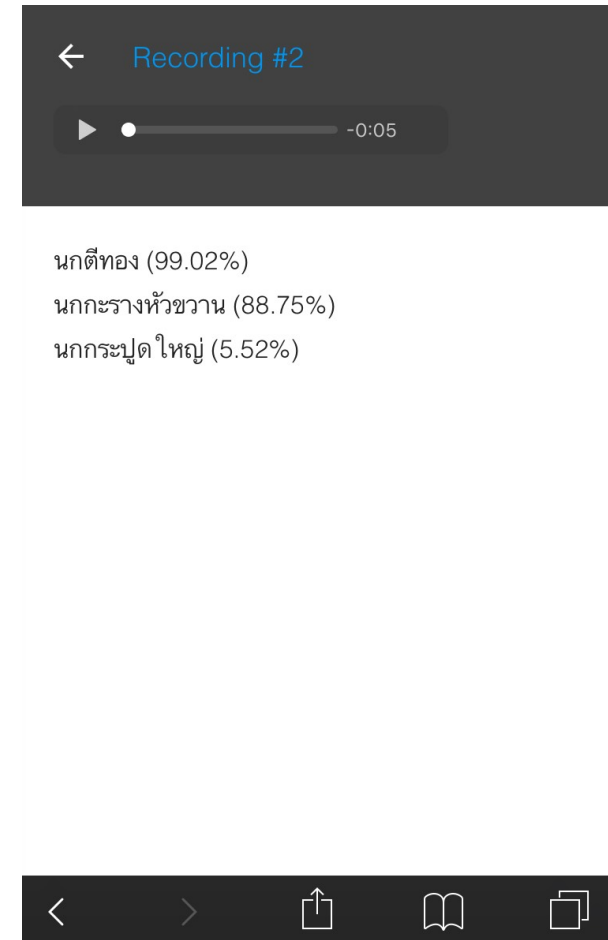
Deployment
Diagram

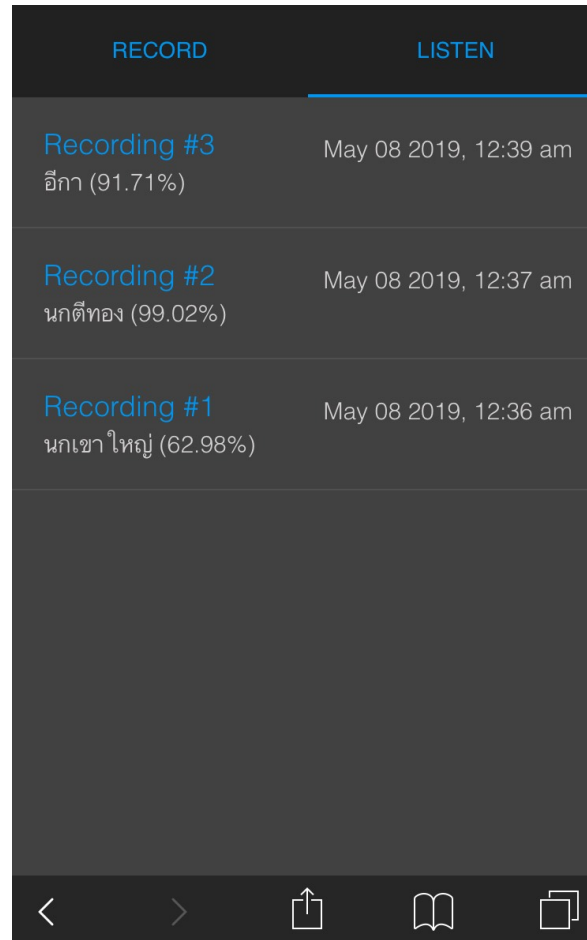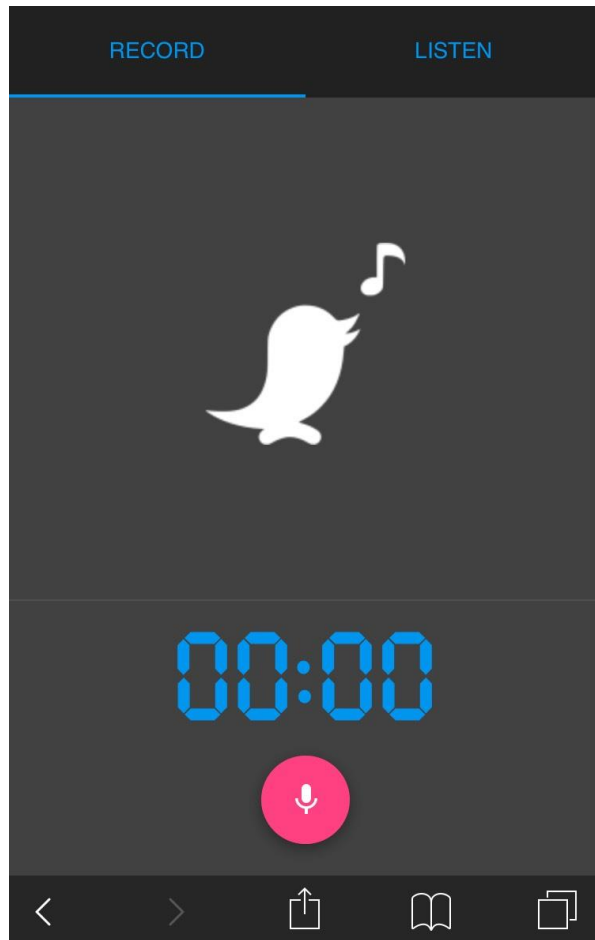# Deployment
Tools

- recordRTC

- react.js

- flask

- localforage cache

- google cloud platform

Deployment
Sample UI

Demo

- https://www.youtube.com/watch?v=vwPgfybdAHY นกอีวาบตั๊กแตน

- https://www.youtube.com/watch?v=ZAsUX0lN6n4 อีกา

- https://www.youtube.com/watch?v=Hfu6b29FhtA นกตีทอง

- https://www.youtube.com/watch?v=UoWQGzJUXJ4 นกเขาใหญ่

- https://www.youtube.com/watch?v=pM5ll1VLceo นกเขาชวา

- https://www.youtube.com/watch?v=Qpjz2oE_Awg นกกระจิ๊ดธรรมดา

- https://www.youtube.com/watch?v=9BMb4NseG4A นกเอี้ยงสาริกา

- https://www.youtube.com/watch?v=qsBtyLAlx_w นกกินปลีอกเหลือง

Q&A