Music Synthesization with RNNs

Simon Fraser University

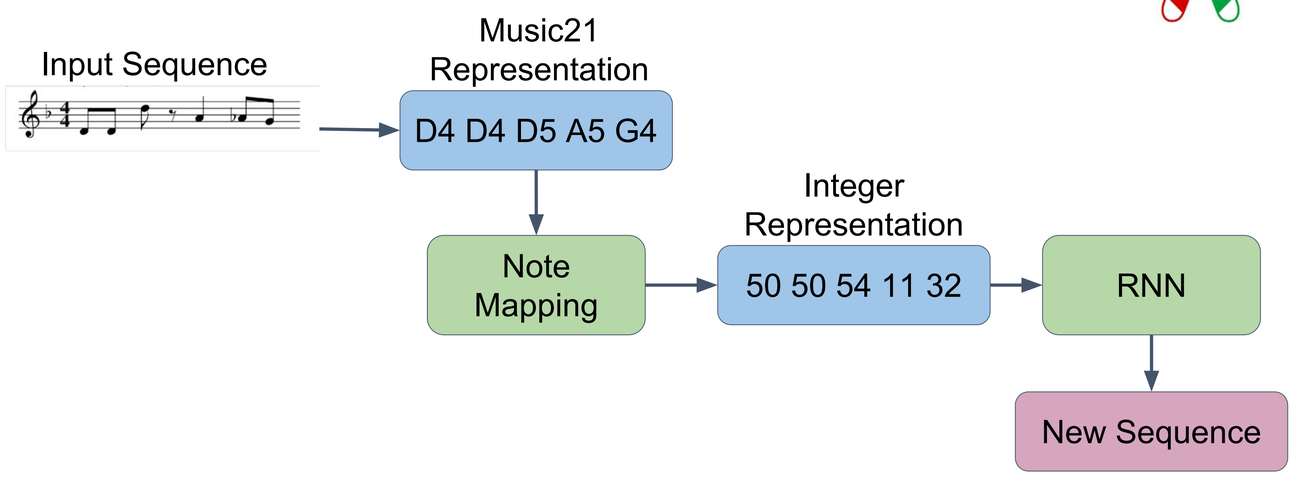
Tyler Pham, Jacob Jensen, Rulai Hu, Kinnan Kwok

OBJECTIVE

Build a model that takes in a fragment of music and outputs a predicted continuation of the music. The objective is analogous to sequence prediction, where given a segment it predicts the subsequent segments using learned weights from training examples. Musicians often improvise songs based off of a given musical fragment. We'd like to see if a machine can do the same.

