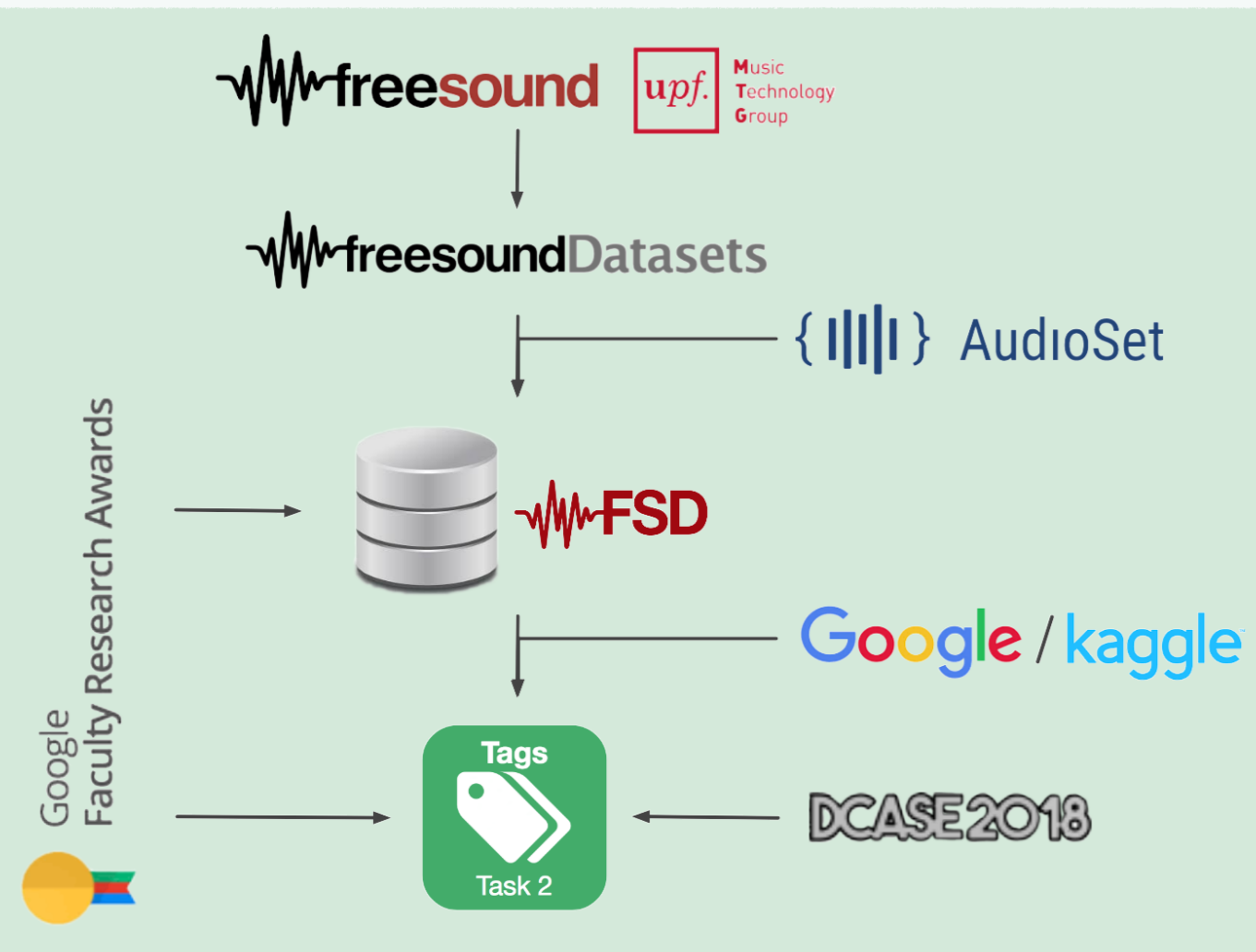


T2 General-purpose audio tagging of Freesound