Musical Notes Recognizer

Sakshamdeep Singh
50475857
ssingh86@buffalo.edu

Shashank Vivek Baratwaj
50479686
svivekba@buffalo.edu

1 Scope of the project

1.1 The Application

The proposed project aims to develop a computer vision system capable of accurately recognizing and interpreting musical notes. The objective is to train a model to recognize a variety of musical symbols and convert them into a machine-readable format. Additionally, the system will enable the computer to play music based on the interpreted notes.

This proposed model can be used in the music industry to aid in the transcription and digitization of sheet music. Currently, manual transcription of sheet music is a time-consuming and error-prone process. With the computer vision system, sheet music can be quickly and accurately transcribed into a machine-readable format, which can save both time and effort.

Secondly, the system can be used in music education to facilitate the learning process for students. Students can easily practice playing musical pieces which can greatly enhance their learning experience and improve their overall musical skills.

1.2 State of the Art

The current state-of-the-art implementation analyzed in this paper is a likelihood-frequency-time (LiFT) analysis using support vector machines (SVMs). Features are extracted from the isolated note samples, and classification of instruments and notes is performed where the results of the study indicate that the LiFT analysis approach achieved correct classification ratios for 19 different instruments and 36 notes.

We plan on developing a computer vision model which can additionally enable the computer to play music based on the interpreted notes.

1.3 Inputs and Outputs

For this proposed project, the inputs will consist of sequences of images containing musical notes, where each note on the sheet represents a primary input for recognition. The output of the system will be the accurate recognition of the notes and the ability to play the entire sheet as music.

Specifically, the input data will be a 30-second sequence of images containing musical notes that will be generated and input into the network for accurate prediction and interpretation.

Furthermore, if we integrate a combination of Generative Adversarial Networks (GANs) and Long Short-Term Memory (LSTM) networks, the system will have the ability to generate and play high-quality melodies. This approach represents a significant advancement in the field of automatic music composition and can have various real-world applications, such as music education and composition.

Preprint. Under review.

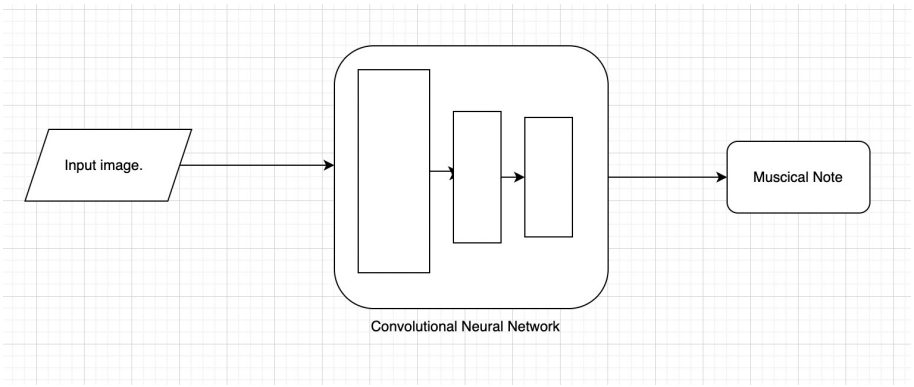

Figure 1: Inputs -> CNN Model -> Outputs

2 Data

We will utilize the dataset Musical-Notes for our proposed project, which comprises five distinct musical symbols designated for classification, namely: Whole Note, Half Note, Quarter Note, Eight Note, and Sixteenth Note. Each classification category includes precisely 1000 data samples, bringing the total number of images in the dataset to 5000, with each image sized at 64*64 pixels.

3 Coding Resource Requirements

The implementation of this model will require several coding resources, including Python programming language, deep learning framework like PyTorch, and relevant libraries such as OpenCV, midiutil and simpleaudio. Using these resources, we will build the whole project on our own.

4 Computational Resource and Effort Requirements

The development of a reliable computer vision model for recognizing and interpreting musical notes requires a decent amount of computational resources which we can provide on our own. This will include a pre-processing a large dataset of sheet music, along with powerful computing hardware and software tools for training and testing and optimizing the model.

