THE DEVELOPMENT OF A DRUM IDENTIFIER AND TRANSCRIBER

Louis Larsen
Dr. Adam Finkelstein

AUTOMATIC DRUM TRANSCRIPTION

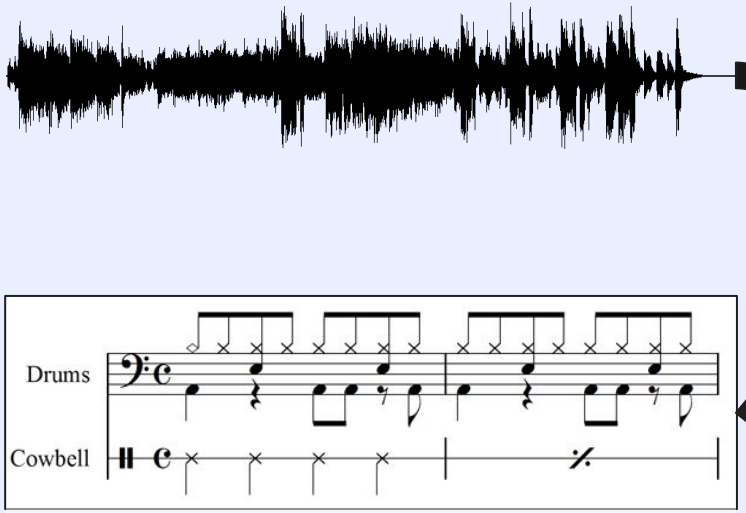


Figure 1. An Audio Waveform Turned into Drum Sheet Music

- **Audio File to Symbolic Representation**
  - **Sheet Music or MIDI File**
- **Differentiate Noise-Based Instruments**

PARTS OF A DRUM SET

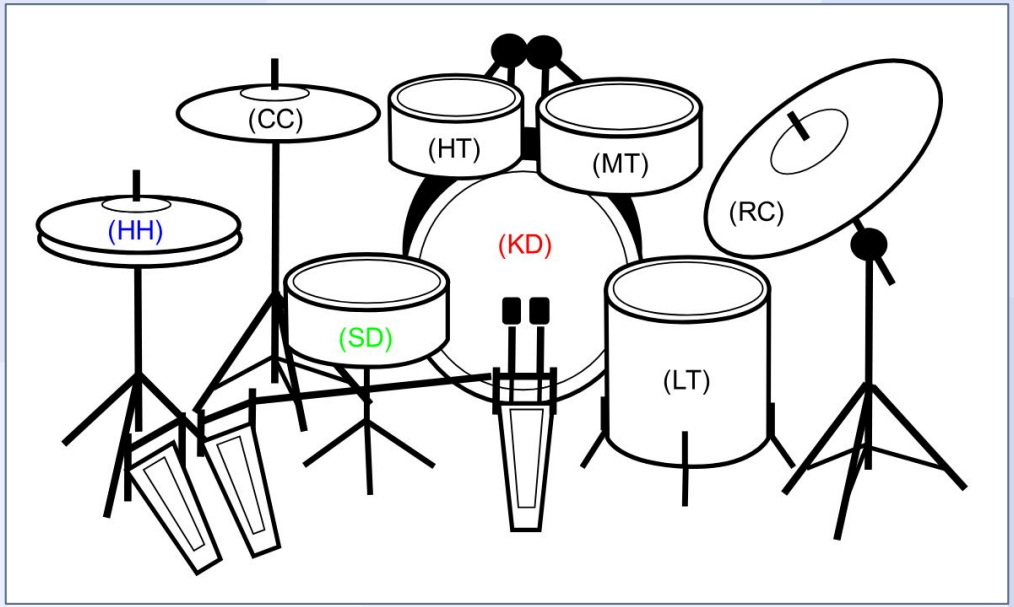


Figure 2. Most Important Parts of a Drum Kit. Adapted from "A Review of Automatic Drum Transcription" by Wu et al, 2018, *IEEE*

RELATED WORK

- Onset & Frames (Callender et al., 2020)
- DT Ensemble (Vogl et al., 2018)