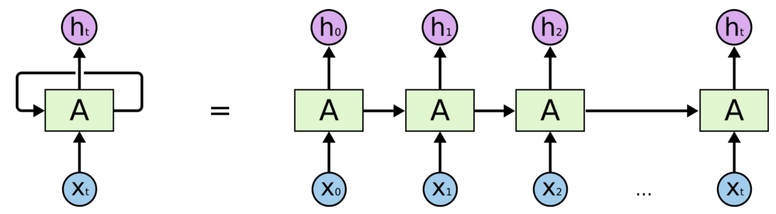
DATA REPRESENTATION

Midi Dataset	Sequences	Unique notes	Files
Holiday Music	50168	324	82

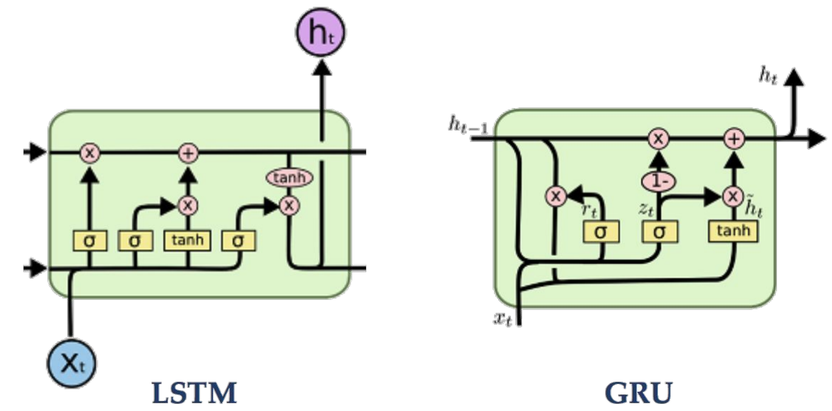


MODEL

Recurrent Neural Network



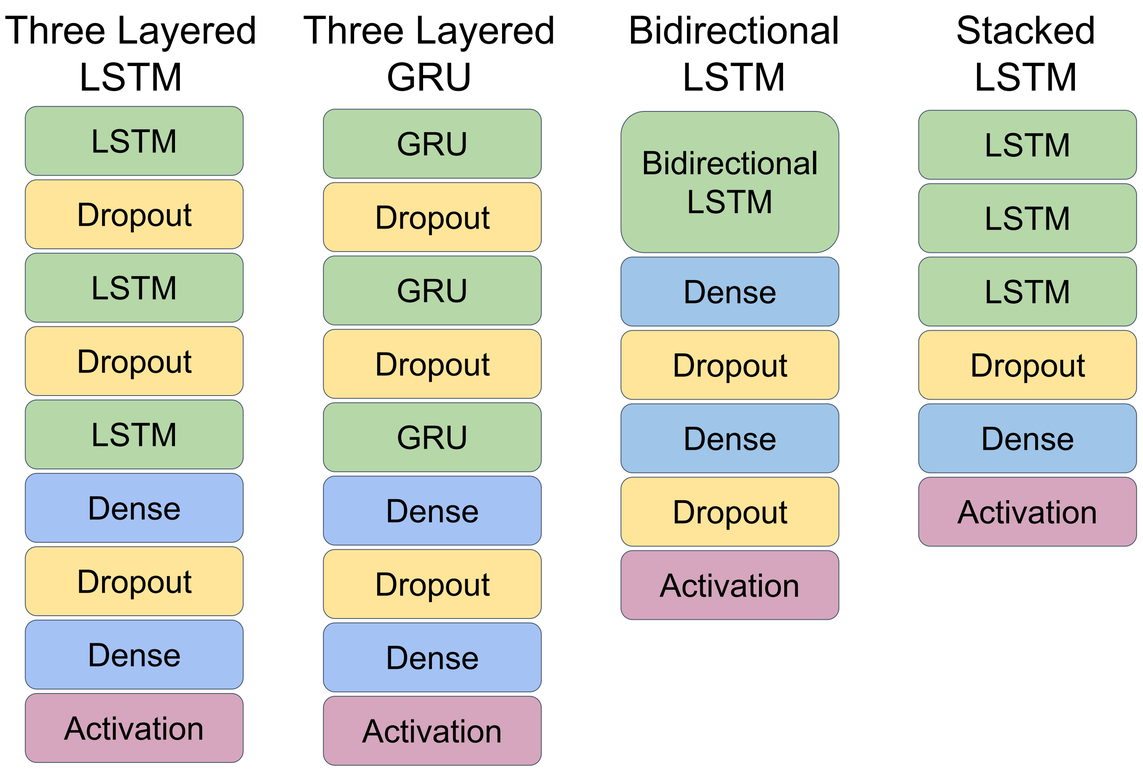
LSTM and GRU



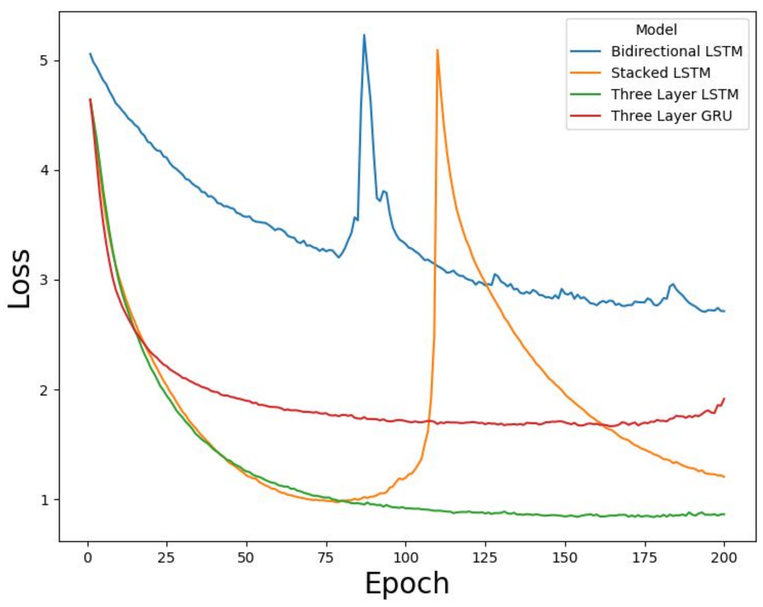
Categorical cross-entropy

$$H(p,q) = -\sum_x p(x)\log(q(x))$$

EXPERIMENTS



Comparisons were made on four different types of RNN units: three layered LSTM, three layed GRU, bidirectional LSTM and stacked LSTM.



