Mozart Vision

Arjun Ramesh Kaushik
UB ID - 50413327
kaushik3@buffalo.edu

Tilak Sharma
UB ID - 50479412
tilaksha@buffalo.edu

Abstract

The goal of the project is to develop a Computer Vision system that automates - interpretation of guitar sheet music and reproduces a written version of it in English.

1 Problem statement

In today's world, music-document digitization tools are not very robust. Artists usually have a non-intuitive user-experience when they manually create/write music. Our solution aims to solve this by detecting different pitches/notes and expressing them in simple English for a layman to understand.

2 Input and Output Data

The below images are examples of input and their corresponding outputs.

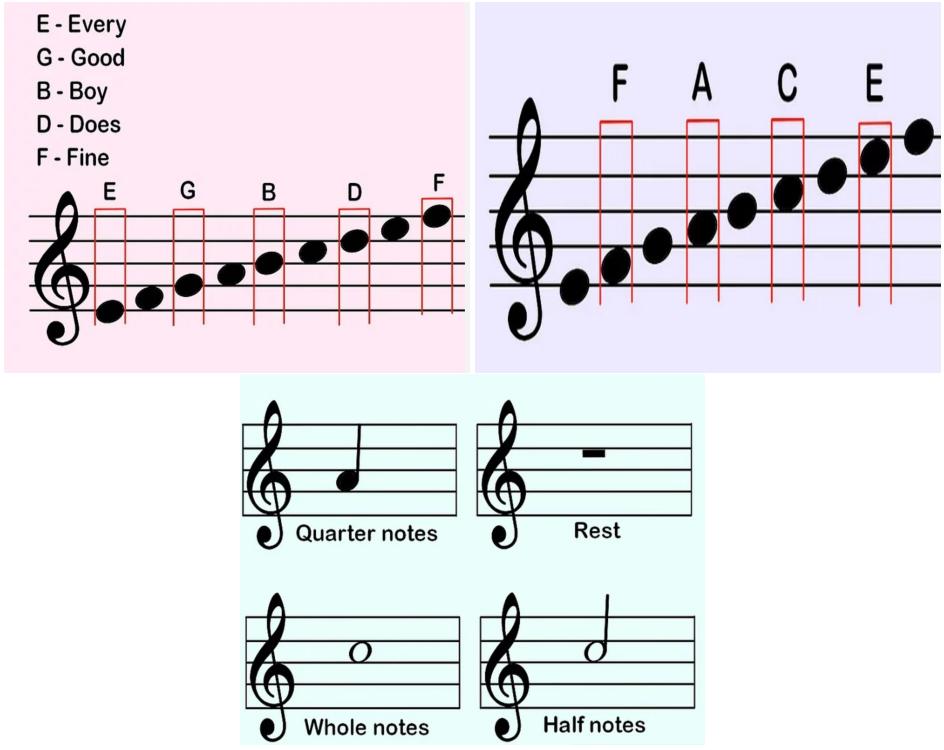

Figure 1: Input and Output

3 Dataset

1. `https://apacha.github.io/OMR-Datasets/`
2. `https://www.kaggle.com/datasets`

4 Libraries

1. OpenCV
2. TensorFlow

5 How to run the project

1. Install the dependencies as listed in the library
2. Run *main.py* file. By default, a new model is trained. But, if you have a pre-trained model run *main.py* with the argument *train_classifier=false*.

6 Algorithm

We have developed the project completely without the use of online resources/code.

The algorithm used in the project is detailed below -

1. Hough transform on the image dataset to detect and store coordinates of the extreme points of horizontal staff lines. This will be used to find the relative position of the pitch with respect to the staff lines.
2. Use morphology to remove the staff lines. Removing the lines resulted in a patchy image, this is fixed by performing another morphological operation.
3. Break a single bar into individual notes using image segmentation
4. Find corners of the notes in segmented image
5. Find centroids of the blobs of the notes. If blob is on a line, then associate them to the nearest staff line. Else, associate to the space between the lines.
6. Based on the position of the note - in space or on line - we derive the pitch, fret and string.
7. Convolutional Neural Networks are used to classify the notes as whole note/half note/quarter note.

7 Results

With our approach and algorithm, the system has an accuracy of 89%. That is, on an average, the system detects 32/36 notes in the input image. The implemented CNN model isnt completely robust, as it detected only 32/36 notes despite it having a training accuracy of 97.98%. Lastly, for the scope of this project, we have modelled the project such that only music sheets with full note/half note/quarter note can be converted.

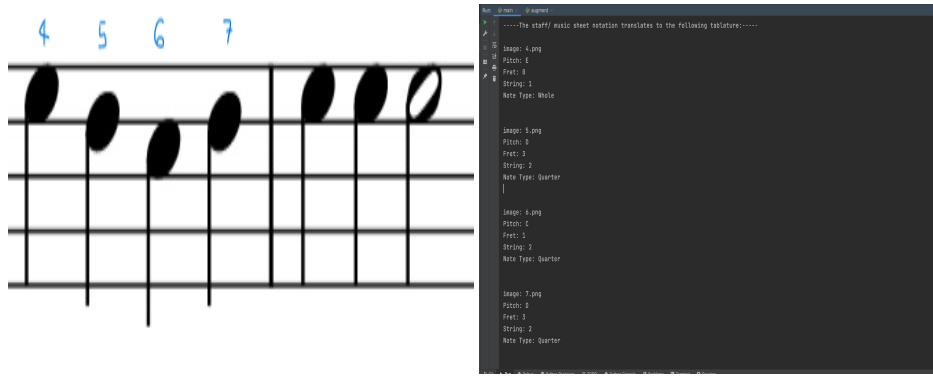

Figure 2: Input and Output

8 Analysis

We are looking to generate tablature for basic sheet music played on guitars - this covers music like "Ode to Joy', "Mary had a little lamb", or mobile ringtones. We expect our system to fail for handwritten sheet music,complex scores and data with class imbalance.

We have also only covered cases that have full note/half note/ quarter notes in them. So, our classifier and system can only read music sheets with such notes.

9 Bibliography

1. https://arxiv.org/abs/2006.07885
2. https://archives.ismir.net/ismir2019/paper/000033.pdf
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9184194/