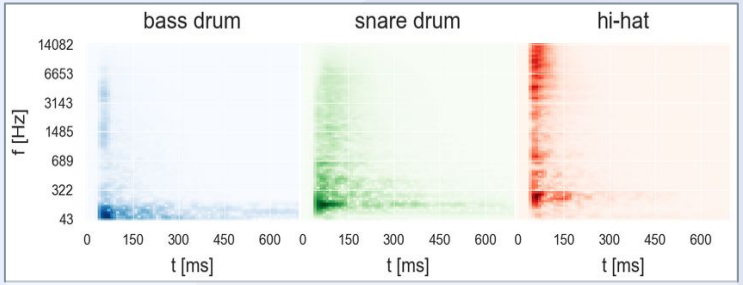


Figure 3. Spectrograms for drum sounds. Adapted from "Deep Learning Methods for Drum Transcription" by R. Vogl, 2018, Nov, *Computational Perception*.

RELATED WORK

- Seq2Seq
  (Hawthorne et al., 2021)
- hFT-Transformer
  (Toyama et al., 2023)
- MT3
  (Gardener et al., 2022)

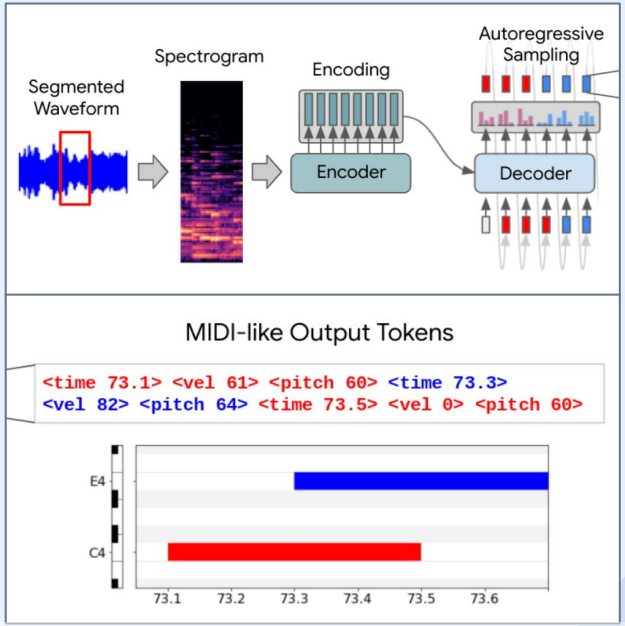


Figure 4. Generic Encoder-Decoder Architecture. Adapted from "Sequence to Sequence Piano Transcription with Transformers" by Hawthorne et al., 2021, *ISMIR*.

TO WHAT EXTENT CAN A TRANSFORMER-BASED MODEL ASSIST IN THE TRANSCRIPTION OF DRUM SET AUDIO?

