

General-purpose audio tagging of Freesound content with AudioSet labels

DCASE 2018 Task 2

Eduardo Fonseca, Manoj Plakal, Frederic Font, Daniel P. W. Ellis

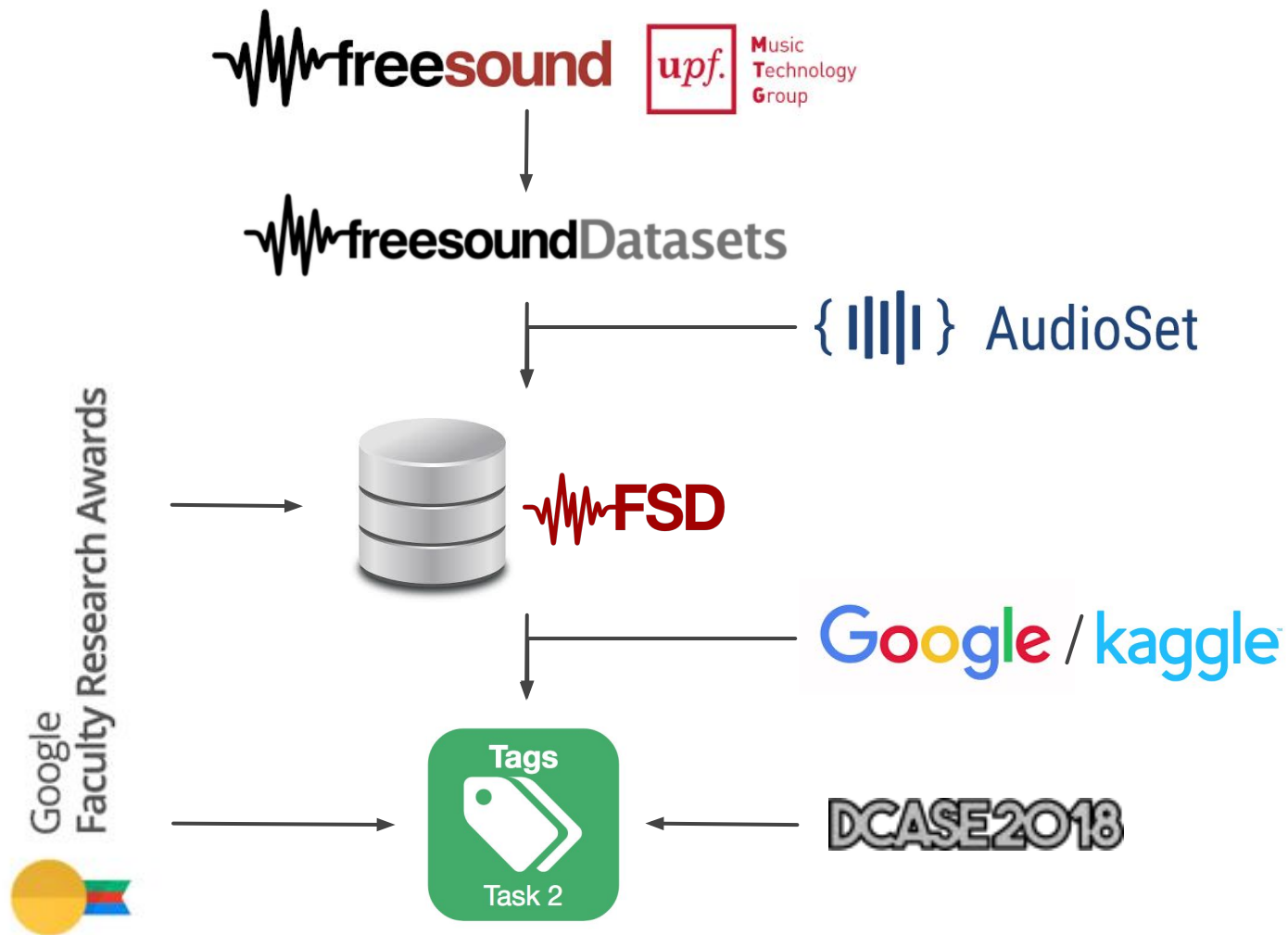


**Universitat
Pompeu Fabra**
Barcelona

MTG
Music Technology
Group



A bit of history

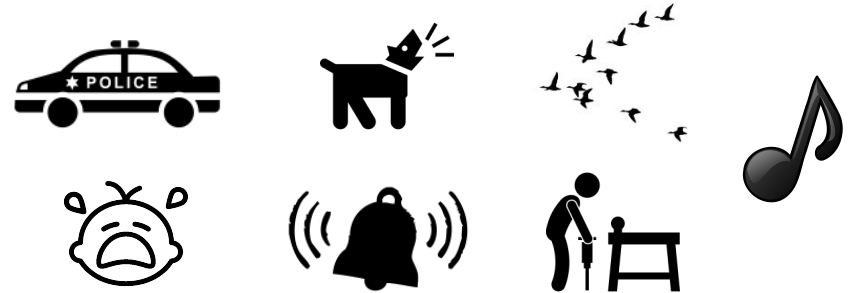


Motivation - Task setup

