Exemplar-Based Window-wise Onset/Offset Detection Model for Multi-Pitch Extraction

XIAOHE TIAN

Key Designs

- Self-built audio files, MIDI data sheets, and data directories
- Window-based dataset with exemplars
- Note event matrices
- Deep Salience-based window-wise onset/offset detection (WOOD) model

Motivations

- Versatile usages of extracted sheet data
- Difficult nature of multi-pitch extraction

Why Exemplar?

- Able to obtain and annotate my machines themselves from full audio
- Provide information for audio reconstruction
- Advanced techniques in instrument separation → easy to produce single-instrument audio

Dataset

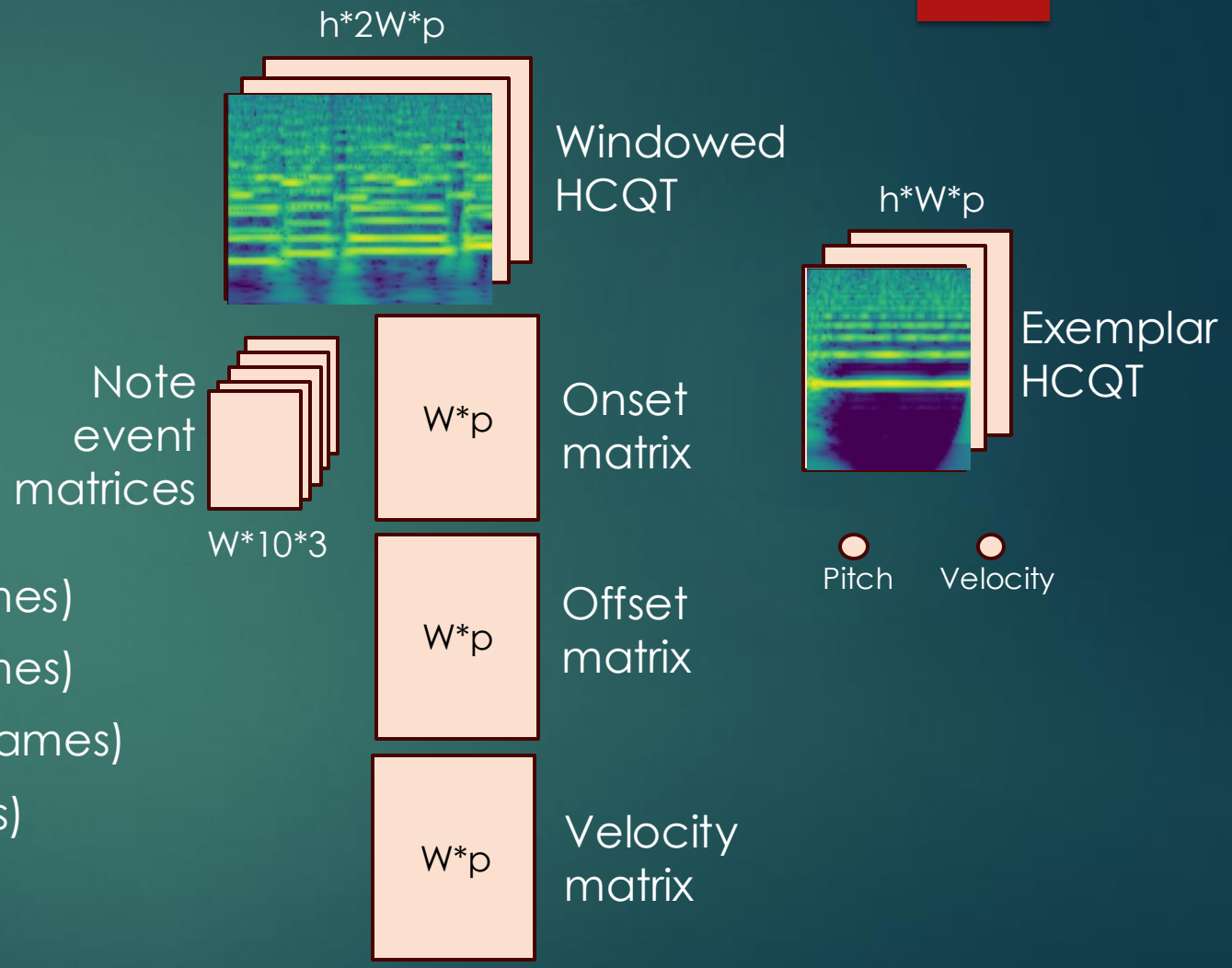
- Exemplar-based dataset built by ourselves!
- Audio rendered with Reaper with random VST instruments
- Directory structure
  - full audio
  - exemplar audio
  - sheet information