DATASET - E-GMD

444	45,537	22
Hours of Drum Set Audio	Total Sequences	Instrument Mappings

DATASET - IMST

2	104	3
Hours of Drum Set Audio	Total Sequences	Instrument Mappings

MODEL STRUCTURE

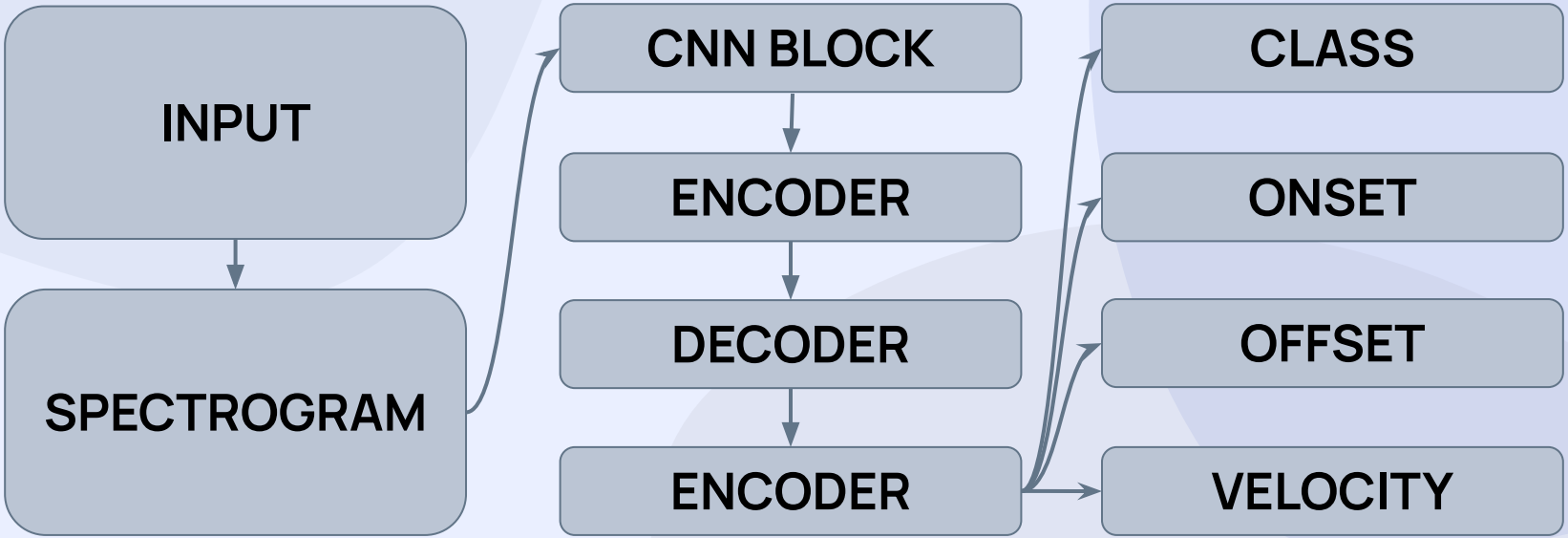


Figure 5. Flowchart for the structure of the model.

SPECTROGRAM INPUT

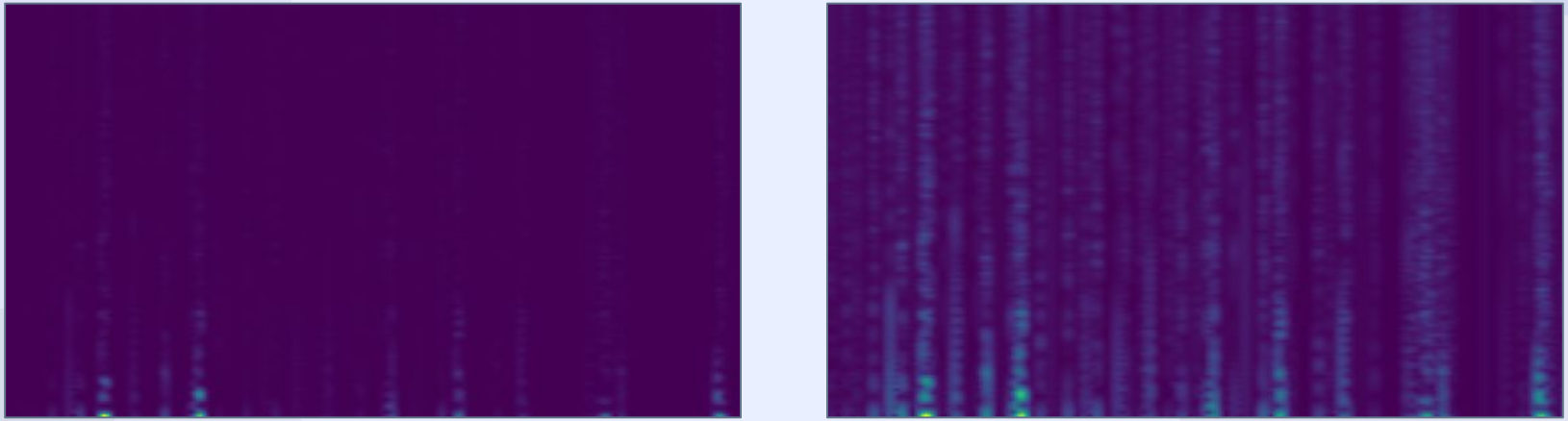


Figure 6. Example of Spectrogram Input. Left is unmodified. Right is square-rooted for visualization.