content with AudioSet labels

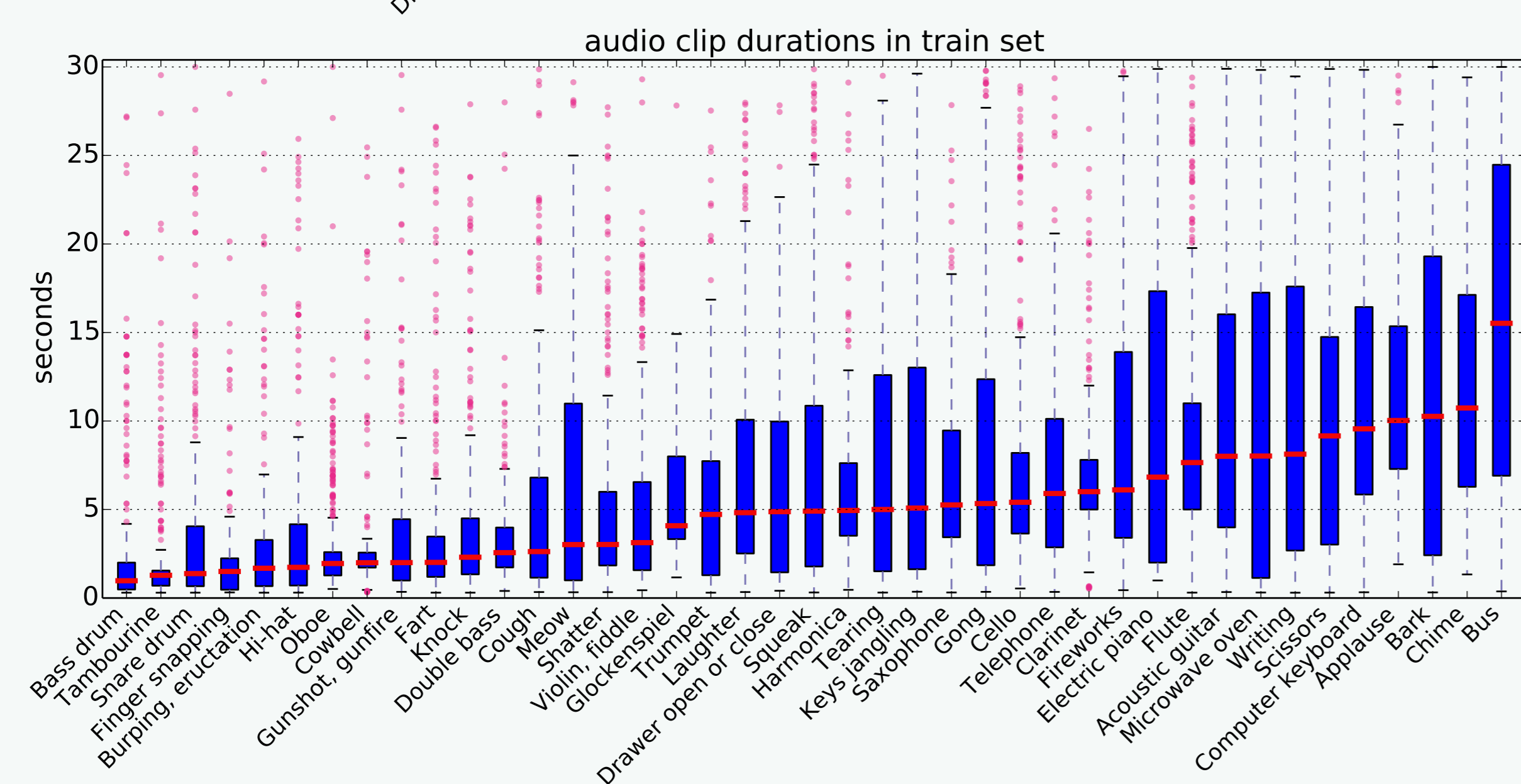
