SUMMER RESEARCH INTERNSHIP

Extraction of melodies with pitch classes from images of HCM compositions in Bhatkhande Notation System

Manjot Singh*, **Dr. Pratik Shah†**

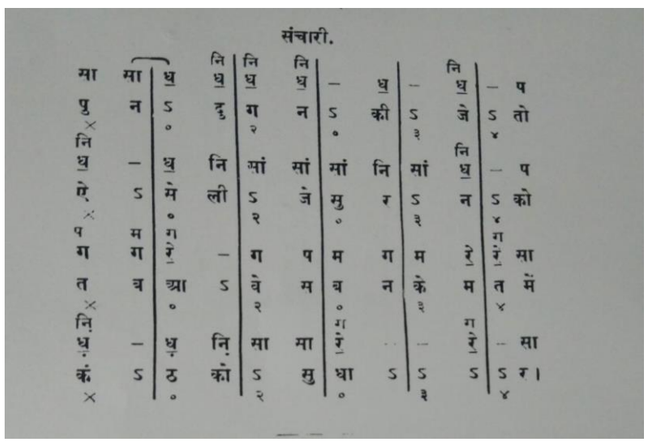
*201951090@iiitvadodara.ac.in, †pratik@iiitvadodara.ac.in

Introduction

Motivation: Bridging the gap between written medium of musical notation system and picturing the beauty of the melodies, to be able to convert your favorite melodies instantly from unexciting texts in your old books to audio files on your phone or laptop.

Objective: Conversion of Music Notes from images of Bhatkhande Notation System into MIDI file (audio file) involving Images Processing using *OpenCV*, Deep convolutional neural network modeling using *TensorFlow* and *Keras* and extracting those notes to be converted into a playable audio file using *music21* toolkit for computer-aided musicology.

Problem Statement



Task: From images of HCM compositions in Bhatkhande notation, extract the melodies with pitch classes.

Conclusion

- Only one two parameters were changed to achieve these results in the test (Search Window and Kernel Size) and once changed these parameters were kept the same for all the test cases(all 4). So different values would significantly improve the model.
- Accuracy could have been improved a lot, the new dataset really did wonders for this model, only hits taken by the accuracy were by the accidental detection of numbers as notes.
- If it were only the notes and not the number, the accuracy could have been above 90%. So, more tuning needs to be done in order of the hard coded values to be automatically computed instead of relying on the user input.

Key References

[1] Sha, P., & Dong, X. (2021, August 21). Research on adolescents regarding the indirect effect of depression, anxiety, and stress between TikTok use disorder and memory loss. International journal of environmental research and public health. Retrieved July 25, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8393543/

[2] Lee, S. (2022, January 30). Lines detection with Hough Transform. Medium. Retrieved July 25, 2022, from https://towardsdatascience.com/lines-detection-with-hough-transform-84020b3b1549

[3] Canny edge detector. OpenCV. (n.d.). Retrieved July 25, 2022, from https://docs.opencv.org/4.x/da/d5c/tutorial_canny_detector.html

[4] Alpaydin, E. (2021). Machine Learning, revised and updated edition. MIT Press.

[5] Springer. (2015). Image Processing, *Computer vision*.

[6] Connectedreams.com. (2016, December 1). Beginner's Guide to Computer Vision. Medium. Retrieved July 25, 2022, from https://medium.com/readers-writers-digest/beginners-guide-to-computer-vision-23606224b720

[7] Kang, N. (2019, February 4). Introducing deep learning and neural networks - deep learning for rookies (1). Medium. Retrieved July 25, 2022, from https://towardsdatascience.com/introducing-deep-learning-and-neural-networks-deep-learning-for-rookies-1-bd68f9cf5883

Proposed Methodology

The final approach ended up being divided into 3 parts:

❑ NOTATION RECOGNITION AND PROCESSING *(What to read in this image?)*

OpenCV is used for preprocessing of the image in which all the extra noise on the image is removed, like the color of the paper, different shades of color etc., so that it would not interfere with the detection. After this we end up with a binary version of the image with only Black (0,0,0) and White(255,255,255). Bounding Rectangles are drawn on the Edge Detected Contours on the respective co-ordinates and saved on another blank image. Then, selective filtering of the contours is done for classification and thus removing the unnecessary clutter like vertical lines, heading, numberings etc. These elements are cut out of the original binary images in order for the classification.