MODEL STRUCTURE

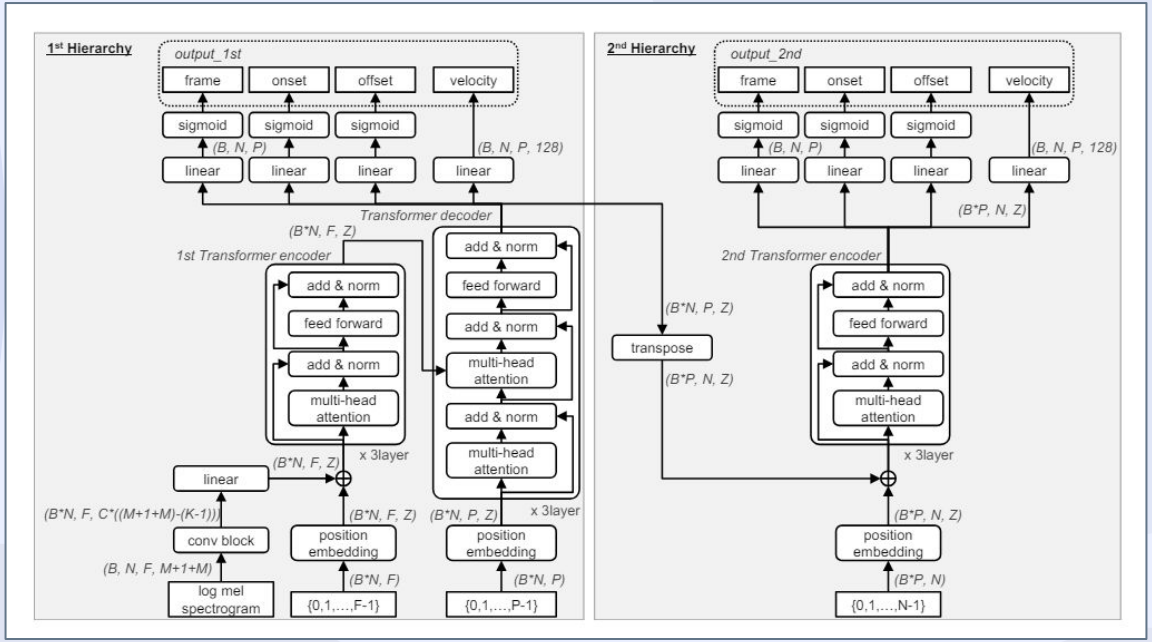


Figure 7. More Detailed Flowchart of the Model Architecture. Adapted from "Automatic Piano Transcription with Hierarchical Frequency-Time Transformer" by Toyoma et al., 2023, *ISMIR*