- previous tasks: domain-specific



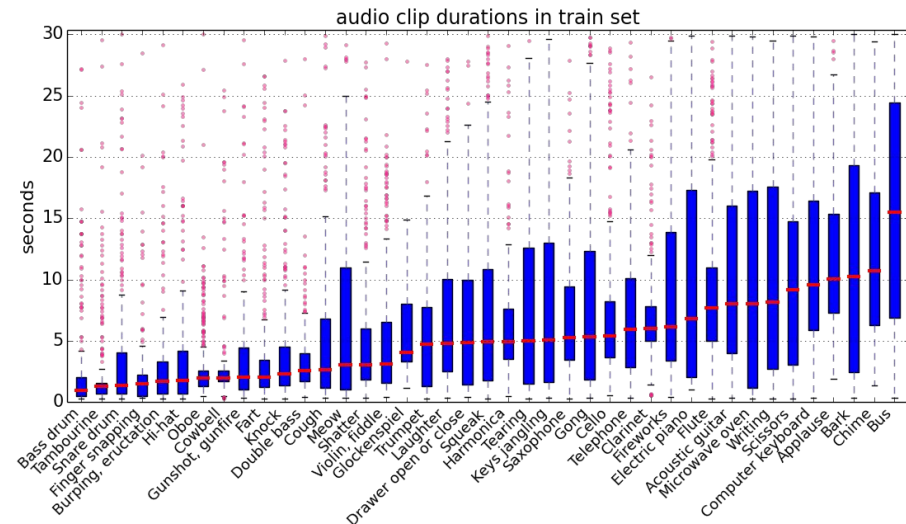
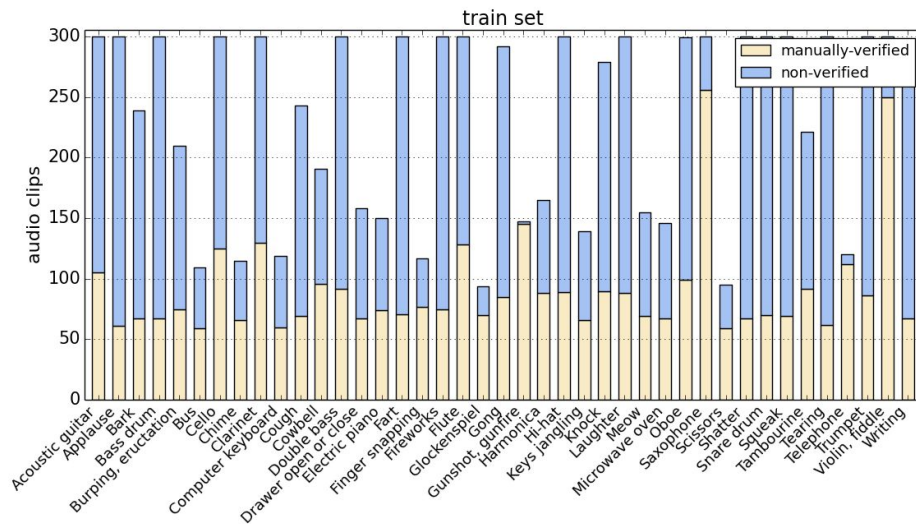
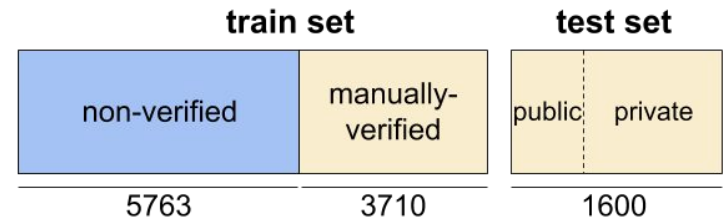
- our task: general-purpose



- goal: recognize the category among 41 categories
- Simplicity: multi-class classification problem -> audio tagging
- annotations of varying reliability