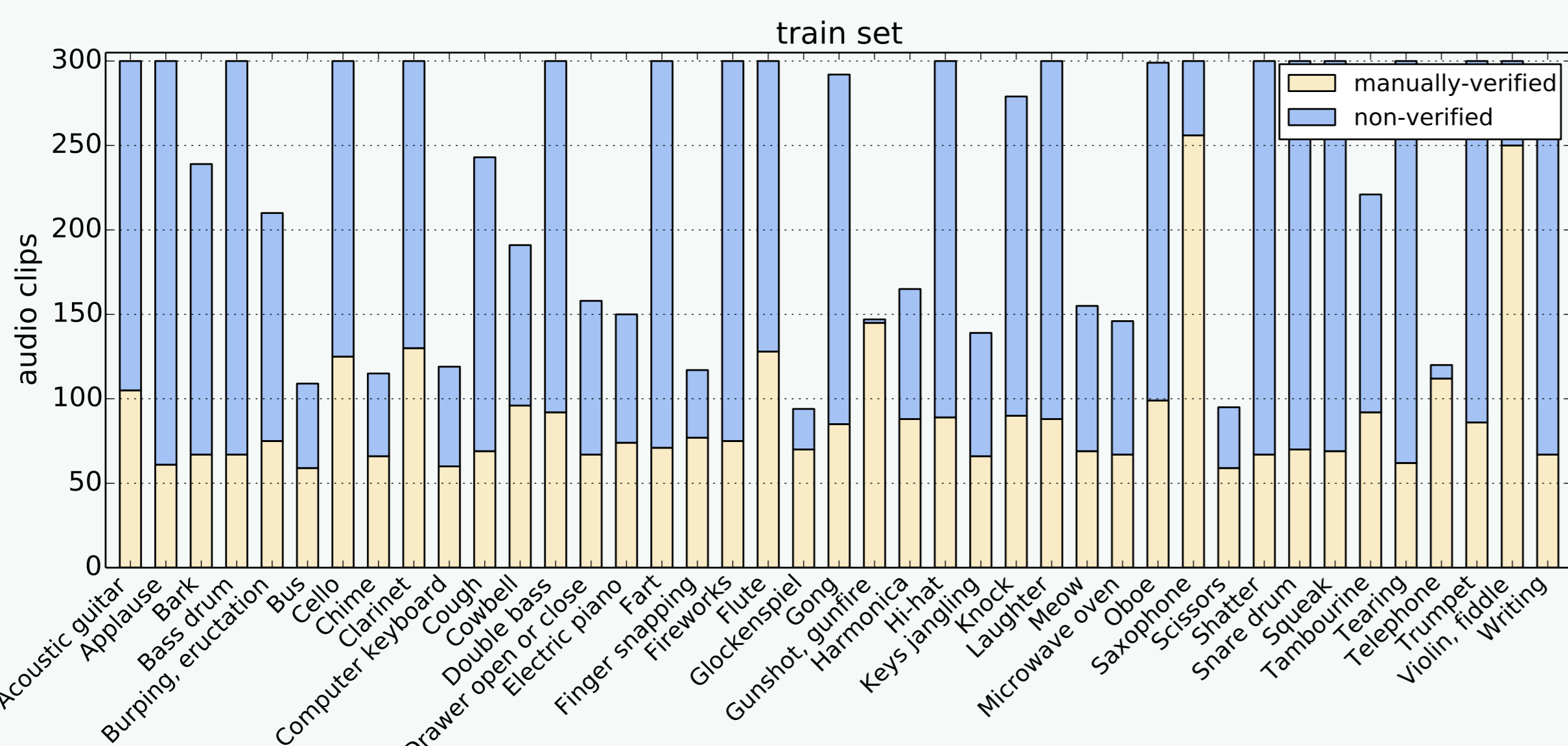
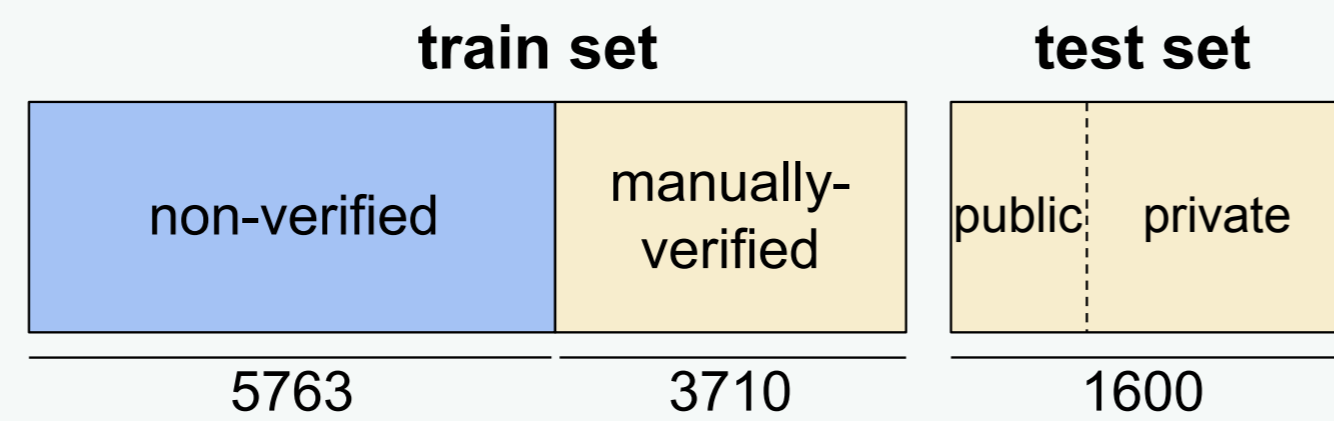
Motivation



