Milestone 3 Report

Team members: Juliana Choi (88867312), André Santiago Correia (59366682), Harman Sihota (99773566), Charles Wang (87971487)

Main goal of milestone

Decide on how the dataset will be represented and how the model we create will be trained and then create the samples and labels accordingly. Start to create/research the model.

Current state of project

We decided how to represent the data and how we would train a model to generate sheet music from audio. Our model will be a convolutional neural network. We will convert the audio to spectrograms and these images will be the input to our model. The labels are one hot encoded vectors/matrices.

More specifically, samples of our dataset will consist of images of the spectrogram representation of a song and labels that are matrices where each row represents a second of the piece of music and each column represents a certain note. So if the sample is 90 seconds long, the label will be a matrix with 90 rows and 128 columns since there are 128 midi notes.
If note j is played at time i, label[i][j] = 1.
If this note becomes "off" at time k, label[k][j] = -1
Otherwise, each position of the matrix is 0.
The time a note is played and the time it becomes "off" are both available from the spectrogram of an audio sample.

The code for creating this dataset is complete and tested with two minor exceptions:

- We need to trim the blank white edges and plot axes from the spectrogram image.
- We need to find a more efficient way to convert .wav files into spectrogram images.

We have also begun considering more of the details of the training process and the model. As stated above we are planning on using a CNN. From the feedback that we received on milestone 2, we are strongly considering an already trained CNN such as the resnet models. This source also has pretrained models specifically for audio applications that we could potentially use.

The output from the model will be a matrix with the same format as the label. Since this will contain which note is played at what time, and for how long, we will be able to create sheet music from this.

Technical Challenges

Finding a suitable pretrained model or creating our own CNN
Connecting python code to backend of website
Finding a more efficient way to convert .wav files into spectrogram images
Create the front end of the website
Deciding on loss function

Team tasks

Juliana Choi

- Research on how our model can tackle different aspects at once (DONE)
- Decide on how to represent data, labels, output based on research (DONE)
- Research what type of model would be best for our implementation (DONE)
- Write the Final Report (IN PROGRESS)
- Write the code for the model based on pretrained models (IN PROGRESS)

Andre Correia

- Continue working on the non-ML component of the project by learning more about HTML, JS and CSS. (IN PROGRESS)

Harman Sihota

- Decide on how to represent data, labels, output based on research. (DONE).
- Write the code for creating dataset that is ready to be used for training (DONE).
- Do any other necessary preprocessing (converting midi files into .wav, converting stereo audio files into mono)
- Research potential pretrained models that we could use (IN PROGRESS).
- Document code (IN PROGRESS).
- Write Milestone 3 report (DONE).

Charles Wang

- Finding website host and creating template for website (DONE)
- Research into backend implementation of python code (IN PROGRESS)
- Research on identifying sounds as notes in sheet music for model (IN PROGRESS)