Optical Music Symbol Recognition (sheet-to-music and music-to-sheet)

Neha Shah
Department of Information Technology
K.C. College of Engineering and Management Studies and Research
Thane, India
nehashah@kccemsr.edu.in

Tanvi Sawant
Department of Information Technology
K.C. College of Engineering and Management Studies and Research
Thane, India
tanvisawant@kccemsr.edu.in

Bhakti Machha
Department of Information Technology
K.C. College of Engineering and Management Studies and Research
Thane, India
bhaktimachha@kccemsr.edu.in

Sheetal Jadhav
Department of Information Technology
K.C. College of Engineering and Management Studies and Research
Thane, India
sheetal.jadhav@kccemsr.edu.in

***Abstract*— Optical Music Recognition (OMR) technology aims to transcribe Sheet music into a format that computers can understand and vice versa, allowing Musicians to write, perform, and edit music more easily by simply Capturing a photograph of a music sheet. This study proposes a system that addresses the challenges of existing OMR systems by converting music sheets into a machine-readable format and generating MIDI files. Additionally, the system can convert instrumentation music into musical sheets. The system takes an image of a music sheet as input and uses Convolutional Neural Network to identify symbols, calculate pitch and duration, and output a string corresponding to each line in the music sheet. This string can be played in MIDI or MP3 format. The major steps involved in this process include staff line removal, symbol extraction, symbol recognition, and conversion of musical scores into MIDI format. This study aims to create melodic music that can be performed by humans via a piano interface. The selection of instrument is the piano, which will be the same for both the input and output. Overall, this proposed system has the potential to revolutionize the music industry by making music composition, playing, and editing more accessible to everyone.**

Keywords—Musical Recognition, MIDI, Paino Music staff Lines, CNN

I. INTRODUCTION

Most music manuscripts are currently available only in paper form, lacking digital format. As a result, sheet music can be difficult for non-experts to read and understand. However, computers offer a potential solution to this problem. With the aid of music recognition technology, musical notes from a sheet of music can be converted into an electronically readable format, allowing a computer to play music directly from the sheet. This project aims to digitize music notation automatically by offering an automated method for translating photographs of sheet music into machine-readable form. By identifying the notations, reconstructing the musical information, and converting it into an electronically readable format, such as MIDI, this work aims to create a custom music player capable of playing the resulting music.

The playable format can be divided into several modules, including Preprocessing, Frequency Distribution, Detection ofMusical Symbols, and Reconstruction of Notes.

As far as we know, there is currently no system that can automatically recognize a wide range of real-world data accurately and with high usability. However, various methods have been developed to improve specific steps of this linear process. Digitizing a large collection of music scores could provide many benefits and be made easily accessible and searchable. To promote such projects, we propose a novel strategy where the system learns to extract features and suitable rules on its own instead of manually specifying them.

II. MATERIALS METHODS (CONVERT SHEET2MUSIC)

To create an effective Optical Music Recognition (OMR) system, the ability to read and comprehend a music sheet is crucial. This involves subjecting synthetic musical notes to preprocessing techniques, followed by identifying the necessary parameters to achieve accurate music recognition.

A. Input Details: This work indicates that openly available music sheets were utilized for analysis. A sample image of the music sheet that was downloaded is included in the study.

Fig 1: Sample image of music sheet

© SSGMJSE, 2023

B. Nosie Filtering: Noise filtering refers to a set of techniques employed to remove unwanted noise from data captured at building and infrastructure sites. The noise elimination algorithms smooth the entire image, except for areas near contrast boundaries, to reduce or eliminate noise visibility.

C. Segmentation: To partition an image into various regions, each representing a distinct object or part of the image, is known as image segmentation. The purpose of image segmentation is to make an image simpler and more comprehensible for analysis. It is a widely used technique in computer vision, including object detection and tracking, medical imaging, and scene comprehension. The outcome of image segmentation is typically an image with labels or masks that.
indicate each area's uniqueness

Fig 2: Segmentation

B. Staff Line Detection and Removal: The staff, which is composed of five horizontal lines and four spaces, is used to represent musical pitches or percussion instruments. The staff lines play an important role in determining symbol size, scanning quality, and image skew. However, they can also interfere with the recognition of musical symbols on the score. Therefore, it is necessary to remove the staff lines before proceeding to the next step.