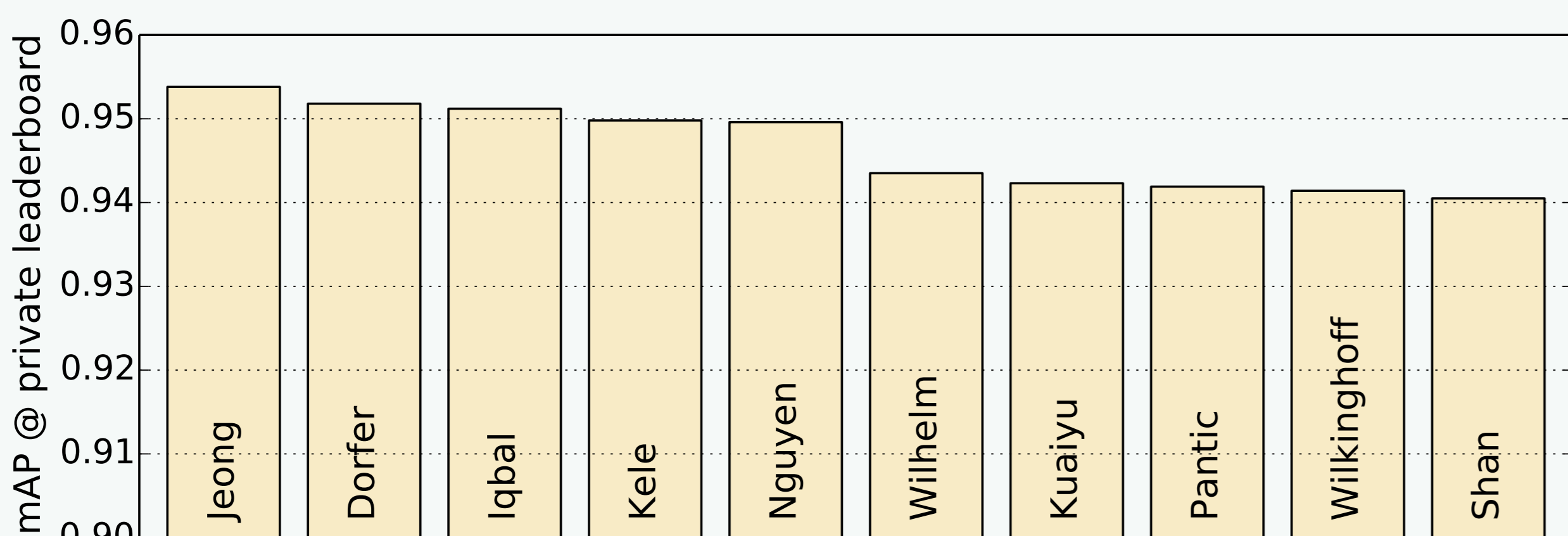
- ▶ from domain-specific tasks (domestic audio, smart cars)
- ▶ to general-purpose sound event recognizers