CHALLENGES

Figure 8. A GPU Server Room

- Overfitting Tests
- Training Processing Power
- Multi-GPU Training

RESULTS - LOSS

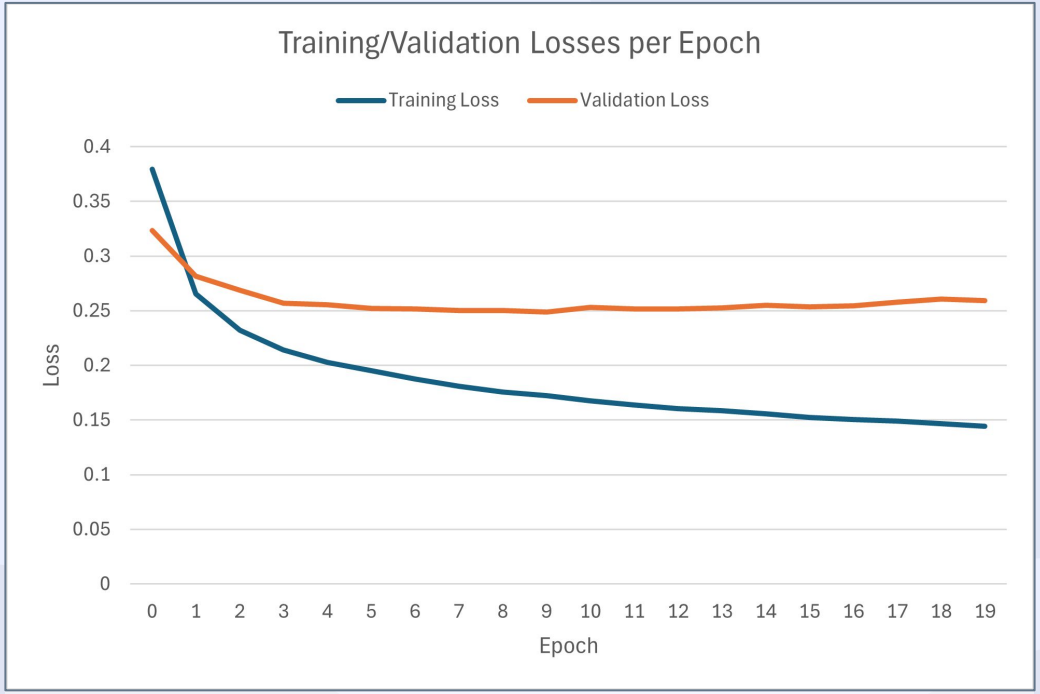


Figure 9. Shows Training and Validation Loss per Epoch Number

RESULTS - BEST EPOCH

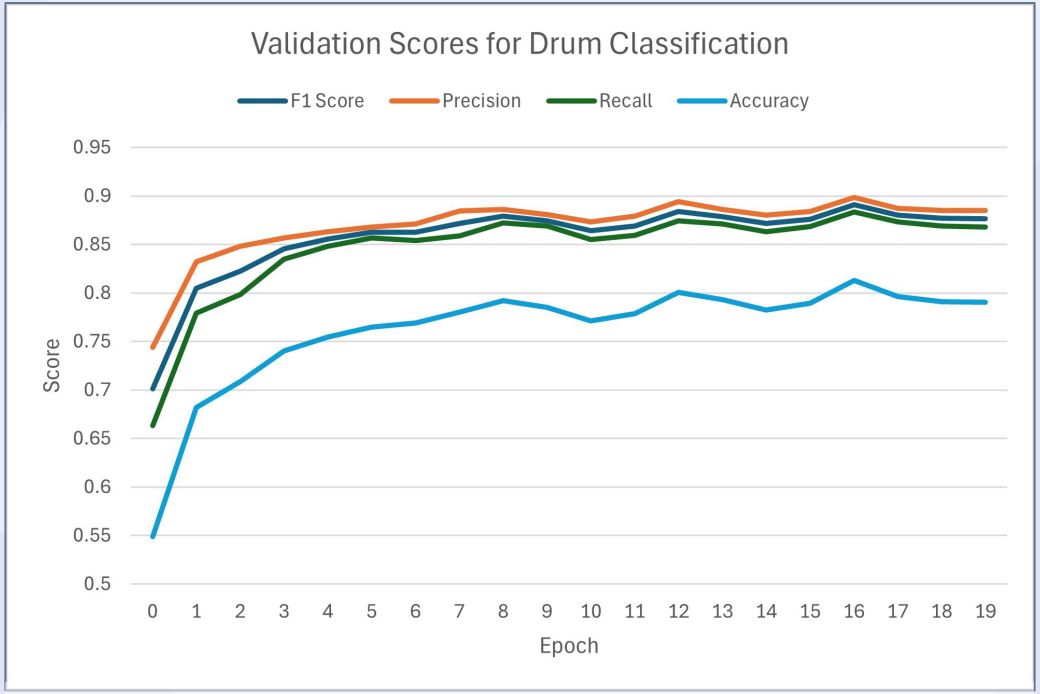


Figure 10. Shows Classification Scores per Epoch Number

