NeuroJazz

MIE324 Project
Final Presentation

Jiyu Nam and Ryan Do

GOAL AND MOTIVATION

Motivation

RNNs have been used previously for music generation.

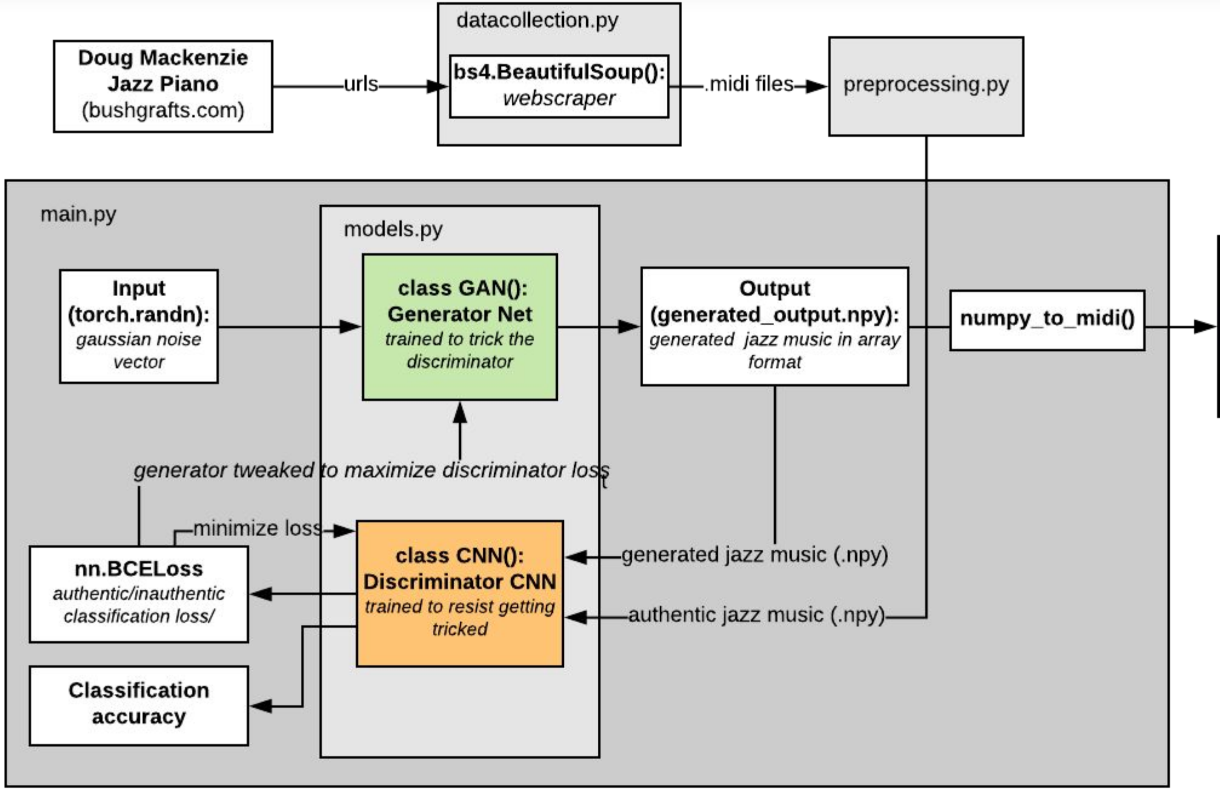
Jazz contains subtle patterns in multiple dimensions.

Maybe another way would be better?

Goal

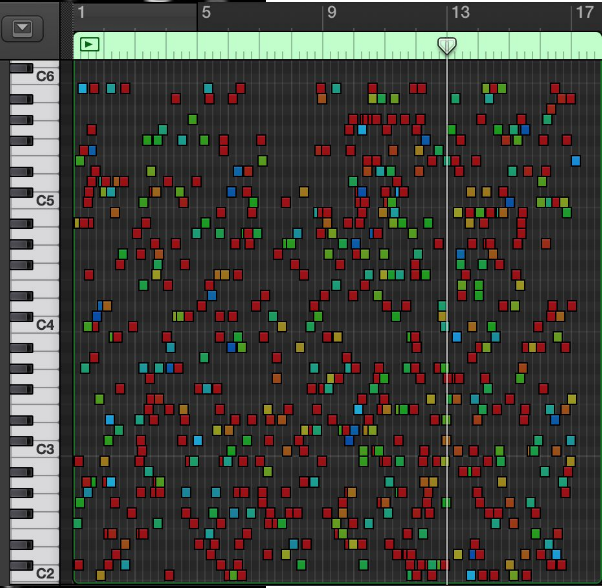
To implement a Deep Convolutional Generative Adversarial Net (DCGAN) for generation of jazz music.

