Generating Complex Melodies on the Edge TPU
Introduction

• On device inference allows user to build safe, private, fast products.

The picture is taken from https://coral.ai/
Motivation

• Generative Music Model, such as RNN is a good candidate.
• Local AI is interesting and beneficial!
Overall architecture

System Overview
Overall system
Background

• Machine Learning in Music Generation
  – Markov Model.
  – RNN
  – LSTM
  – Sequence to Sequence
  – GAN
Background

• Edge TPU
  – Coral Platform
  – Dev Board

An Edge TPU, the picture is taken from https://coral.ai/products/dev-board/
Background

• TensorFlow Lite
  – Converter
  – Interpreter
Design details

- Generative Model
- Data processing
- Dev Board Preparing
Design details

• Generative Model
  – LSTM
  – Model training
  – Model Converting
Design details

- Data Processing
  - 0-127 Note-on event
  - 128-256 Note-off event
  - 256-288 Velocity event
  - 288-388 Time-shift event
Design details

- Dev Board
  - Required hardware and software
  - Flashing the system image
  - Connecting the Dev Board with host computer
  - Ready to go!
Experiment

- Converted Model
- Dev Board
- Keyboard
- External speaker equipment
- Video
Video
Evaluation

• Model evaluation
  – Loss

Loss value of the LSTM model
Evaluation

• Converted Model with quantization VS Converted Model without quantization
  – Size
  – Speed

<table>
<thead>
<tr>
<th>Model</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model_with_quan</td>
<td>5M</td>
</tr>
<tr>
<td>Model_without_quan</td>
<td>19M</td>
</tr>
</tbody>
</table>

Sizes of the Model without Quantisation and the Model with Quantisation

<table>
<thead>
<tr>
<th>Model</th>
<th>Average time</th>
<th>Sum time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model_with_quan</td>
<td>0.334</td>
<td>2.313</td>
</tr>
<tr>
<td>Model_without_quan</td>
<td>3.336</td>
<td>23.132</td>
</tr>
</tbody>
</table>

The iteration time for the Model without Quantisation and the Model with Quantisation
Conclusion

From the results of our experiment

– Generative LSTM Model
– Optimized TensorFlow Lite Model
– A working system
Further work

• Generative Model
  – More complex model
  – Larger dataset

• Self-contained Instrument
  – More Friendly
  – Business use
Acknowledgement

• Many thanks to my supervisor, Dr. Charles Martin for his helping in my project. I can’t complete my project and achieve the success without his guidance.
Question

• Thank you for listening!
• Any questions?