Pitch	Start time	End time	Velocity
69	0	0.277778	106
46	0	0.555556	127
70	0.277778	0.694444	104
70	0.833333	1.25	104
41	0.833333	1.388889	127
74	1.388889	1.666667	104
75	1.666667	1.944444	106
…	…	…	…

Format of .csv sheet data

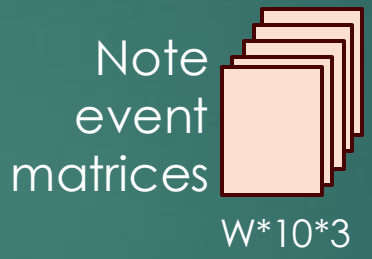
Dataset

- PyTorch Dataset class
  - Windowed HCQT
  - Exemplar HCQT
  - Exemplar pitch
  - Exemplar velocity
  - Target onset matrix (of current frames)
  - Target offset matrix (of current frames)
  - Target velocity matrix (of current frames)
  - Note event matrices (of last frames)
- Parameters: number of harmonics; window length; number of pitches



Dataset

- Note event matrix
  - Active events: (pitch, velocity, duration)
  - Number_of_pitches represents no pitch
  - Sorted



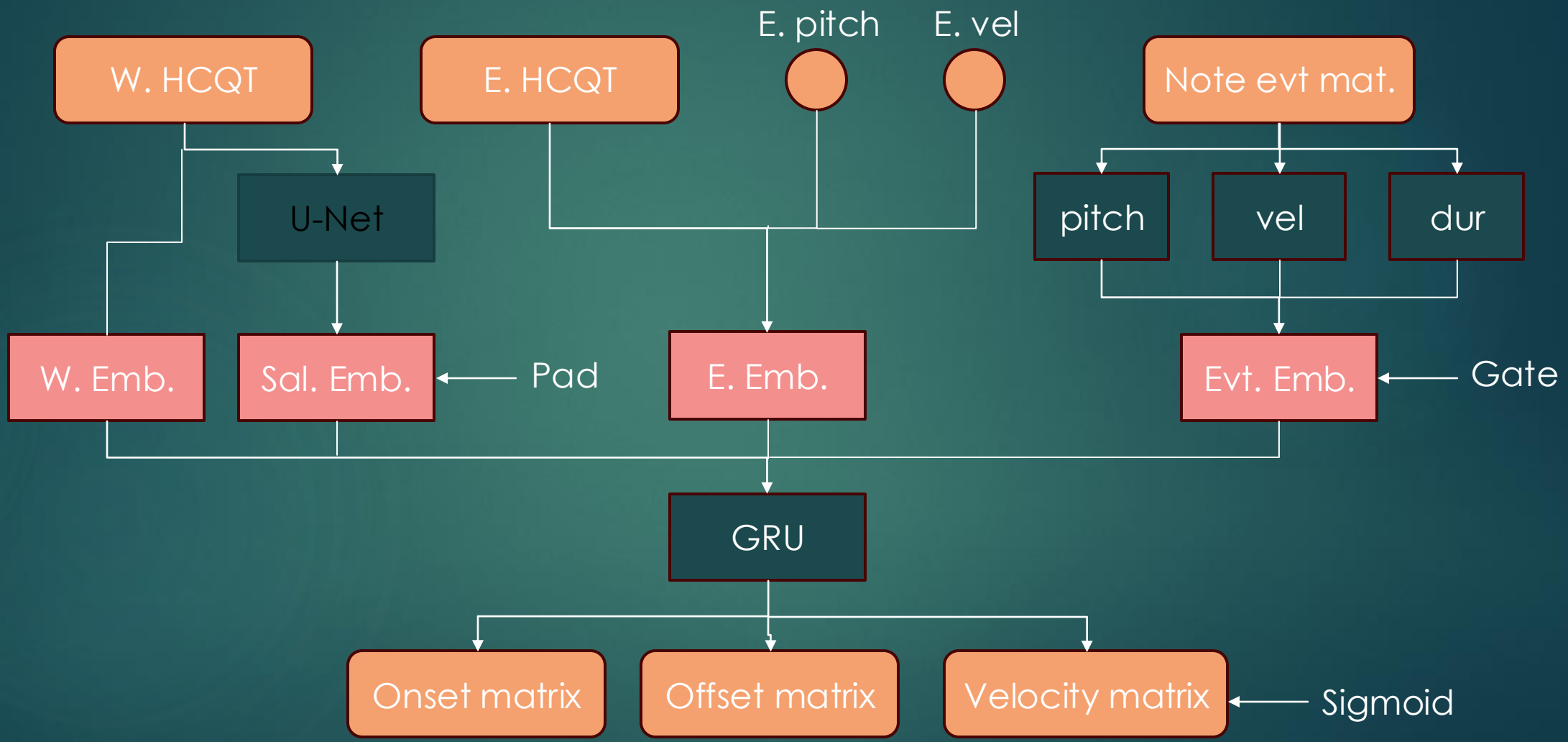
Pitch	Velocity	Duration
20	120	7
51	82	5
56	65	15
84	0	0
84	0	0
84	0	0
84	0	0
84	0	0
84	0	0
84	0	0

Format of note event matrix

WOOD Model Structure

- Deep Salience method is applied to windowed HCQT
- Make embeddings for each of:
  - Windowed HCQT
  - Windowed HCQT salience
  - Exemplar HCQT + exemplar pitch + exemplar velocity
  - Note event matrices (columns separately encoded and then combined together)
- Multi-head outputs:
  - Onset matrix
  - Offset matrix
  - Velocity matrix

