

Input: Single-source as well as mixes

(Mel-Spectr. or "Harmonic Error")

Model (pre-trained) + fine-tuned

- Harmonic CNN
- Short Chunks CNN
- RNNs

separate Parts (layers) (take on Attention might be better)

Git | Vox | Strings | . | . | . | Drums Perc

Instrument-Family space

Activity/Loudness
Event-Rel.

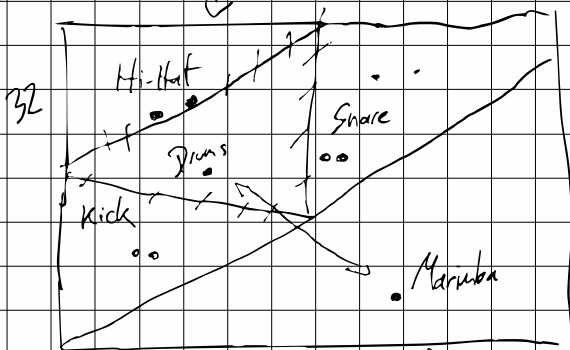
Norm

Norm

Activity/Loudness/
Event Detection

Normalise

Viola
Violon
String Orchestra
Cello
CB



MFCFS (single-frame?)
Pooling

Zero-Crossing Rate

Spectral Centroid

Deep Clustering

Neighborhood Component Analysis:

$$\mathcal{L} = \frac{1}{|B|} \sum_{i \in B} \log \frac{\sum_{j: y_i = y_j} \exp(-\|z_i - z_j\|^2)}{\sum_{k \neq i} \exp(-\|z_i - z_k\|^2)}$$

How to sample Batches?