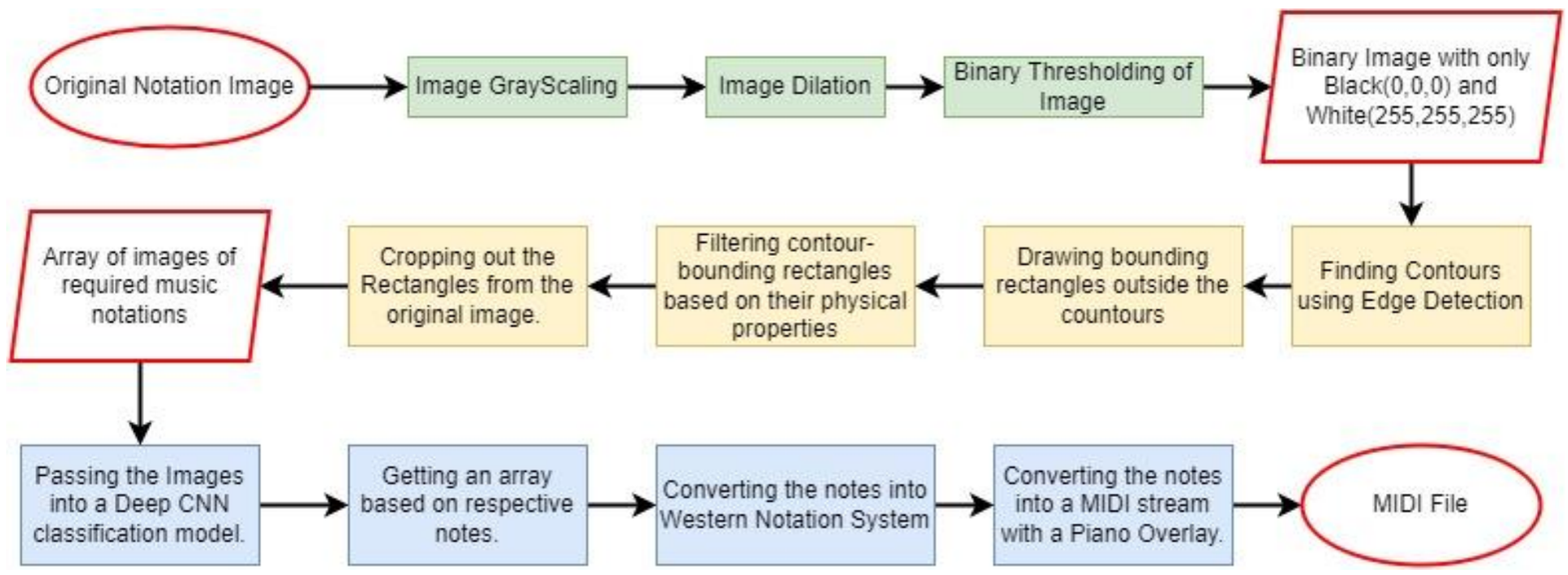
❑ NOTATION CLASSIFICATION *(What is the meaning of the read part in the image?)*

Since there was no dataset available for the Hindi music notation, screenshots of a lot of images of notations of each notes namely सा रे गा मा प ध नि were taken from the internet (600 to 900 each). After that, Data Augmentation is used to generate more synthetic images with random flips, random rotation and random zoom, converting the image count of each notes to 3 times the original (1800 to 2700 each). This Dataset is then used to generate a Deep CNN Classification model for the notes.

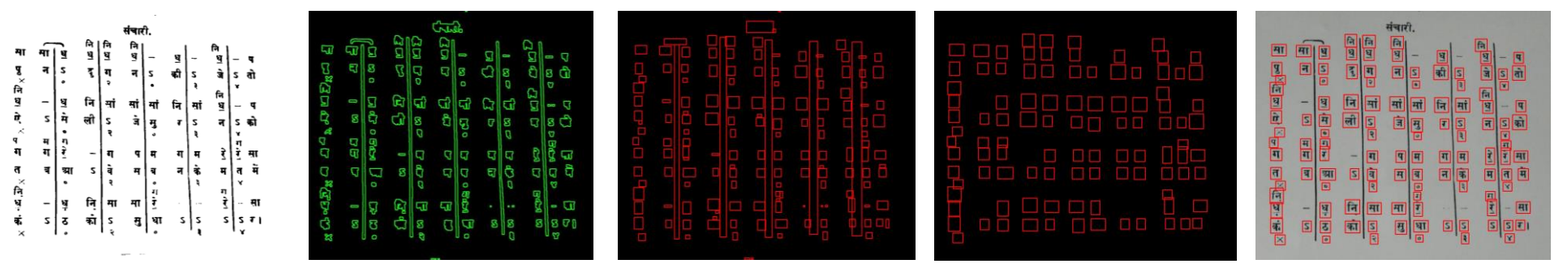
❑ AUDIO EXTRACTION *(What sound to produce in response to what was just read.)*

The image array is then used here to classify the acquired images from value 0 to 6 and then using a dictionary convert to Sa, Re, Ga, Ma, Pa, Dha and Ni. This array is then converted to stream using music21 library and converted to MIDI file and used as audio file.

Flow Chart of the Proposed Algorithm



Process of Image Preprocessing and Recognition



1. Binary Converted | 2. Contours Drawn | 3. Bounding Rectangles | 4. Filtered Rectangles | 5. Extracting Notes

Testing & Results

The model needs to be sufficient enough to be used on any notation image that contains the notes. So to check that eligibility, 4 test cases were ran on notation images from internet and the results were promising

- Classifications are made using the model.
- Then those values are checked one by one using an incremental iterator and the value is stored in an array/list in order to be cross checked against the predictions.
- After that, heat map and midi files are generated through these test cases resources and data of which can be found on the GitHub Repository of the project. *(https://github.com/mickee00000/201951090_Research_Internship_2022)

	Accuracy
TEST 1	**69**.33333333333334%
TEST 2	**78**.94736842105263%
TEST 3	**76**.66666666666667%
TEST 4	**78**.84615384615384%