Dataset: FSDKaggle2018

- ▶ 41 classes, 11,073 files
- ▶ 18h of training data
- ▶ one ground truth label/clip



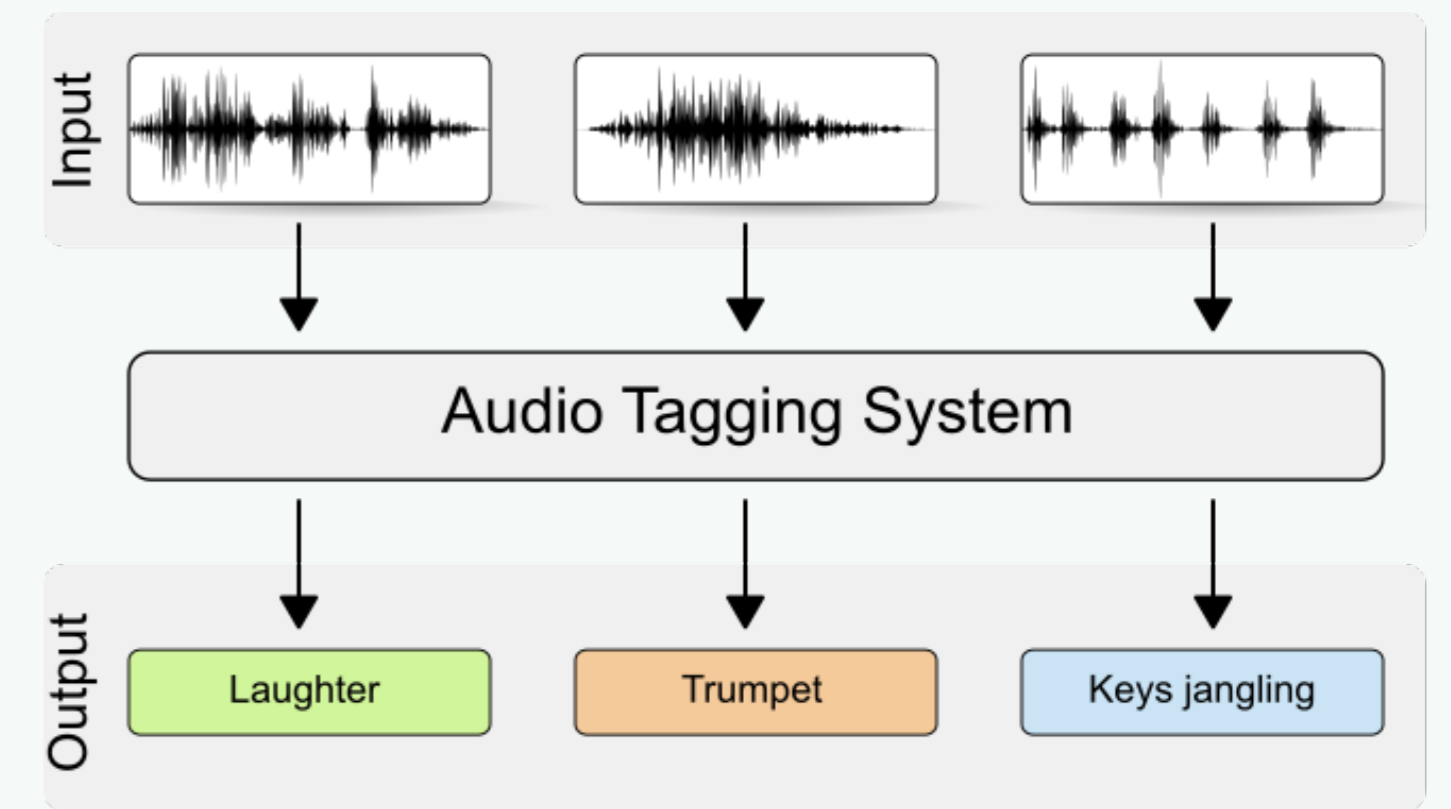
Results, Top-10



System	Features	Classifier	mAP@3
Jeong	log-mel energies, waveform	CNN, ensemble	0.9538
Dorfer	mel-weighted spectrogram, log-power spectrogram	CNN, ensemble	0.9518
Iqbal	log-mel energies	CNN, GCNN, CRNN, GCRNN	0.9512
Kele	log-mel energies	CNN, GBM, ensemble	0.9498
Nguyen	log-mel energies	CNN, ensemble	0.9496
Wilhelm	log-mel energies	CNN, ensemble	0.9435
Kuaiyu	log-mel energies, MFCC, waveform	CNN, ensemble	0.9423
Pantic	mel-spectrogram, CQT	CNN, ensemble	0.9419
Wilkinghoff	PLP, MFCC, mel-spectrogram, waveform	CNN, ensemble	0.9414
Shan	log-mel energies, MFCC	CNN, ensemble	0.9405