OVERALL SOFTWARE STRUCTURE

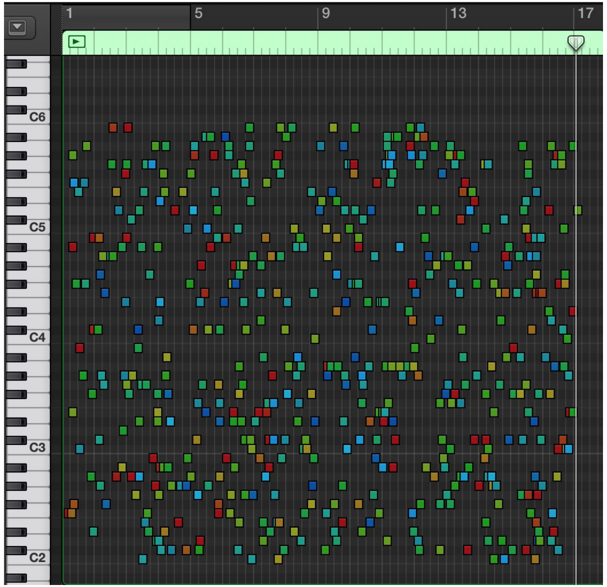


LET'S LISTEN TO THE RESULTS!

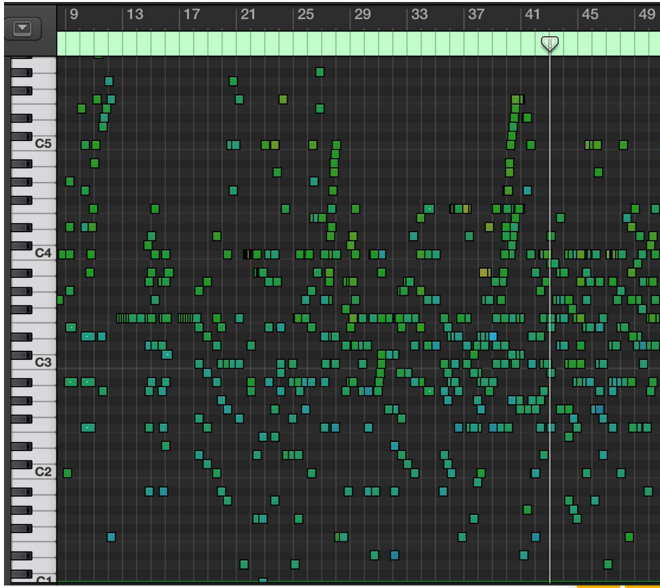
Untrained Generator Output



Trained Generator Output



Training Data (Authentic Music)



DATA COLLECTION & PREPROCESSING

- **Source (authentic data for training):** https://bushgrafts.com/midi/

- **Pre-Processing:**

1. Filter to only piano components
2. Parse midi files into stream object
3. Extract pitch, offset, duration and velocity information
4. Convert to 2D numpy arrays
5. Crop to relevant pitches and to shorter length
6. Label as ‘authentic’
7. Save instance

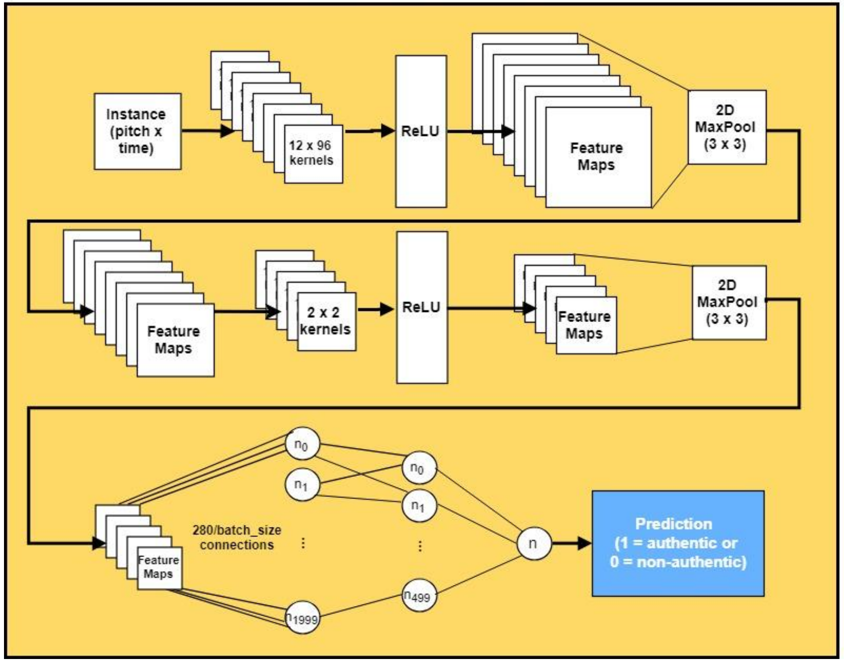
PREPROCESSED DATA

Timestep (1/24 of a beat)

