Grand Digital Piano: Multi-Modal Transfer of Learning

Sanches Shayda Ernesto Evgeniy (esanches), Ilkyu Lee (lqlee)

Introduction

Reproduction of the touch and sound of real instruments has been a challenging problem throughout the years. For example, while current digital pianos attempt to achieve this task via simplified, finite sample-based methods, there is still much room for improvement in regards to mimicking an acoustic piano.

In this study, we aim to improve upon such methods by utilizing machine learning to train a physically-based model of an acoustic piano, allowing for the generation of novel, more realistic sounds.

Our Approach

We utilize a multi-modal transfer of learning method, wherein digital piano data regarding touch can ultimately generate realistic sound of an acoustic piano:

1. A laser sensor is used on digital piano to gather touch data via key velocity and is used to train a model to predict intermediary modalities, such as MIDI.
2. The model is then applied on the acoustic piano by inputting touch data to predict the intermediary modalities that is not naturally available on acoustic pianos.
3. With these intermediary modalities as input, another model is trained to predict the sound of the acoustic piano.
4. Finally, the previous model is executed on the digital piano to generate realistic sound from sensor data only available on the digital piano.

Implementation

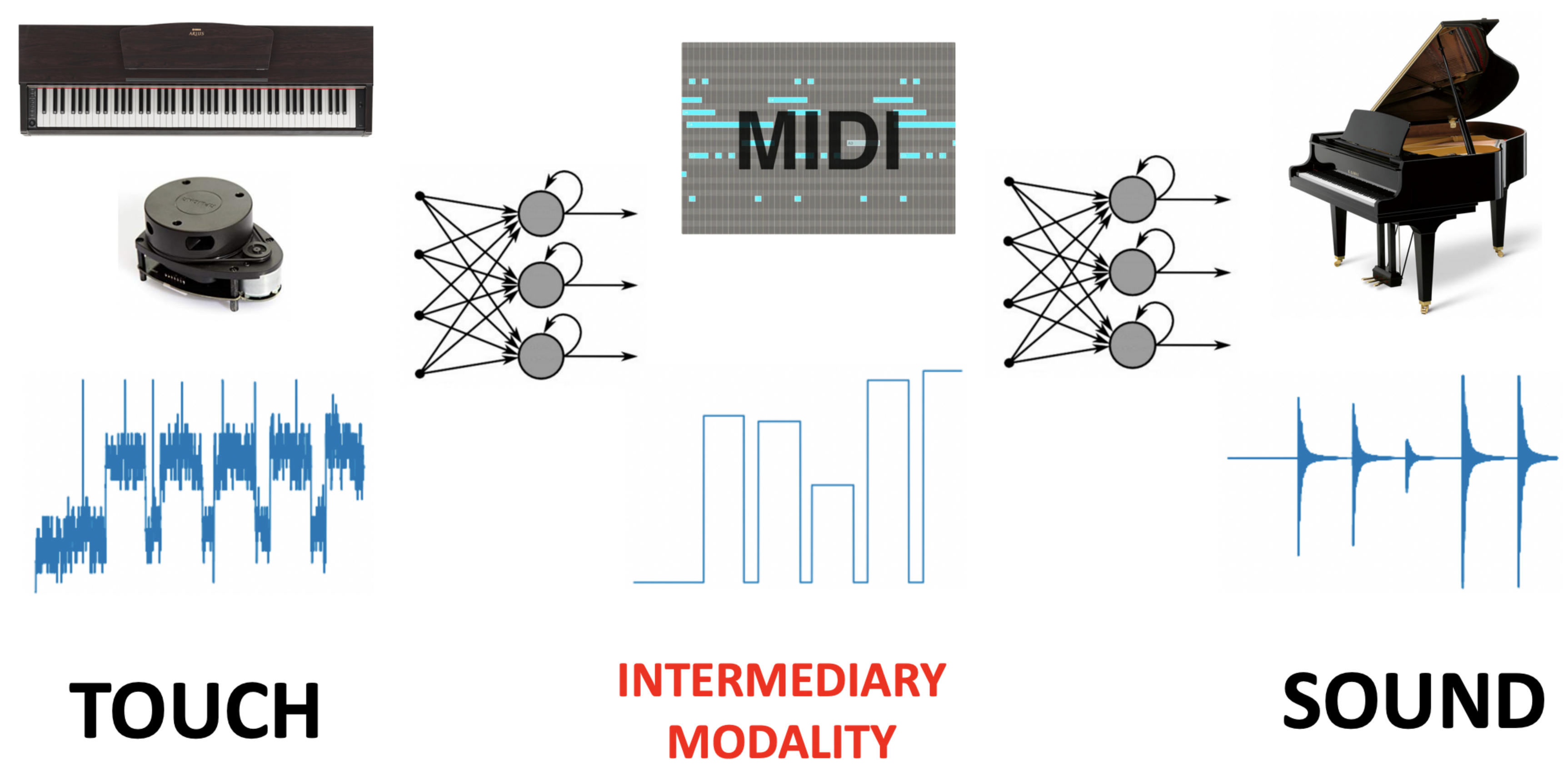

Figure 1: Full model architecture

Experimental Results

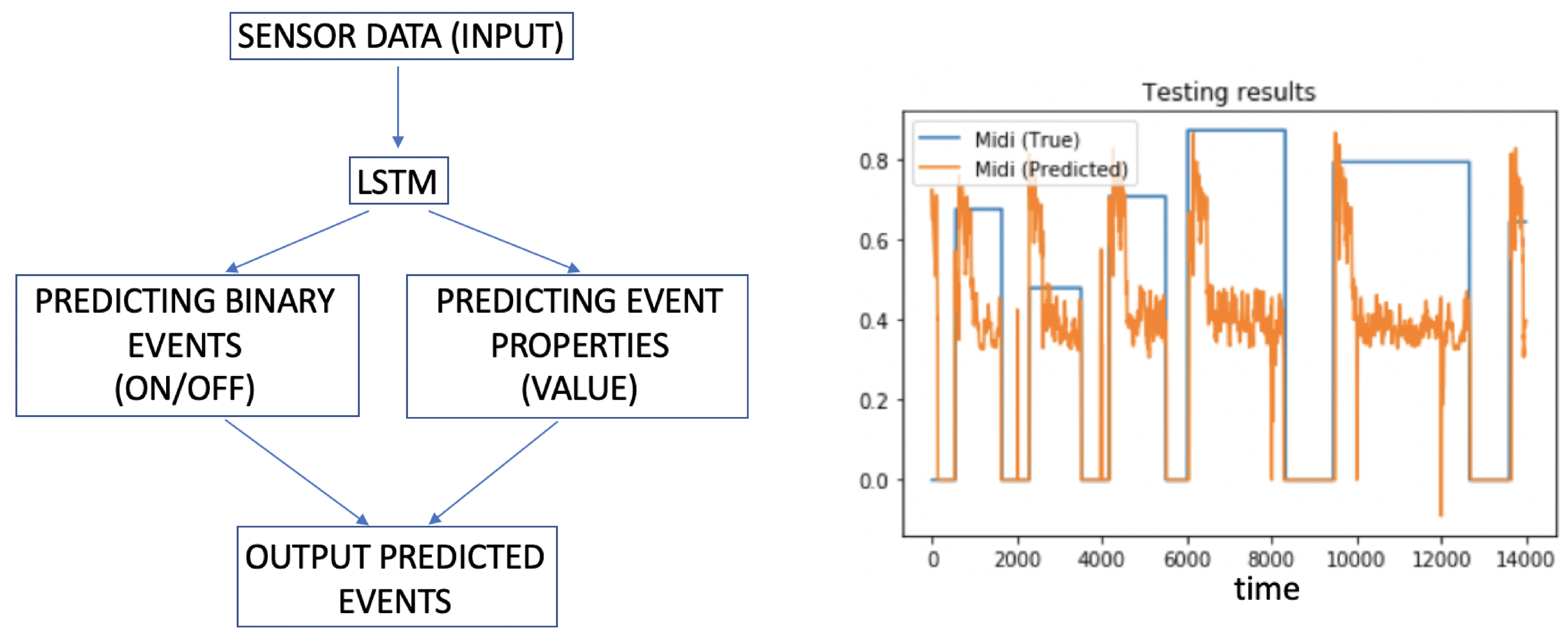

Figure 2: RNN architecture for Laser → MIDI training (Left). RNN training result (Right).

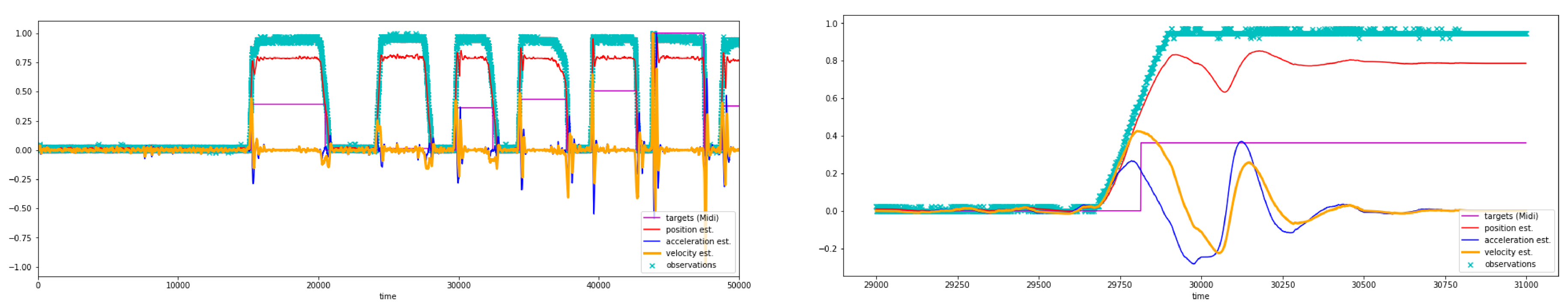

Figure 3: Using Kalman filter for Laser → MIDI task. Right is a zoomed-in sight.

Conclusions

We have utilized a multi-modal approach to the problem of predicting intermediary modalities such as MIDI velocity from touch data using a binary output RNN, with greater success by using a Kalman filter.

While the touch → MIDI training showed signs of success, more work must be done on the sound generation aspect, with conditioning to pitch and MIDI velocity. WaveNet provides a good starting point; however, more research and modifications are necessary for full completion of the entire model.

Future Work

- Utilize a greater dataset for sound generation training and explore parallelization and other tools for time efficiency.
- Work on real-time conversion of the process from key press to sound generation.
- Consider other advanced models such as conditional Generative Adversarial networks for sound generation.

Acknowledgments

We would like to thank the CS229 Course Staff and Anand Avati for helpful suggestions.