Task Description

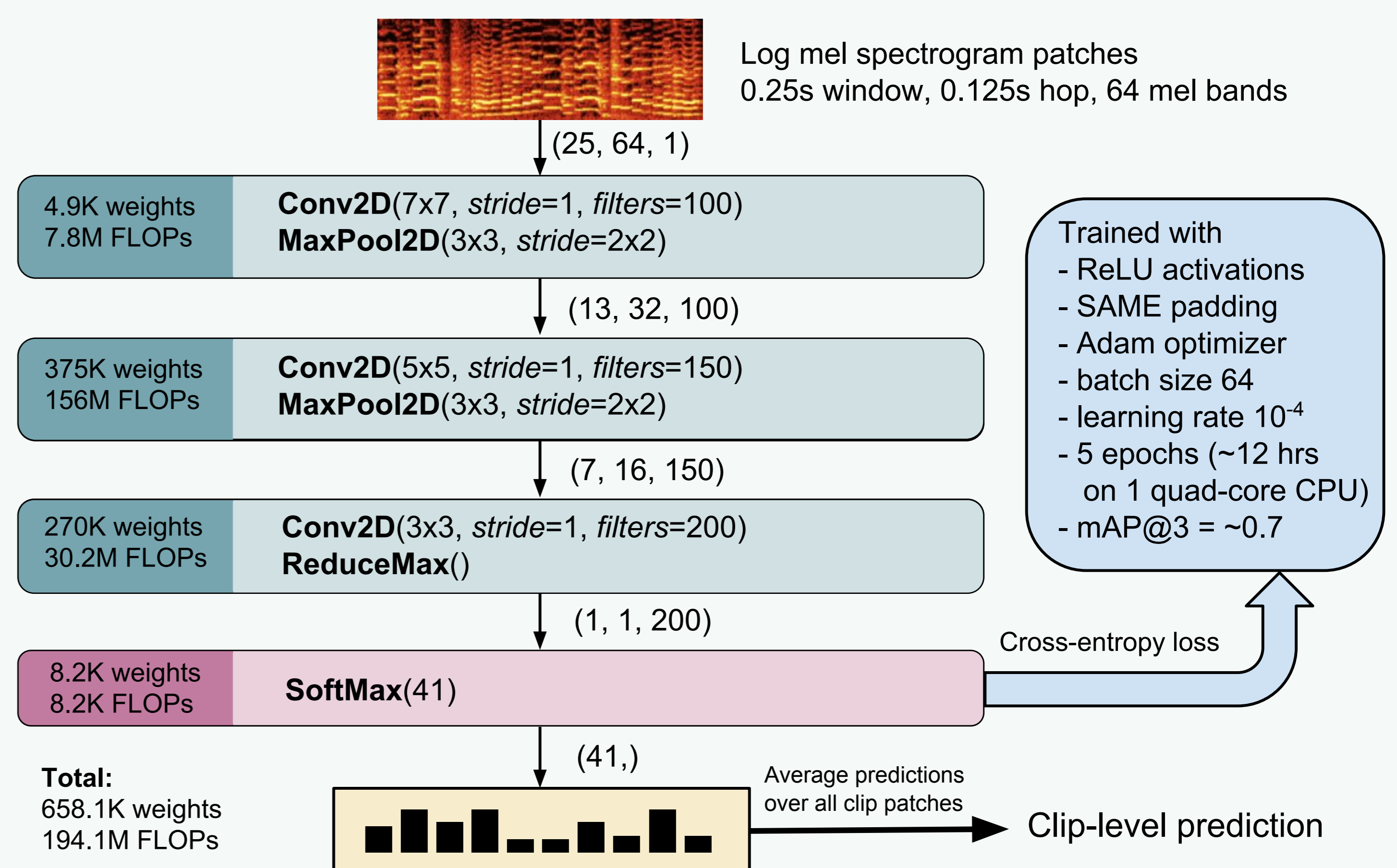
- ▶ **Goal:** recognize the category among 41 diverse categories
- ▶ multi-class classification problem \Rightarrow audio tagging
- ▶ Freesound content: variable length & recording techniques
- ▶ annotations of varying reliability



Task Setup & Some Numbers

- ▶ public/private leaderboards: 19% / 81% of test set
- ▶ evaluation metric: mAP@3 (partial credit for almost correct predictions)
- ▶ maximum of 2 daily submissions
- ▶ top-3 winners are required to publish the code
- ▶ Judges' Award to foster novel, problem-specific, efficient approaches
- ▶ 558 teams (20 of them submitting 36 systems to DCASE) & 5684 entries

Baseline System



Discussion

- ▶ Variable-length clips: random slicing vs. **fixed-length sliding windows**
- ▶ **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 \Rightarrow 30)
- ▶ Noisy labels: **semi-supervised learning** (self-supervision) & robustness measures (**label smoothing, loss masking**)

Open Knowledge: discussion forum & sharing kernels

