PROGRESS 3: CNN - BIRDCLEF 2016

BIRD SONG IDENTIFICATION

Thanyaporn Phinthuphan

21 Feb 2019

Outline

- Recap
- Literature Review
- Methodology
- Preliminary Result
- Next step

Recap

Introduction

- trained on 11 bird sound class with manual labeled 120 wav file (10 per class)
- → download all sound of 11 class with API → 1,847 file (100-200 per class)
- goal: find the model that work on this data before append number of class

Recap

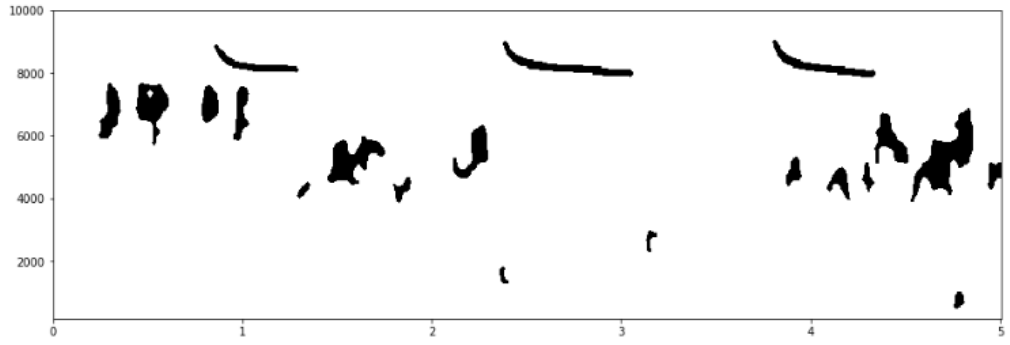
Training data problem

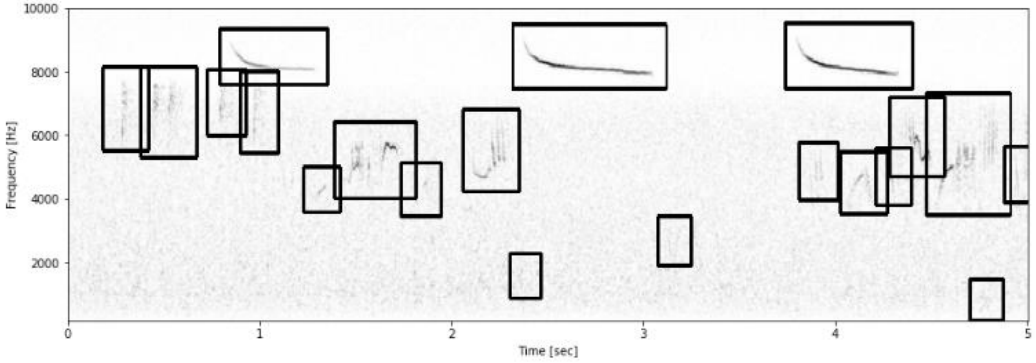
- label only 1 class per file but want multi-label result
- mp3 to wav stereo (multi-channel) sound: node-lame → mean
- quality of sound: using **all** / only A (~701) / A & B (~1,600)
- sound type (call/song): **combine** / separate class
- API download problem

Recap

Last algorithm problem

- former model : feature = max correlation with segments in spectrogram
- 4,143 feature → 67,991 feature
- prediction time ~ 1 times of sound length → ~ 20 times
- too slowly !!!!!