Fig 3: Staff Line Removal from the musical sheet of the piano.

C. Symbol Detection and Recognition: The process of symbol identification involves matching each node with a set of around 32 types available in the database.

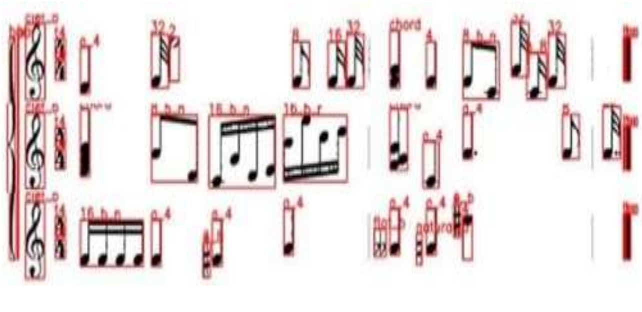

Fig 4: Symbol Detection and Recognition

D. Converting text notes to sound with python library piano

E. Testing Results:

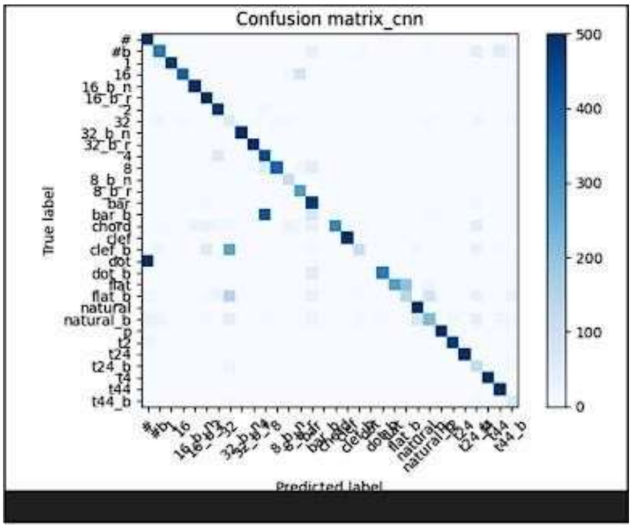

Fig 5: Confusion Matrix

- Classification Accuracy on Test dataset: 0.783699724323178

- Classification Report:

	precision	recall	f1-score	support
#	0.47	1.00	0.64	500
#b	0.82	0.70	0.75	500
1	1.00	0.95	0.97	500
16	0.99	0.81	0.89	500
16_b_n	0.96	1.00	0.98	500
16_b_r	0.83	0.97	0.91	500
2	0.85	0.52	0.91	500
32	0.11	1.00	0.18	500
32_b_n	1.00	1.00	1.00	115
32_b_r	0.47	0.96	1.00	500
4	0.98	0.89	0.62	500
8	0.77	0.78	0.87	500
8_b_n	0.74	1.00	0.87	500
8_b_r	0.63	1.00	0.85	114
bar	0.00	0.96	0.76	286
bar_ b	1.00	0.00	0.00	500
Chord	1.00	0.66	0.79	500
Clef	0.85	1.00	1.00	500
Clef_ b	0.00	0.22	0.35	500
Dot	0.99	0.00	0.00	500
Dot_ b	1.00	0.88	0.93	500
Flat	0.41	0.56	0.72	396
Flat_ b	0.85	0.29	0.34	500
Natural	0.66	1.00	0.92	500
Natural b	1.00	0.44	0.53	500
P	0.96	1.00	1.00	500
T2	1.00	0.95	0.96	500
T24	0.38	1.00	1.00	500
T24_b	0.38	0.79	0.51	155
T4	1.00	1.00	1.00	500
T44	0.37	1.00	0.51	500
T44_b	0.46	0.84	1.00	81
accuracy			0.78	14147
macro avg	0.75	0.79	0.74	14147
weighted avg	0.78	0.78	0.76	14147

© SSGMJSE, 2023

III. MATERIALS METHODS (CONVERTMUSIC2SHEET)

The conversion of sound to music sheets involves a multi-step processing approach. The process includes various modules which are listed below.

