NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

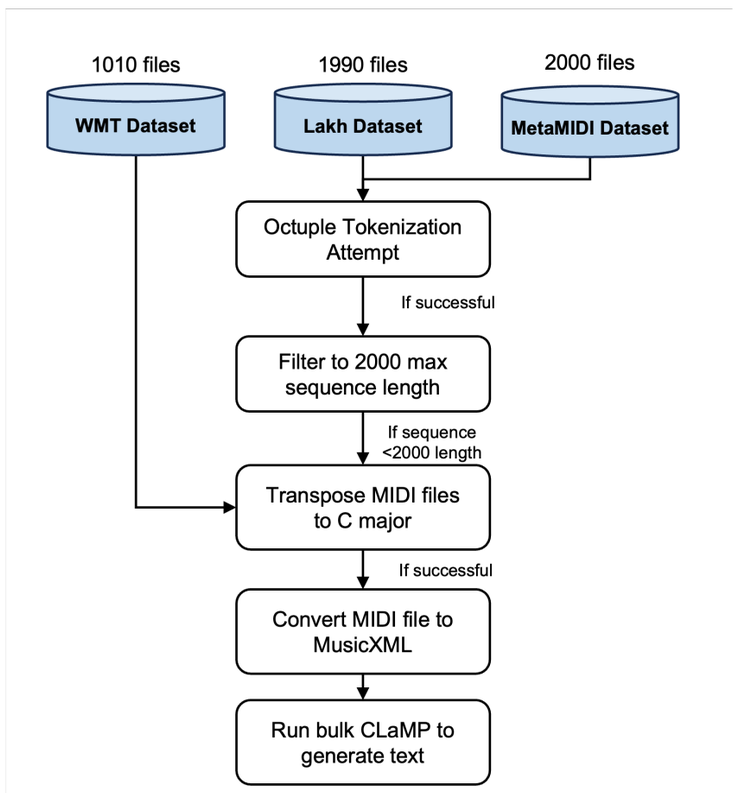
Music Generation with Deep Learning Techniques

Generating expressive and varied multimodal music from textual prompts

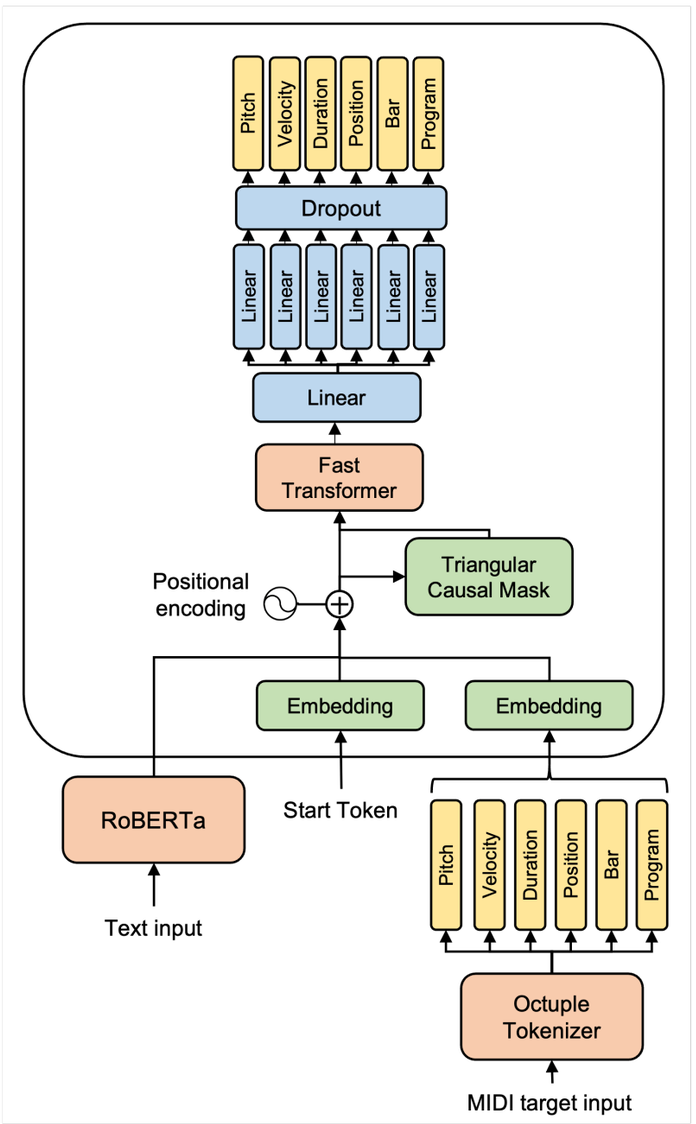
Student: Paul Solomon Low Si En Supervisor: Dr. Alexei Sourin

Recent research on music generation has shifted towards transformers and the exploration of multimodal relationships between text and music. Hence, this project explores the use of a Fast Transformer with neural network architectures to generate music from a textual prompt.

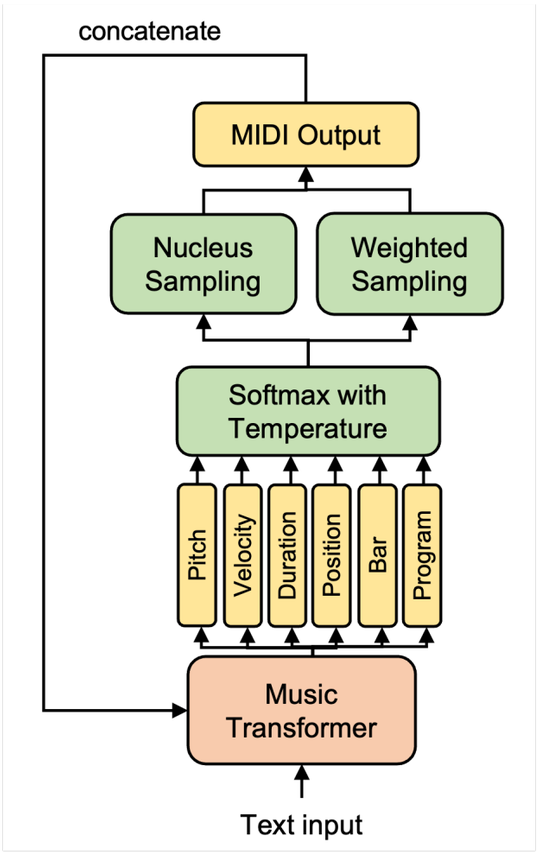
Dataset Processing



Transformer Design



Hyperparameter Evaluation



CLaMP generated the top 30 textual descriptors for each of the 5000 MIDI files used in the dataset. The textual input and musical scores were first processed by RoBERTa and Octuple before being fed into the Music Transformer with a fast transformer base. The lowest average loss values for 6 Octuple parameters defined the final model, with hyper parameter training done to promote coherent and varied outputs.

Results

- ✓ Model generates music that is likeable or expresses the textual prompt
- ✓ Model generates varied music with different styles and instrumentation
- ✓ Generated music was adapted and performed by researcher

Listen to Excerpts with their prompts by scanning this QR code