WOOD Model Structure

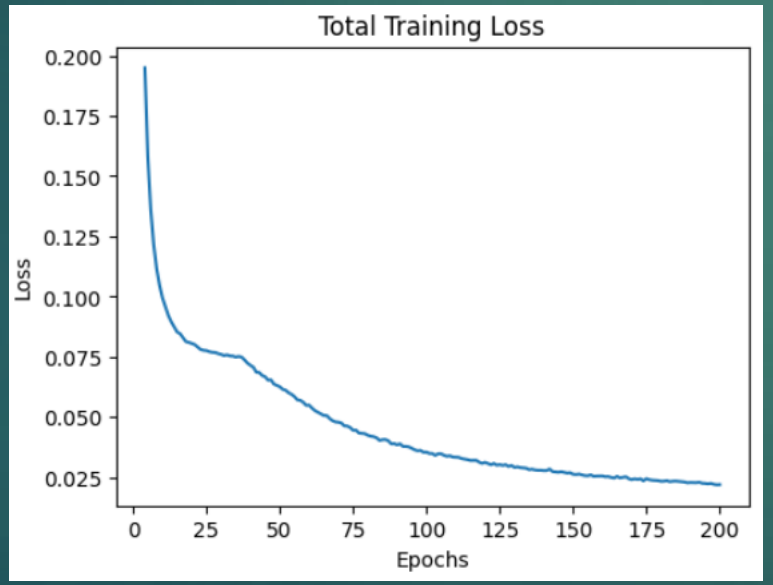


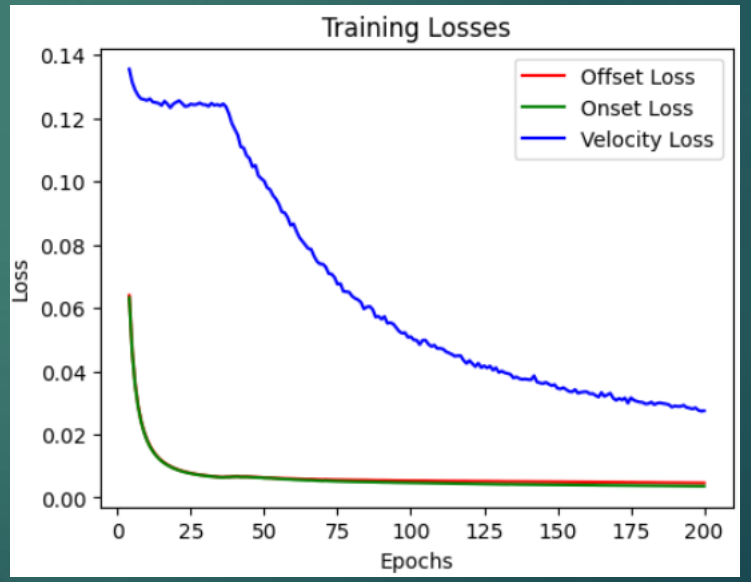
WOOD Model Structure

- Loss calculation
  - BCEWithLogitsLoss for onset and offset matrices
  - $L = \frac{\sum(abs(\boldsymbol{V_p} - \boldsymbol{V_t}) * \boldsymbol{O_t})}{\sum \boldsymbol{O_t} + \varepsilon}$ for velocity matrix

Results & Discussions

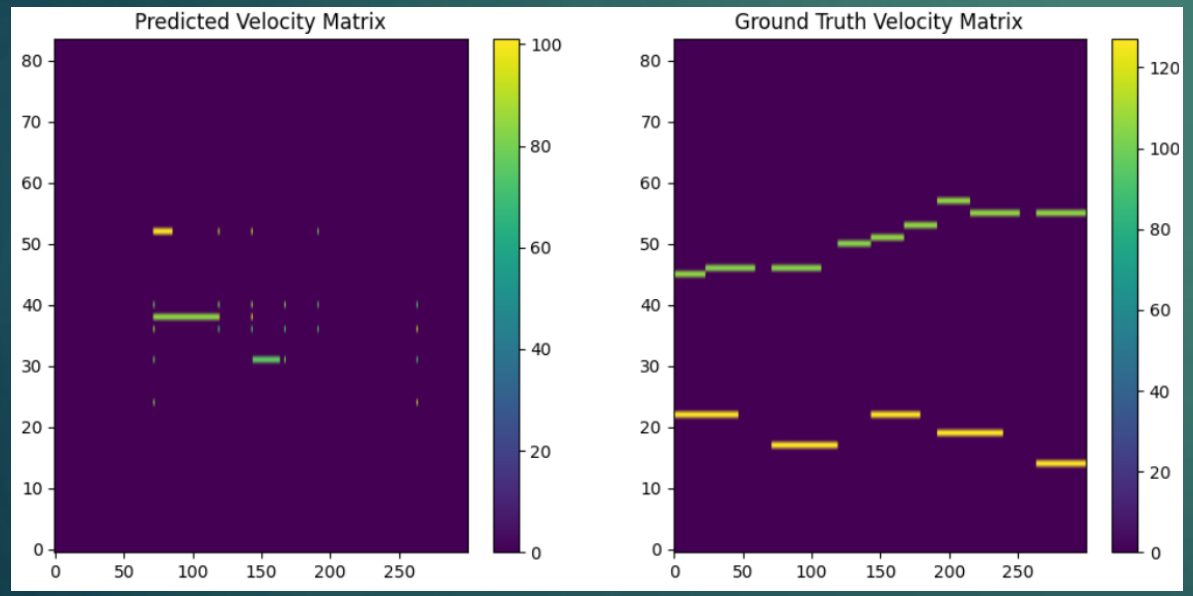
- Trained on Google Colab with T4 GPU
- Total and individual losses of 200 epochs are plotted