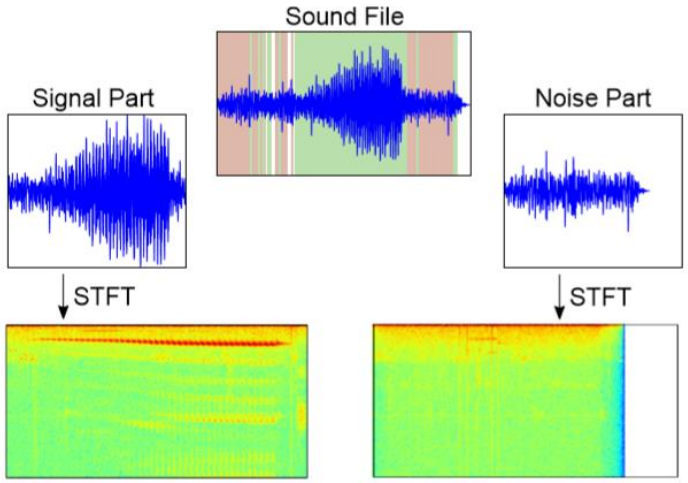
Literature review

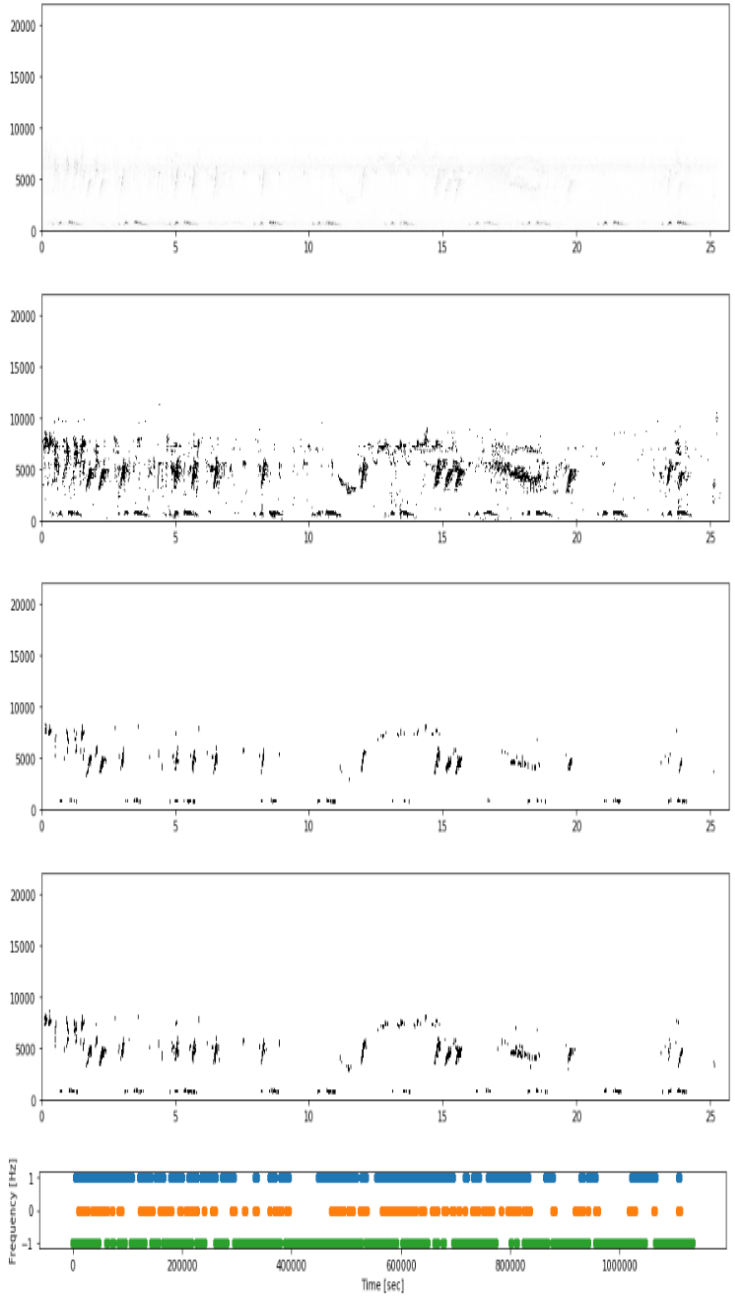
- Audio Based Bird Species Identification using Deep Learning Techniques (Elias Sprengel and teams, 2016)
- winning solution of BirdCLEF 2016
- data 33,000 recording with 999 difference species (25 files/class)
- mean average precision score of 0.686 (include background species 0.555)

Methodology

Feature generation

- 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

Methodology

Feature generation

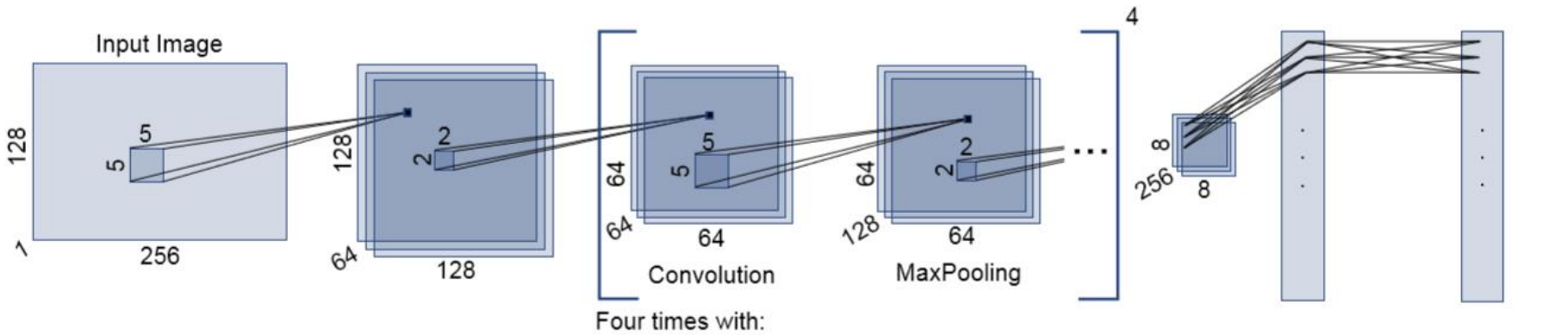
- data augmentation → avoid overfitting
- time shift
- pitch shift
- combining same class
- adding noise

