Few-Shot Drum Transcription in Polyphonic Music

Yu Wang[1], Justin Salamon[2], Mark Cartwright[1], Nicholas J. Bryan[2], Juan Pablo Bello[1]

[1]Music and Audio Research Laboratory, New York University
[2]Adobe Research

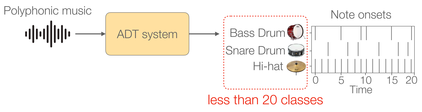
Highlights

- Apply few-shot learning to automatic drum transcription (ADT)
- Outperforms SOTA supervised ADT under fixed transcription vocabulary
- Supports open vocabulary ADT with a small cost of minimal human input

SCROLL

1. Motivation & Goal

- Current ADT systems have **small** and **fixed** transcription vocabulary
- Standard supervised learning requires a lot of data to expand the vocabulary

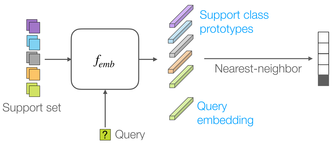


Can we perform open vocabulary ADT on any percussive sound with few data?

2. Method: Metric-based Few-Shot Learning

- Recognizing novel classes from very few labeled examples
- **Prototypical networks** [Snell, 2017]
  - Learn a discriminative embedding space
  - Robust representation (prototype) for a novel class based on few examples
  - Classification: finding nearest prototype

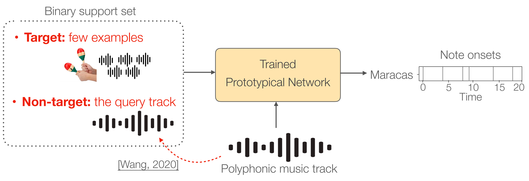
- Training objective: ***C-way K-shot* classification**
  - Ex: *5-way 2-shot*



- **Episodic training:** Sample different set of classes in each training episode

3. Proposed Paradigm: Few-Shot Drum Transcription

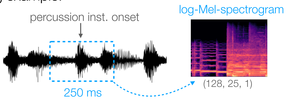
- Given a target percussion instrument and a music track:

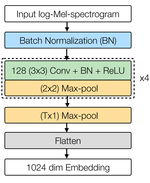


Trained prototypical network + Minimal human input = Transcribe any percussion instrument

4. Experimental Design: Training

- Dataset: Slakh2100 [Manilow, 2019]
  - Define 282 percussive classes
- Training example:



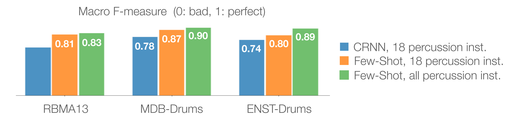


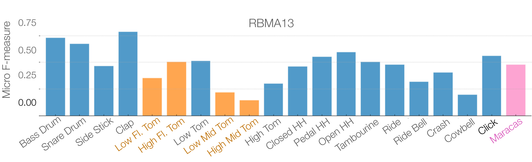
- 10-way 5-shot classification with CNN embedding model

5. Experimental Design: Evaluation

- Three real music datasets
  - ENST-Drums *(20 percussion inst.)*, MDB-Drums *(21)*, RBMA13 *(23)*
- Transcription vocabulary
  - Fixed — 18 percussion instruments
  - Open — all percussion instruments within each dataset
- Target examples in the support set
  - Randomly sample **5** target examples from each track to simulate human input
- Baseline: Supervised CRNN [Vogl, 2018]

6. Results





- Outperforms supervised approach under fixed vocabulary setting
- Supports open vocabulary ADT
- Supports finer-grained class labeling and/or extended vocabularies
- **Future work:** automatic and human-in-the-loop target example selection

Paper: bit.ly/fewshotADT

Contact: wangyu@nyu.edu
y-wang.weebly.com
@yuwang_tw