Dataset: FSDKaggle2018

- 11k clips, 41 categories
- 18h of training data



Issues encountered

- building the dataset:
 - our first dataset built of Freesound
 - early snapshot of the larger FSD
- engagement:
 - 558 teams and 5684 entries
 - 20 DCASE submissions
 - Need to work on getting more submissions to DCASE

Task 2 Takeaways

- **log-mel energies**, waveform, MFCC, CQT, ...
- **Mixup** for data augmentation
- Mainly **CNN/CRNN**: **VGG**, **DenseNet**, **ResNe(X)t**, gated convs
- Heavy usage of **ensembles** (2 -> 30)
- Noisy labels: **Semi-supervised learning** (self-supervision) & robustness measures (**label smoothing**, **loss masking**)

Open knowledge

5			Huge thanks to Kaggle, DCASE organizers and everyone Il-Young Jeong 3 months ago	last comment by Peter A. I. Forsyth 1mo ago	6
3			8th place solution Sainath Adapa 3 months ago	last comment by Sainath Adapa 3mo ago	0
0			SVM using MFCC features anmour last run 7 months ago	last comment by Kevin Degila 17d ago	0
264			Beginner's Guide to Audio Data Zafar last run 7 months ago	last comment by Avi Shapiro 1mo ago	40
13			4th solution Kele 3 months ago	last comment by scut_wzm 3mo ago	2
4			Best Single Model: 0.915 (Private LB score) with nearly zero pre... Gyat 4 months ago	last comment by thomeou 3mo ago	4
3			Using large amount of non-deep features George Oblapenko 3 months ago	last comment by daisukelab 3mo ago	1

Branch: master
New pull request

Create new file
Upload files
Find file
Clone or download

Latest commit 5a3c9af on Sep 29

dcase_task2	added MIT LICENSE, added live-audio-tagger	2 months ago
docs	self_verification figure updated	2 months ago
.gitignore	first commit	3 months ago
LICENSE	added MIT LICENSE, added live-audio-tagger	2 months ago
README.md	paths in githubpage and readme updated, latex formulas removed	3 months ago
requirements.txt	readme, requirements.txt and setup.py added to project	3 months ago
setup.py	readme, requirements.txt and setup.py added to project	3 months ago

README.md

Training General-Purpose Audio Tagging Networks with Noisy Labels and Iterative Self-Verification

This repository contains the corresponding code for the 2nd place submission to the first [Freesound general-purpose audio tagging challenge](#) carried out as Task 2 within the [DCASE challenge 2018](#).



Zafar

Beginner's Guide to Audio Data

last run 7 months ago · IPython Notebook HTML · 18,941 views
using data from [multiple data sources](#) · ...

264
voters



Notebook Code Data (4) Output Comments (40) Log Versions (15) Forks (624)

Fork Notebook

Tags beginner data visualization deep learning cnn multiple data sources

1. Exploratory Data Analysis

- Loading data
- Distribution of Categories
- Reading Audio Files
- Audio Length

2. Building a Model using Raw Wave

- Model Description
- Configuration
- DataGenerator class
- Normalization
- Training 1D Conv
- Ensembling 1D Conv Predictions

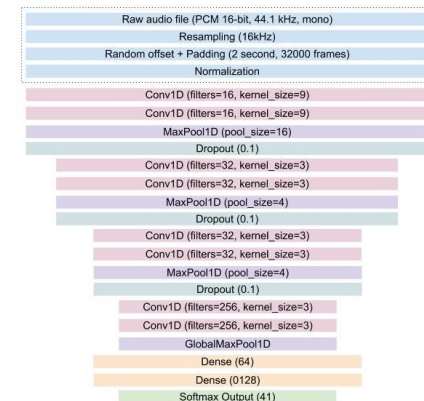
2. Building a Model using Raw Wave

We will build two models:

- The first model will take the raw audio (1D array) as input and the primary operation will be Conv1D
- The second model will take the MFCCs as input. (We will explain MFCC later)

Keras Model using raw wave

Our model has the architecture as follows:



General-purpose audio tagging of Freesound content with AudioSet labels

DCASE 2018 Task 2

Thank you!

<https://www.kaggle.com/c/freesound-audio-tagging>

<http://dcase.community/challenge2018/task-general-purpose-audio-tagging-results>

Eduardo Fonseca, Manoj Plakal, Frederic Font, Daniel P. W. Ellis



Universitat
Pompeu Fabra
Barcelona

MTG
Music Technology
Group