	MAP (FG only)	MAP (FG & BG)
Baseline	0.842	0.728
w/o Noise	0.831	0.731
w/o Same Class	0.839	0.730
w/o Time Shift	0.801	0.701
w/o Pitch Shift	0.828	0.725
w/o Noise and Same Class	0.768	0.661

Methodology

Network architecture

- 5 CNN and 1 dense with max pooling and ReLU activation function
- batches of 8 / 16 examples
- train : test = 90 : 10



Preliminary result

- signal part audio file (5 Gb) on Google drive
- trained only A/B quality files
- using Pytorch on Google Colab
- there was some bug…


```
Saving..
1 / 20 Batch Loss: 8.056486402665199
Saving..
2 / 20 Batch Loss: 8.033975657492153
Saving..
3 / 20 Batch Loss: 8.033110983582379
4 / 20 Batch Loss: 8.03337495861993
Saving..
5 / 20 Batch Loss: 8.032938653365814
6 / 20 Batch Loss: 8.068400501552459
7 / 20 Batch Loss: 8.110126499089532
8 / 20 Batch Loss: 8.11012649685261
9 / 20 Batch Loss: 8.110126496106968
10 / 20 Batch Loss: 8.11012649685261
11 / 20 Batch Loss: 8.110126495361328
12 / 20 Batch Loss: 8.110126495361328
13 / 20 Batch Loss: 8.110126495361328
14 / 20 Batch Loss: 8.110126495361328
15 / 20 Batch Loss: 8.110126495361328
16 / 20 Batch Loss: 8.110126495361328
17 / 20 Batch Loss: 8.110126495361328
18 / 20 Batch Loss: 8.110126495361328
19 / 20 Batch Loss: 8.110126495361328
20 / 20 Batch Loss: 8.110126495361328
== Training Finish ==
```

Next Step

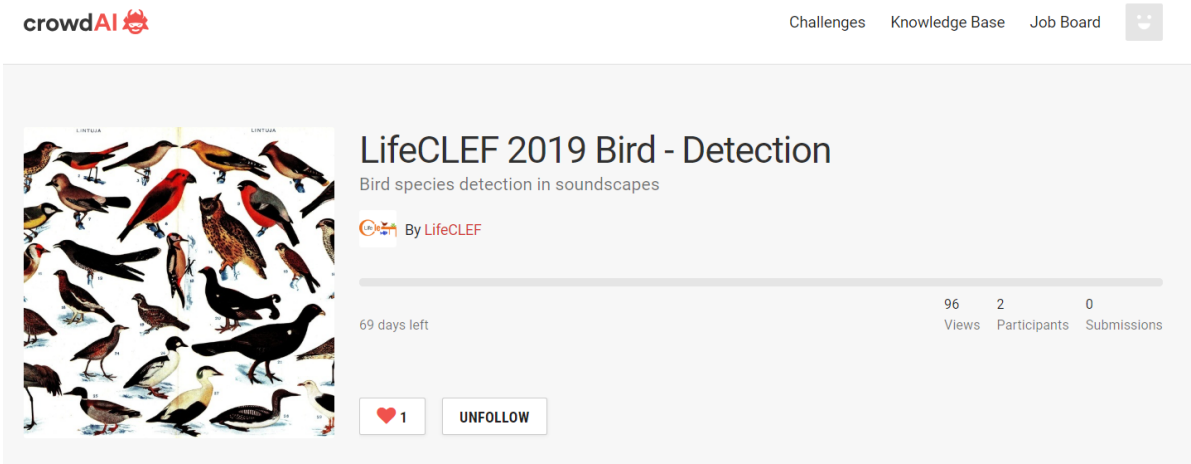
- find the best model on 11 class: good AUC / recall / prediction time
- write UI to use on mobile phone
- define new scope and test
- test with real data (manual record / data from other website)
- test in the park on mobile phone

Experiment

BirdCLEF 2019

- Task1 - Bird species detection in soundscapes
- Task2 - Birds counting in soundscapes

- Jan 2019: training data release on www.crowdai.org → **updated on 20 Feb !!**
- March 2019: test data release
- 1st of May 2019: deadline for submission of runs by the participants
- 13th of May 2019: release of processed results by the task organizers
- 24th of May 2019: deadline for submission of working notes papers



Q&A