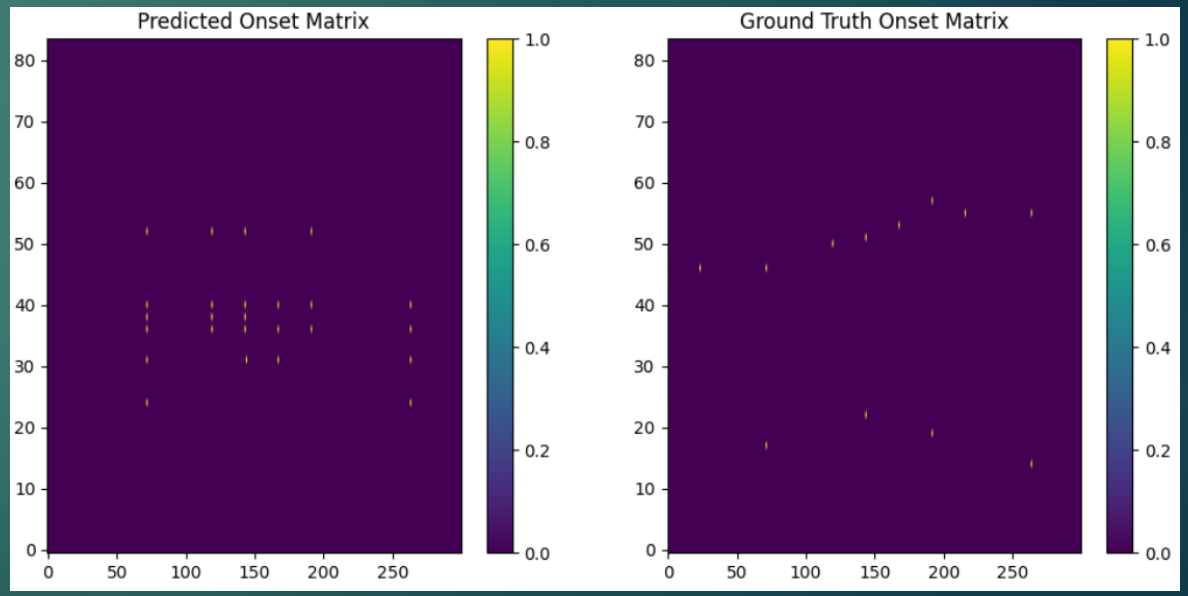




Results & Discussions

- Tested on whole song; recurrent use of note event matrices
- Constructed note event matrices with output prediction matrices
- Constructed velocity and onset matrices from note event matrices





Results & Discussions

- How to decide the gate of note event embedding?
- Exemplar prediction?
- How to evaluate long-term loss of model?
- Expanding the dataset
- Loss weight in training & threshold in testing

Thank you!

- Q&A!

References
[1] Bittner R M, McFee B, Salamon J, et al. Deep Salience Representations for F0 Estimation in Polyphonic Music. *ISMIR. 2017: 63-70*. 2017.
[2] Hawthorne C, Elsen E, Song J, et al. Onsets and frames: Dual-objective piano transcription. *arXiv preprint arXiv:1710.11153*, 2017.
[3] Engel J et al., Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. http://arxiv.org/abs/1704.01279. 2017.
[4] J. Engel et al. "DDSP: Differentiable Digital Signal Processing". *International Conference on Learning Representations*. 2020.