A. Frequency Generation:

- Raw Music File: This is the sample music file which is said to be the wav file. This is a table footnote.

- Normalization: Audio normalization, also known as the norm, is the process of applying a constant gain to an audio recording to adjust the amplitude to a desired level. This technique is used to reduce the volume of background noise in the recording.
- Fast Fourier Transform: It breaks down a signal into its distinct spectral components, giving frequency information about the signal in the process.

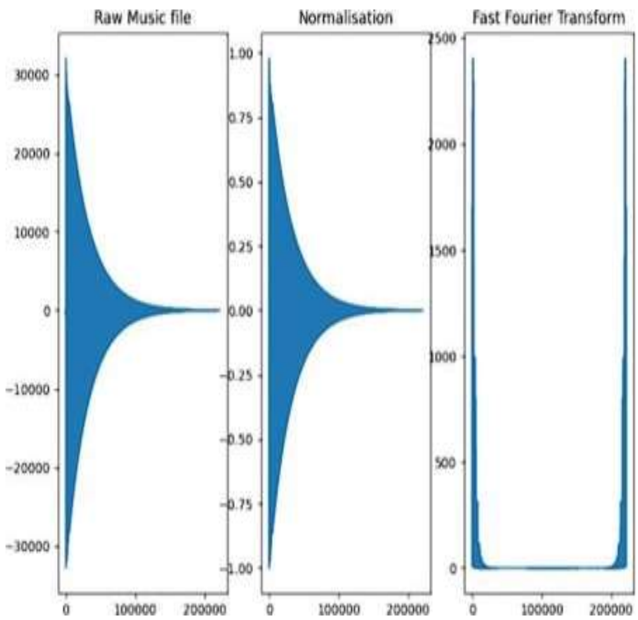

Fig 6: Frequency Detection

B. Creating lookup table and notes detection:

- The creation of a lookup table is a crucial step in the process of converting sound to music sheets. This involves creating frequency ranges for different types of piano notes, which is necessary for comparing the frequencies from the input music file to the frequency database.

  Frequency database (SAMPLES):

Notes	Frequencies
C0	16.35
C#0	17.32
D0	18.35
D#0	19.45
E0	20.60
F0	21.83
F#0	23.12
G0	24.50
G#0	25.96
A0	27.50
A#0	29.14
B0	30.87
C1	32.70
C#1	34.65
D1	36.71
D#1	38.89
E1	41.20
F1	43.55
F#1	46.25
G1	49.00
G#1	51.91
A1	55.00
A#1	61.74
B1	65.41
C2	69.30
C#2	73.42
D2	77.78
D#2	82.41
E2	87.31
F2	92.50
F#2	98.00
G2	103.83
G2#	110.00
A2	116.54
A2#	123.47

- Notes Detection: The subsequent step involves comparing the note frequencies of the sample music with those in the database. The frequencies that closely match are then selected.

IV. CONCLUSION AND FUTURE WORK

In this paper, a visual music recognition technique utilizing a convolutional neural network (CNN) is proposed. The results of the study are promising, indicating that single-staff music notation can be recognized as a sequence. However, more complex musical compositions with multiple voices will require alternative approaches. Future research will explore transfer learning techniques to recognize handwritten music scores.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to our guide, Mrs. Prof Sheetal Jadhav, and our project coordinator, Prof Kiran Bhandari, for giving us the wonderful opportunity to work on the topic of Optical Music Recognition. Their guidance and support were invaluable in our research, and we learned so much from them. We would also like to thank our Head of Department, Mrs. Amarja Adgaonkar, for her direct and indirect help throughout our project work. We extend our appreciation to the I/C Principal, Dr. Vilas N Nitnaware, for providing us with the opportunity to implement our project. Without their support, our project would not have been possible. Finally, we would like to express our gratitude to our parents and friends who helped us immensely in completing this project within the limited timeframe. We are grateful for their unwavering support and encouragement.