RESULTS - BEST EPOCH

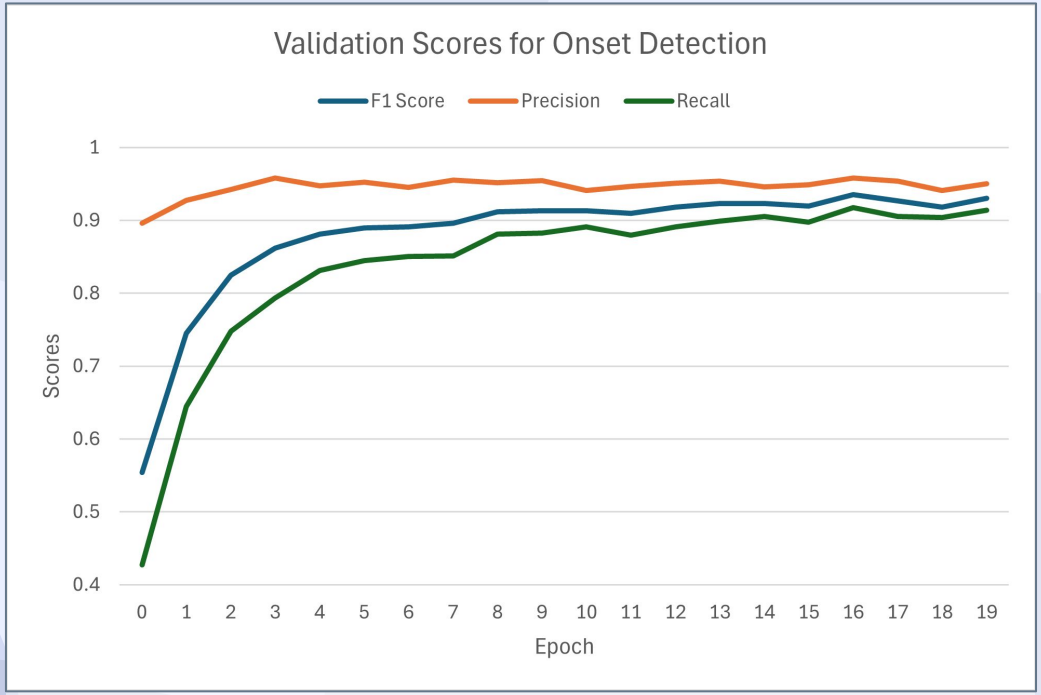


Figure 11. Shows Onset Scores per Epoch Number

RESULTS - BEST EPOCH

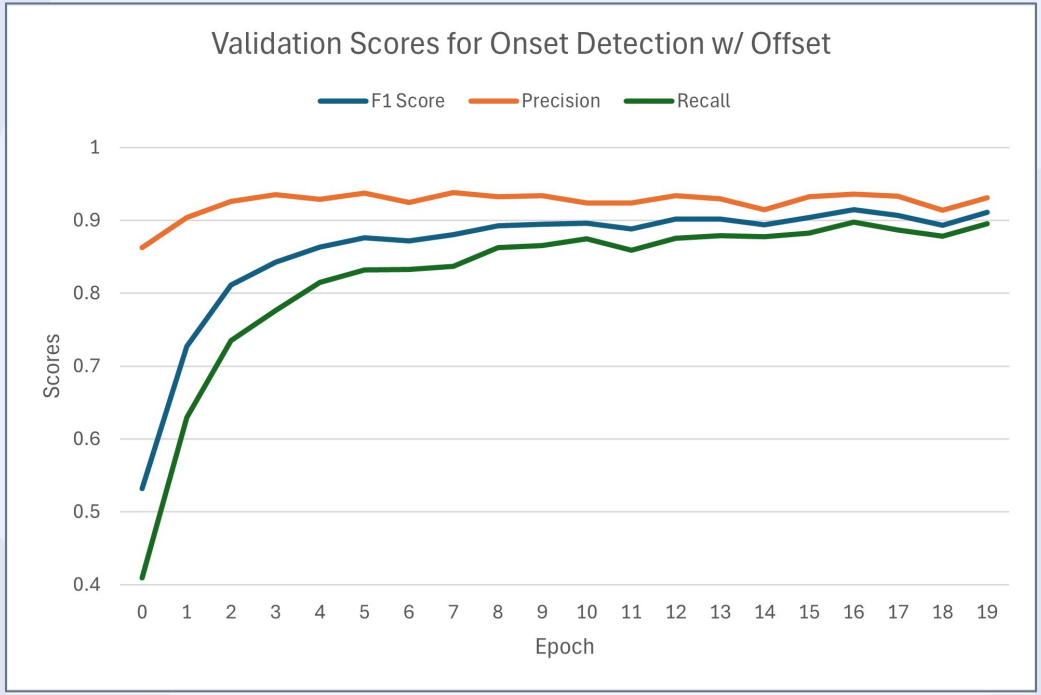


Figure 12. Shows Onset with Offset Scores per Epoch Number

RESULTS - BEST EPOCH

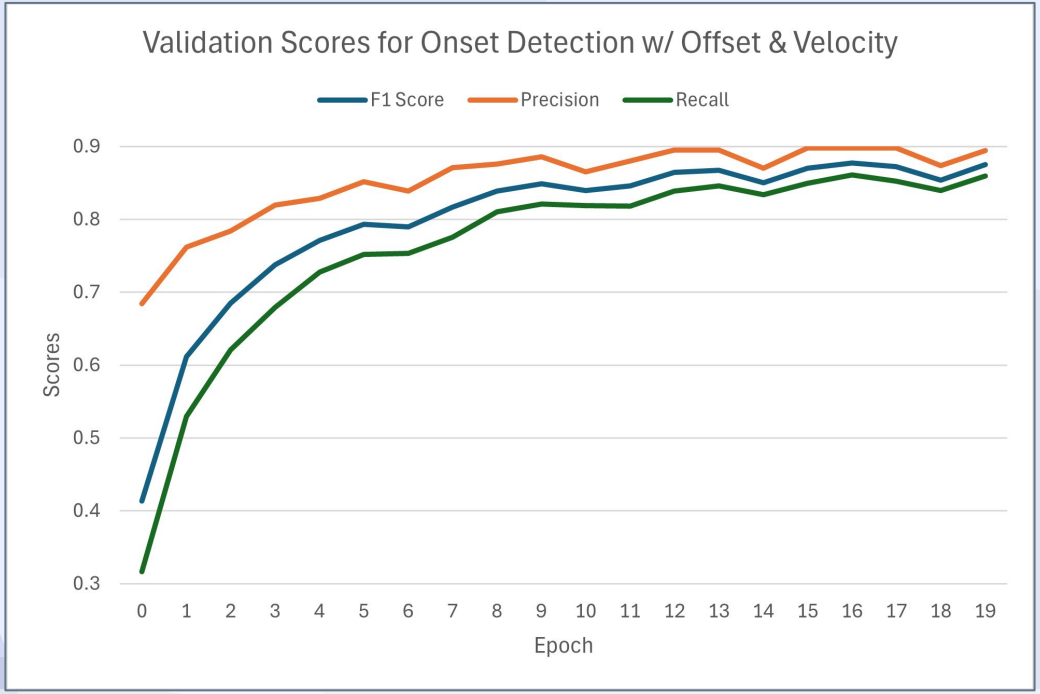


Figure 13. Shows Onset with Offset and Velocity Scores per Epoch Number

RESULTS - EVALUATION ON E-GMD

83.28	86.02	83.67	80.86
Class F1 Score (%)	Onset F1 Score (%)	Onset w/ Offset F1 Score (%)	Onset w/ Offset & Velocity F1 Score (%)

RESULTS - COMPARISON ON E-GMD

MODEL	ONSET F1 (%)	VELOCITY F1 (%)
My Model	86.02	83.11
Onset & Frames	83.40	61.70
DT Ensemble	64.98	–

