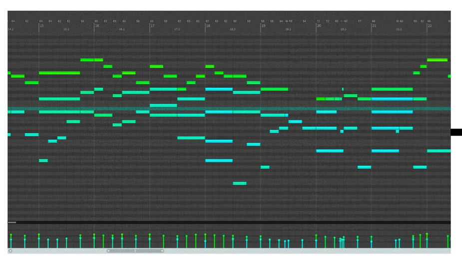
Performance MIDI-to-Score Conversion by Neural Beat Tracking

23rd International Society for Music Information Retrieval Conference (ISMIR 2022)

Lele Liu[*1,2], Qiuqiang Kong[3], Veronica Morfi[1], and Emmanouil Benetos[1,2]

[1] Queen Mary University of London, UK [2] The Alan Turing Institute, UK [3] ByteDance Shanghai, China
* The author conducted this work as an intern at ByteDance

I. Performance MIDI-to-Score (PM2S) Conversion



PM2S Conversion

Recording from MIDI keyboard or Transcribed MIDI performance

Music score in **MIDI**/MusicXML/Lilypond/… format

- **Applications**: Music improvisation, complete music transcription, music performance analysis
- **Subtasks**: Rhythm quantisation, note value prediction, key estimation, voice separation, and possibly score type-setting such as beaming and playing techniques annotation

II. Contributions

- A deep learning (CRNN) model with a compact output for PM2S
- A new rhythm quantisation method by tracking beats on a MIDI note sequence
- Comparisons between different input encodings of MIDI note sequences for tracking beats from performance MIDI
- Ablation studies on input features and data augmentation methods for tracking beats from performance MIDI
- A PM2S conversion toolbox

III. Methodology

Performance MIDI in note sequence of length N:

$$X = \{(p_n, o_n, d_n, v_n)\}_{n=1}^{N}$$

where for the n-th note:

p_n - MIDI pitch
o_n - onset in seconds
d_n - duration in seconds
v_n - velocity in MIDI

PM2S Conversion Model

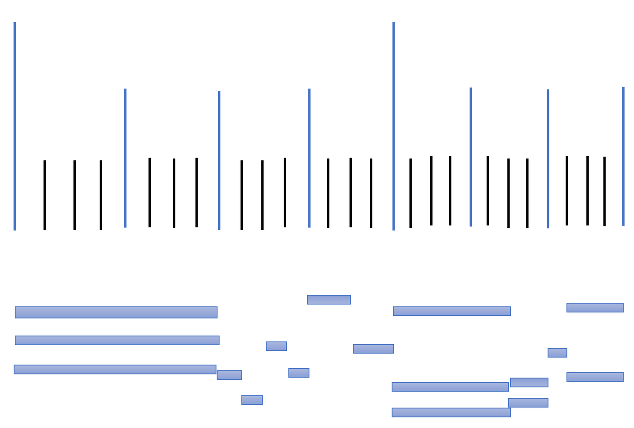
Additional score information to get the **MIDI score**:

$$Y = (Y_n, Y_t, Y_k)$$

where

Y_n - musical onset, note value and hand part for all notes
Y_t - time signature changes
Y_k - key signature changes

Rhythm quantisation: the prediction of musical onset and note value for each note in the note sequence



Rhythm quantisation:

Beat/downbeat tracking
+
Beat subdivision prediction

Beat tracking on note sequences:

