VirtuosoNet v1.1

1st Dasaem Jeong
Dept. of Art & Technology
Sogang University
Seoul, South Korea
dasaemj@sogang.ac.kr

2nd Taegyun Kwon
Graduate School of Culture Technology
KAIST
Daejeon, South Korea
ilcobo2@kaist.ac.kr

3rd Juhan Nam
Graduate School of Culture Technology
KAIST
Daejeon, South Korea
juhan.nam@kaist.ac.kr

Abstract—We submit a MIDI peformance rendered by VirtuosoNet, which was introduced in 2019, with minor modification. The model takes a sequence of score features encoded in note-level, and generates a sequence of performance features for each note.

I. INTRODUCTION

Our model consists of recurrent neural networks with hierarchical attention and conditional variational autoencoder. The model takes a sequence of note-level score features extracted from MusicXML as input and predicts piano performance features of the corresponding notes.

II. METHODOLOGY

The architecture is the same as ISMIR 2019 version [1], with few modification.

- Used GRU instead of LSTM
- Added additional loss term for first differential of tempo
- Added additional loss term for velocity ratio of each note in a beat
- Trained in a batch size of 32 instead of 1

We did not utilize the measure-level performance planner from the origingal paper.

III. DATASETS

We used the same training dataset as ISMIR 2019 version. To provide the sense of comparison, this dataset later forked to ASAP [2], and performance and music score remain almost the same.

IV. POST-PROCESSING

Although our system is capable of try certain range of diverse style generation, we submitted without any post-processing or manual curation. The result is submitted with the model's initial output without any human evaluation. The initial tempo was strictly followed by given XML. We did not use any style-based modification during the inference. We did not modify or annotate additional information for given music scores.

Sonification was done by basic Studio Grand virtual instrument of Logic Pro X. We applied a simple MIDI FX scriptor to convert MIDI Pedal value to 0 if the pedal value is smaller than 64, as this VSTi applys sustain pedal if the value is larger than zero. We applied no other audio effect except default setting of Logic Pro X. For Mozart/Volodos piece, we trimmed the performance to given time limit, one and half minutes.

REFERENCES

[1] D. Jeong, T. Kwon, Y. Kim, K. Lee, and J. Nam, "Virtuosonet: A hierarchical rnn-based system for modeling expressive piano performance." in *ISMIR*, 2019, pp. 908–915.

[2] F. Foscarin, A. Mcleod, P. Rigaux, F. Jacquemard, and M. Sakai, "Asap: a dataset of aligned scores and performances for piano transcription," in *ISMIR 2020-21st International Society for Music Information Retrieval*, 2020.