DrummerNet –
Deep Unsupervised Drum Transcription

Keunwoo Choi and Kyunghyun Cho
Spotify and New York University
keunwooc@spotify.com

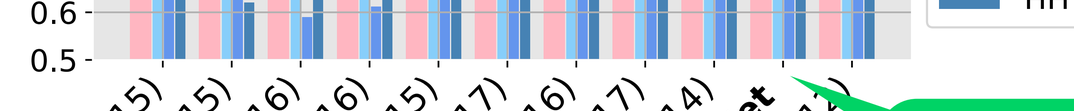
Code + Paper

1. Introduction

- def(Drum Transcription): drum stem → music score with kick/snare/hihat
- Unsupervised transcription: transcribe using audio only!
  - Because it's expensive to get an large annotated dataset
    - Hence supervised learning approaches don't generalize well
      - i.e., fails with unseen kind of drum sound
- Any prior work?
  - Few non-deep unsupervised (drum, or others) transcription systems

2. How is it possible?

- Imagine a structure like this:

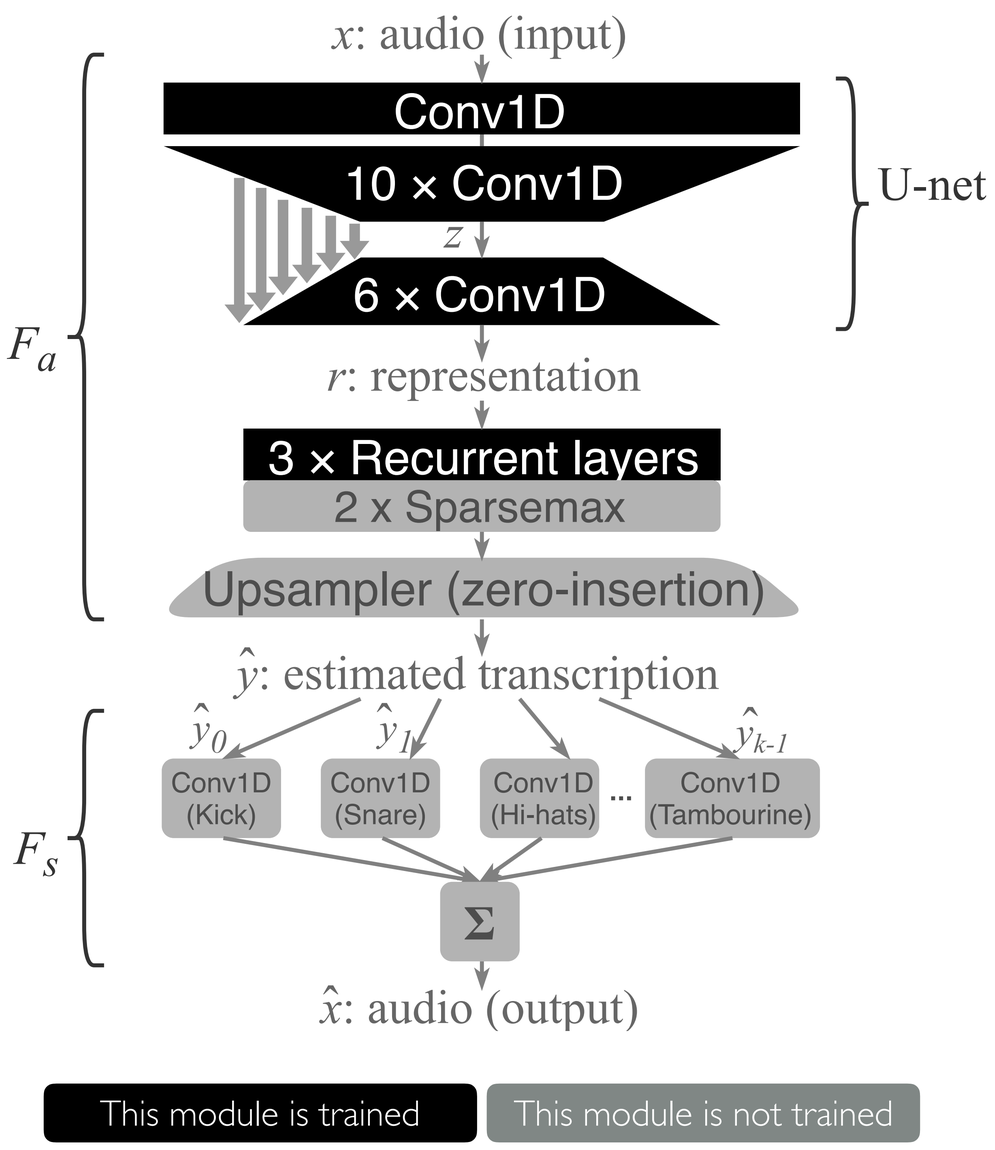


Q. If we train Blackbox so that ***x_hat*** becomes similar to ***x***, what would Blackbox do?

drum rolls

A. Drum Transcription!

3. DrummerNet

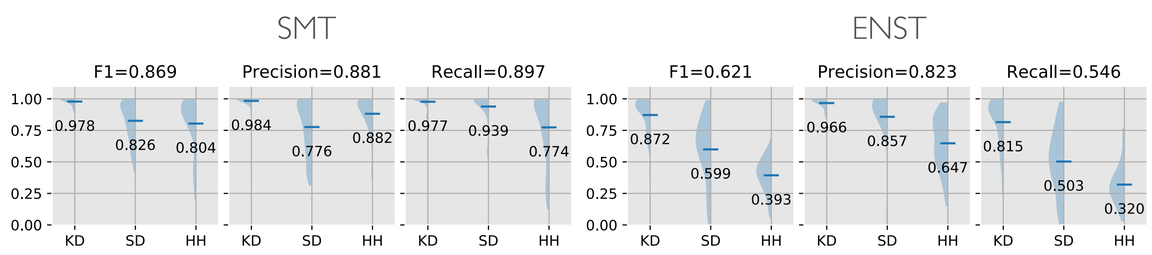


- **Analysis system, *F_a*** : Transcriber
  - Unet learns a representation ***r*** based on the input audio drum stem ***x***
  - Recurrent layers learn to transcribe along time/instruments
  - Sparsemax layers output *really* sparse activation (lots of zeros)
  - Upsampler outputs transcription ***y_hat***