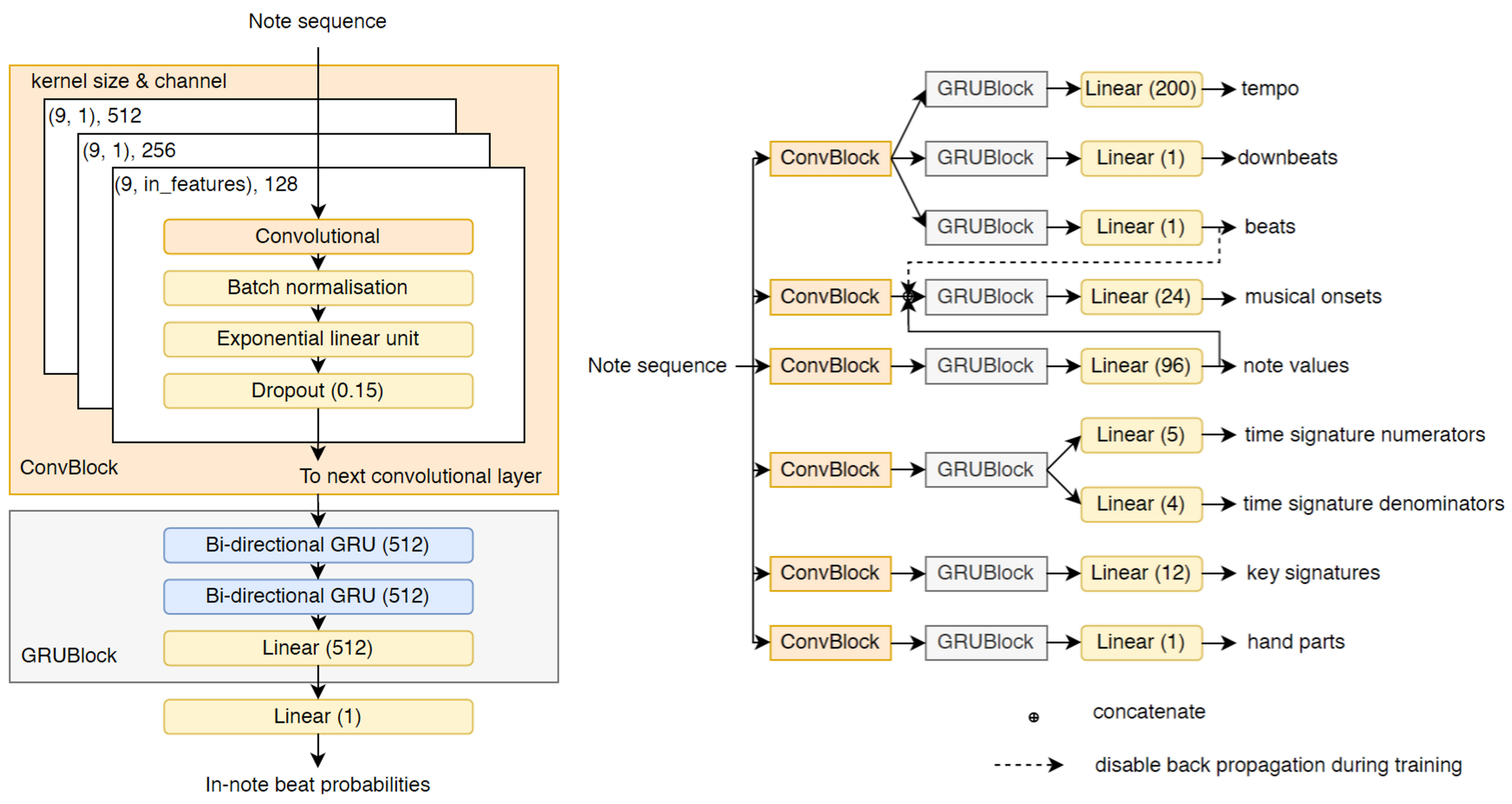
- In-note beats prediction using neural networks
- Out-of-note beats prediction using dynamic programming

- For in-note beat prediction, we use a CRNN model
- For out-of-note beat prediction, we design an objective function to be minimized that encourages a low level of tempo change and add fewer out-of-note beats

$$\mathcal{O} = \sum_{n=1}^{N-2} \left| log\left(\frac{b_{n+2} - b_{n+1}}{b_{n+1} - b_n}\right) \right| + \lambda \times N^o$$

b_n – n-th beat time in the in-note beats list
λ – penalty coefficient for adding out-of-note beats
N^o - number of out-of-note beats

- We further expand the CRNN model to predict a compact output data representation for musical scores



IV. Experiments

- We use a collection of classical piano performances and scores, in total 123.2 hours of piano performances in 504 distinct pieces.
- The train/valid/test splits do not have overlapping pieces

- Comparative experiments show that:
  - Using MIDI pitch, onset shift in one-hot format, duration in raw values, and velocity achieves best performance among difference input data encoding combinations
  - All four input features are helpful
  - Using data augmentation during model training is beneficial
  - Our proposed beat tracking model outperforms a baseline beat tracking model

MV2H evaluation on performance MIDI-to-Score Conversion:

Methods	F_p	F_{vo}	F_{me}	F_{va}	F_{ha}	F
Finale	82.2	54.6	9.9	92.2	86.2	65.0
MuseScore	10.0	65.0	15.3	95.0	84.5	54.0
Proposed	**99.8**	**87.0**	**61.7**	**99.9**	**91.1**	**87.9**

- Results show better performance can be achieved by our model in comparison with two commercial software (Finale and MuseScore), based on the MV2H metric.
- Significantly better performance can be observed in the metrical alignment metric, due to our method's ability to track tempo changes.

Contact:
Email: lele.liu@qmul.ac.uk
Homepage: https://cheriell.github.io

Paper: Code: Example & Demo:

UK Research and Innovation