REFERENCES

[1] W. Ng and X. T. Nguyen, "Improving Deep-Learning-based Optical Music Recognition for Camera-based Inputs," 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS), Incheon, Korea, Republic of, 2022, pp. 118-121, doi: 10.1109/AICAS54282.2022.9869934.

[2] Calvo-Zaragoza, Jorge, Alexander Pacha and Elona Shatri. “Proceedings of the 4th International Workshop on Reading Music Systems.” ArXiv abs/2301.10062 (2022): n. page. DOI:10.1145/3397499Corpus ID: 199543265

[3] C. Zha et al., "Coordinate Embedding Transformer Model for Optical Music Recognition on Monophonic Scores," 2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Baishan, China, 2022, pp. 354-358, doi: 10.1109/CYBER55403.2022.9907643.

[4] C. Zha et al., "Coordinate Embedding Transformer Model for Optical Music Recognition on Monophonic Scores," 2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Baishan, China, 2022, pp. 354-358, doi: 10.1109/CYBER55403.2022.9907643

[5] F. F. De Vega, J. Alvarado, and J. V. Cortez, "Optical Music recognition and Deep Learning: An application to 4-part harmony," 2022 IEEE Congress on Evolutionary Computation (CEC), Padua, Italy, 2022, pp. 01-07, Doi: 10.1109/CEC55065.2022.9870357.

[6] W. Ng and X. T. Nguyen, "Improving Deep-Learning-based Optical Music Recognition for Camera-based Inputs," 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS), Incheon, Korea, Republic of, 2022, pp. 118-121, doi: 10.1109/AICAS54282.2022.9869934.

[7] F. F. De Vega, J. Alvarado, and J. V. Cortez, "Optical Music recognition and Deep Learning: An application to 4-part harmony," 2022 IEEE Congress on Evolutionary Computation (CEC), Padua, Italy, 2022, pp. 01-07, Doi: 10.1109/CEC55065.2022.9870357.

[8] Calvo-Zaragoza, Jorge, Alexander Pacha and Elona Shatri. “Proceedings of the 4th International Workshop on Reading Music Systems.” ArXiv abs/2301.10062 (2022): n. page. DOI:10.1145/3397499Corpus ID: 199543265

[9] K. Bhumichitr, M. Keo and A. Khant Oo, "Musical Pitch Alphabets Generator Using Haar-like Feature," 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)

[10] L. Tuggener, Y. P. Satyawan, A. Pacha, J. Schmidhuber and T. Stadelmann, "The DeepScoresV2 Dataset and Benchmark for Music

[11] K. Bhumichitr, M. Keo and A. Khant Oo, "Musical Pitch Alphabets Generator Using Haar-like Feature," 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)

[12] L. Tuggener, Y. P. Satyawan, A. Pacha, J. Schmidhuber and T. Stadelmann, "The DeepScoresV2 Dataset and Benchmark for Music Object Detection," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 9188-9195, doi: 10.1109/ICPR48806.2021.9412290.

[13] Calvo-Zaragoza, Jorge, Jan Hajic and Alexander Pacha. “Understanding Optical Music Recognition.” ACM Computing Surveys (CSUR) 53 (2019): 1 - 35.Corpus ID: 219687137

[14] Calvo-Zaragoza, Jorge, Jan Hajic and Alexander Pacha. “Understanding Optical Music Recognition.” ACM Computing Surveys (CSUR) 53 (2019): 1 - 35.Corpus ID: 219687137

[15] D. M. Dhanalakshmy, H. P. Menon and V. Vinaya, "Musical notes to MIDI conversion," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 799-804, DOI: 10.1109/ICACCI.2017.8125940.

© SSGMJSE, 2023