**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**

<table>
<thead>
<tr>
<th>System</th>
<th>Features</th>
<th>Classifier</th>
<th>mAP@3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeong</td>
<td>log-mel energies, waveform</td>
<td>CNN, ensemble</td>
<td>0.9538</td>
</tr>
<tr>
<td>Dorfer</td>
<td>mel-weighted spectrogram, log-power spectrogram</td>
<td>CNN, ensemble</td>
<td>0.9518</td>
</tr>
<tr>
<td>Iqbal</td>
<td>log-mel energies</td>
<td>CNN, GCNN, CRNN, GCRNN</td>
<td>0.9512</td>
</tr>
<tr>
<td>Kele</td>
<td>log-mel energies</td>
<td>CNN, GBM, ensemble</td>
<td>0.9498</td>
</tr>
<tr>
<td>Nguyen</td>
<td>log-mel energies</td>
<td>CNN, ensemble</td>
<td>0.9496</td>
</tr>
<tr>
<td>Wilhelm</td>
<td>log-mel energies, MFCC, waveform</td>
<td>CNN, ensemble</td>
<td>0.9453</td>
</tr>
<tr>
<td>Kuali</td>
<td>log-mel energies, MFCC, waveform</td>
<td>CNN, ensemble</td>
<td>0.9423</td>
</tr>
<tr>
<td>Pantic</td>
<td>mel-spectrogram, CQT</td>
<td>CNN, ensemble</td>
<td>0.9419</td>
</tr>
<tr>
<td>Wilkinghoff</td>
<td>PLP, MFCC, mel-spectrogram, waveform</td>
<td>CNN, ensemble</td>
<td>0.9414</td>
</tr>
<tr>
<td>Shan</td>
<td>log-mel energies, MFCC</td>
<td>CNN, ensemble</td>
<td>0.9405</td>
</tr>
</tbody>
</table>

**Task Description**
- **Goal**: recognize the category among 41 diverse categories
- multi-class classification problem ⇒ 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**

- Log mel spectrogram patches
- 0.25s window, 0.125s hop, 64 mel bands
- 4.9K weights 7.8M FLOPs
- Conv2D(7x7, stride=1, filters=100)
- MaxPool2D(3x3, stride=2x2)
- 375K weights 18M FLOPs
- Conv2D(5x5, stride=1, filters=150)
- MaxPool2D(3x3, stride=2x2)
- 270K weights 30.2M FLOPs
- Conv2D(3x3, stride=1, filters=200)
- ReduceMax()
- 8.2K weights 8.2K FLOPs
- SoftMax(41)

**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 ⇒ 30)
- Noisy labels: **semi-supervised learning** (self-supervision) & robustness measures (label smoothing, loss masking)

**Open Knowledge: discussion forum & sharing kernels**

---

**Summary & Results, Task 2**

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

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