	Best Epoch	Max Loss	Total Time
Three Layer LSTM	178	0.8394	4h 30min
Bidirectional LSTM	194	2.7070	2h 45min
Stacked LSTM	79	0.9755	4h 3min
Three Layer GRU	165	1.6670	3h 48min

- Categorical cross-entropy was used for loss
- The musicality of the output is subjected to human interpretation

MUSIC PREDICTION

Modified RNN to predict a segment of music instead of generating music

A test set of 10 midi songs were used to evaluate the output

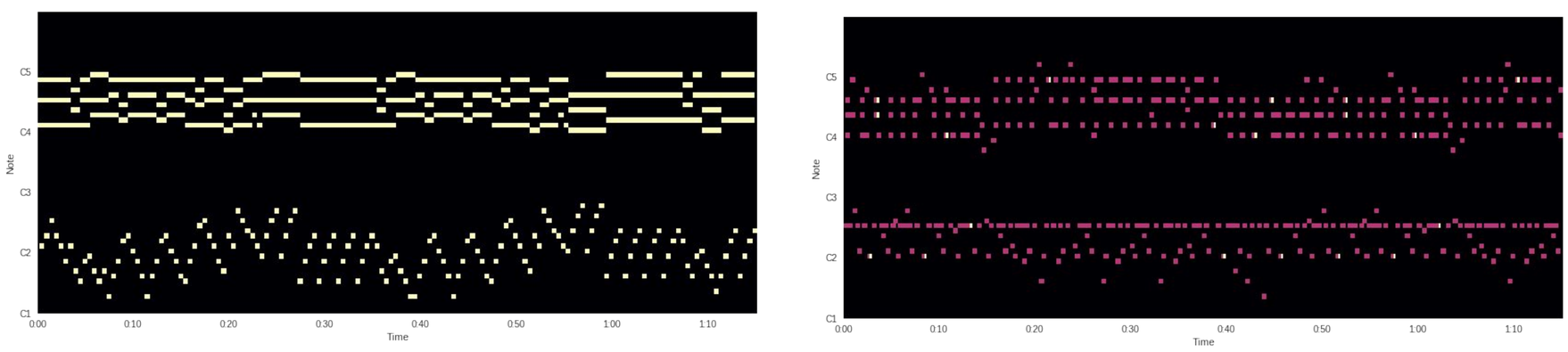
Similarity of predicted versus expected output is calculated by:

$$1 - \frac{EditDistance(expected,\ predicted)}{TotalNumberOfNotes}$$

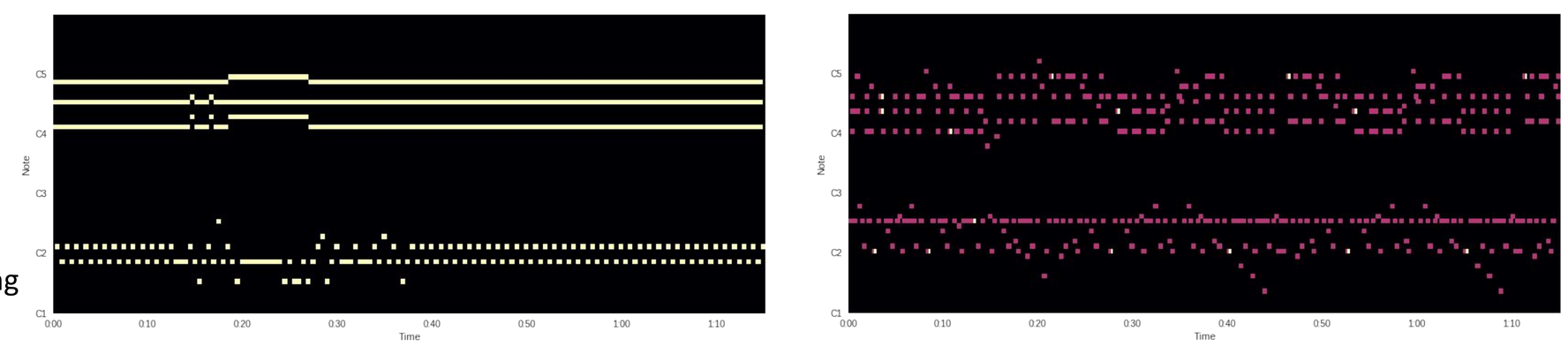
Three Layer LSTM	23.8%
Bidirectional LSTM	23.5%
Stacked LSTM	24.7%
Three Layer GRU	22.2%

The network is unable to produce notes that it has never seen before in the training set

Expected:



Prediction:



Typical outputs from the RNN.