    - Zeros are inserted to compensate downsampling effect of U-net
- **Synthesis system, *F_s*** : Synthesizer
  - Conv1D layers synthesize each drum track by convolving {impulses (***y_hat_k***)} * {drum component waveform_k}
  - Σ sums up each track ***x_hat_k***
- **Training**
  - Minimize reconstruction error (***x*** vs ***x_hat***)

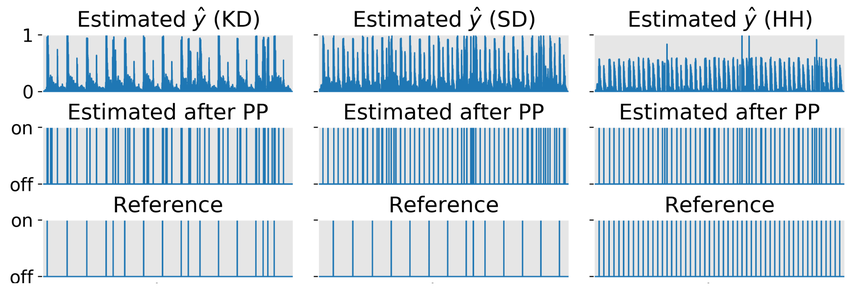
4. Results

- F1 scores on a dataset ('SMT') compared to others
  - DrummerNet uses much larger training dataset and it leads to better generalizability

- SMT vs ENST dataset
  - Because the reconstruction error is measured by audio domain, there's linearity between {transcription, audio} + We mix-use the concepts of velocity vs probability → we get false positives if
    i) different drum components sound (sort of) similar
    ii) the play is nuanced (**`var(velocities)`** is large)→e.g., ENST dataset



- DrummerNet fails like this when it fails



- More results and discussion including ablation study are in the paper!

5. Future work/Code/Paper

Future work — Universal unsupervised transcription!
RL + Unsupervised transcription

code/paper — https://github.com/keunwoochoi/DrummerNet