5 Evaluation

The accuracy of the computer vision model will be evaluated using a range of metrics designed to measure its effectiveness in recognizing and interpreting musical notes. These metrics will include:

- Precision: This metric measures the proportion of musical notes that are correctly identified by the model out of all the notes it identifies. A high precision score indicates that the model accurately recognizes musical notes.
- Recall: This metric measures the proportion of musical notes that are correctly identified by the model out of all the notes present in the sheet music. A high recall score indicates that the model is able to identify a high proportion of musical notes in the sheet music.
- Accuracy: This metric combines precision and recall to provide an overall measure of the model's accuracy in identifying musical notes. A high F1 score indicates that the model accurately recognizes and interprets musical notes with a high degree of accuracy.

6 Project Expectations

The proposed computer must meet several expectations, including accuracy, robustness, scalability, efficiency, and user-friendliness. By meeting these expectations, the model will be able to revolution-

ize the field of music transcription and open up new opportunities for musicians and music enthusiasts alike.

The project expectations comprise four key aspects:

- Accuracy: The computer vision model must accurately recognize and interpret musical notes.
- Robustness: The model must handle variations in musical notation and image quality.
- Scalability: The model must handle large volumes of sheet music data.
- Efficiency: The model must operate efficiently in real-time.

7 Progress Reports

7.0.1 Progress to date - March 28, 2023

As of today, we have made significant progress in our research on existing models, conducted exploratory data analysis on datasets related to music note recognition, and also achieved some progress in playing a tune using Python.

7.0.2 Progress to date - April 11, 2023

Our project has made significant progress to date in achieving our goals of developing a neural network model that can accurately recognize and classify musical notes. We have collected and organized a large dataset of images of musical sheets, normalized it, and labeled the images with the corresponding musical notes. Using deep learning frameworks PyTorch, we have built and trained a CNN model to recognize and classify the musical notes with a high degree of accuracy. We have achieved an accuracy rate of 98% on our validation set, indicating that our model is performing well and is ready for further testing and optimization.

7.1 Changes to original proposal

- Flowchart of the application is added to the proposal.
- Links to the datasets were already added in the original proposal but now we've added a small icon in front.
- Coding resources: Whole project will be coded independently on our own.

7.2 Planned task

7.3 Schedule through project completion

Our proposed timeline for the drowsiness detection deep learning project is as follows:

- Proposal (March 5): Extensive research will be carried out on the current models and an analysis of datasets pertaining to music note recognition will be performed.
- Design and Development (March 5 - April 20): During this phase, we will preprocess the data and develop our project components and features using the Convolutional Neural Network (CNN) architecture.
- Hyperparameter Tuning and Improvements (April 11 - April 25): During this phase, we'll be doing extensive testing and fine tuning of our model.
- Project Submission (April 25 - May 5): We will submit our final report, demo and running code in this week.

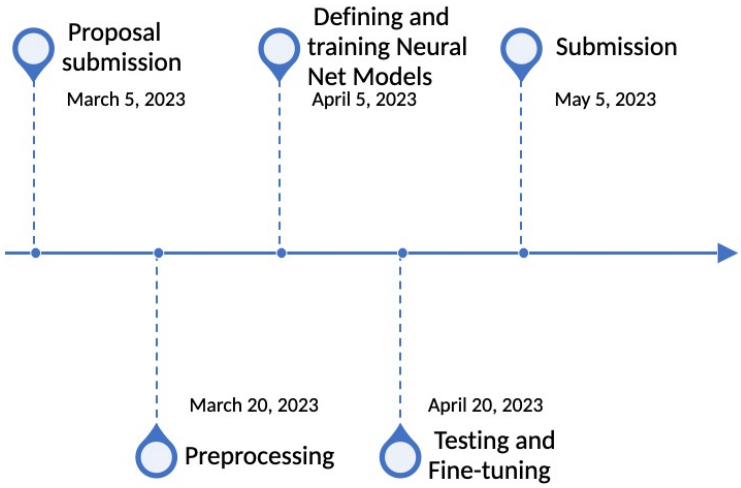

Figure 2: Tasks Timeline