RESULTS - COMPARISON WITH E-GMD

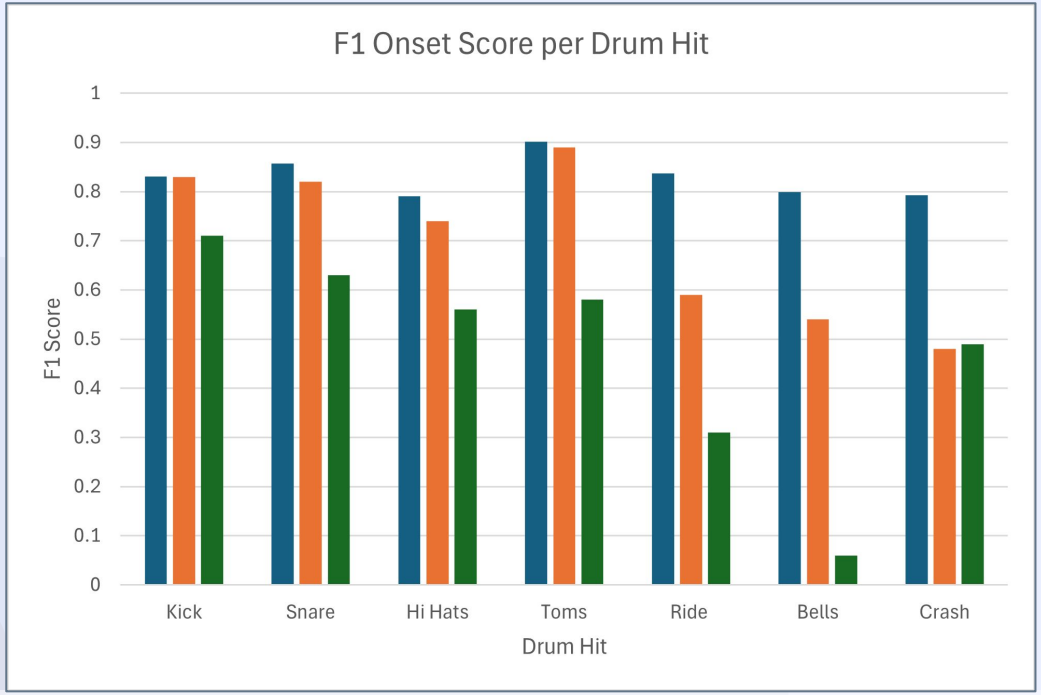


Figure 14. Shows F1 Onset Score per Drum Hit.
Blue is My Model. Orange is Onset & Frames. Green is DT Ensemble.

LIMITATIONS

- Reliance on Isolated Drum Set Audio
- Comparison with IMST

MODEL	ONSET F1 (%)
My Model	77.90
Onset & Frames	85.72
DT Ensemble	91.49

CONCLUSION

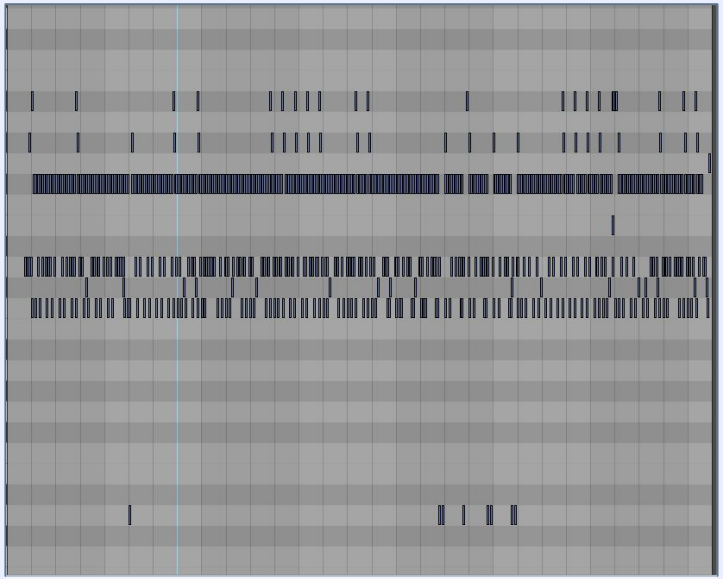


Figure 15. A Drum Set Midi File

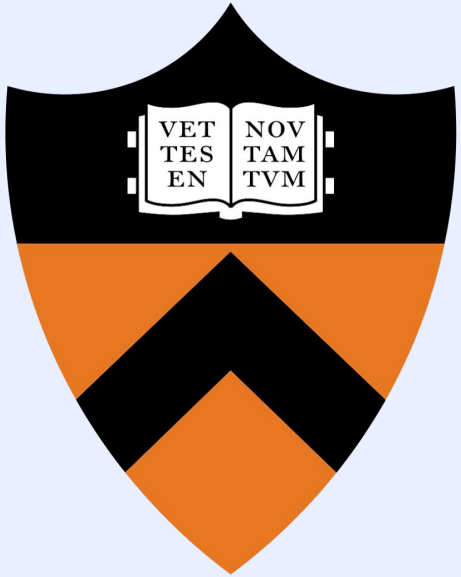
- Model shows promise
- Highest F1 Scores Achieved on E-GMD
- Creative Applications

FUTURE WORK

- Increase F1 Scores on IMST
- Real-Time Application

Figure 15. A Person Playing a Drum Set

ACKNOWLEDGEMENTS



- Dr. Adam Finkelstein
- Mentor David Braun
- Tensordock
- Professors at Princeton