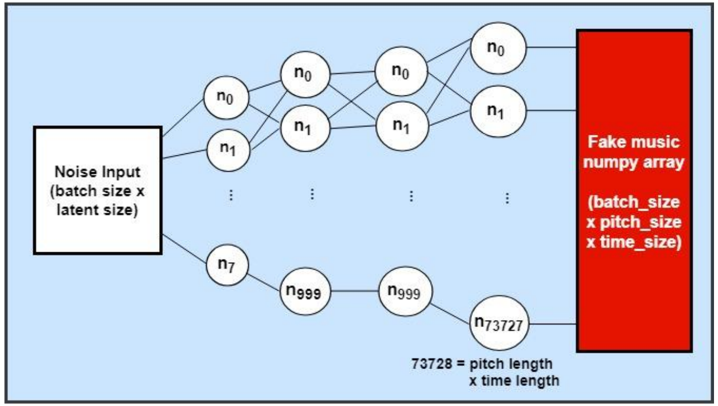
Pitch

	0	1	2	3	4	5
0	0.3779527559055118	0.3779527559055118	0.3779527559055118	0.3779527559055118	0.3779527559055118	0.3779527
1	0.0	0.0	0.0	0.0	0.0	0
2	0.0	0.0	0.0	0.0	0.0	0
3	0.0	0.0	0.0	0.0	0.0	0
4	0.0	0.0	0.0	0.0	0.0	0
5	0.0	0.0	0.0	0.0	0.0	0
6	0.48031496062992124	0.48031496062992124	0.48031496062992124	0.48031496062992124	0.48031496062992124	0.48031496
7	0.0	0.0	0.0	0.0	0.0	0
8	0.4094488188976378	0.4094488188976378	0.4094488188976378	0.4094488188976378	0.4094488188976378	0.4094488
9	0.36220472440944884	0.36220472440944884	0.36220472440944884	0.36220472440944884	0.36220472440944884	0.36220472
10	0.0	0.0	0.0	0.0	0.0	0
11	0.49606299212598426	0.49606299212598426	0.49606299212598426	0.49606299212598426	0.49606299212598426	0.4960629
12	0.0	0.0	0.0	0.0	0.0	0
13	0.0	0.0	0.0	0.0	0.0	0
14	0.0	0.0	0.0	0.0	0.0	0
15	0.0	0.0	0.0	0.0	0.0	0
16	0.0	0.0	0.0	0.0	0.0	0
17	0.0	0.0	0.0	0.0	0.0	0
18	0.0	0.0	0.0	0.0	0.0	0
19	0.0	0.0	0.0	0.0	0.0	0
20	0.0	0.0	0.0	0.0	0.0	0
21	0.5984251968503937	0.5984251968503937	0.5984251968503937	0.5984251968503937	0.0	0

DISCRIMINATOR CNN



GENERATOR GAN



Note: Receptive field of discriminator was designed to span multiple bars of music and a wide pitch range.

MACHINE LEARNING MODELS

RESULTS OF TRAINING - Notable Metrics

Training Accuracy of Discriminator

- **Begins high (~78%) and then stabilises at ~50%**

  i.e. Good at distinguishing non-authentic and authentic music, but as training progresses, is 'fooled' by the generated music

  - Encountered issues where discriminator trains faster than generator

Loss of Discriminator and Generator

- **Both highly fluctuating in the beginning, then stabilises (~1.3 and ~0.7 respectively)**

  i.e. Discriminator is improving (in distinguishing authentic and non-authentic music) at an equivalent rate as the generator

KEY LEARNINGS

Improvised music is extremely diverse - the musical commonalities between samples are subtle.

- ➢ **A larger generator/ discriminator model may be able capture intricate jazz features to a higher degree.**

GANs are difficult to train! Evaluating training quality is subjective.

- ➢ **Surveying multiple people to evaluate results is useful**
  - ■ **Amazon Mechanical Turk?**
- ➢ **Replicating subtle features in the generated data requires a large amount of computational power.**

FUTURE WORK

Use of transposed convolution ("deconvolution") in generator network over fully connected layers

- ➢ **Greater ability to replicate spatial patterns**

Implementing a "conditioner" CNN

- ➢ **Enhances the generator to give it a memory property**

Incorporating more features into training data

- ➢ **Multi-instrumental tracks**
- ➢ **Tempo**

Check us out on SoundCloud!