ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 4, April 2022

Piano Music Transcriber

[1] Vedant Dharane*, [2] Sriaansh Sahu, [3] Rithik Kalla, [4] Chetana Badgujar

[1][2][3][4] Information Technology, Fr. Conceicao Rodrigues Institute of Technology, Vashi Navi Mumbai, India
Corresponding Author Email: [1] dharanevedant1000@gmail.com*, [2] sahusriaansh100@gmail.com, [3] rithikkalla@gmail.com, [4] chetana.badgujar@fcrit.ac.in

Abstract*— Sheet music is a transcribed or printed type of melodic documentation that portrays the rhythms, harmonies, pitches of music with the assistance of melodic images. Printed music is the most essential instrument for artists to learn and play music. An educated musician can, from this sheet, play an entire grand composition. However, making sheet music from scratch is a very long and intricate process. A lot of musical experience is required for making sheet music from music one hears and even experts tend to make mistakes. This makes it almost impossible for beginners who are unable to distinguish between different chords to play the music that they hear.

We aim to create a system that would provide the user with the needed sheet music by taking the input from the user in the form of an audio mp3 file. The proposed system automatically generates sheet music directly from recorded songs. This system proposed is mainly composed of estimation of chords then converting it into MIDI file which is a computer understandable language, and eventually into sheet music. All stereo audio recordings are converted into mono which further resampled to 16 kHz. Music generation is also applied which gives a priming sequence to the model in the types of a MIDI file and then predicts the further notes to help musicians form further melody while the composition of music is taking place.

Beginners and lesser experienced musicians also struggle with procuring sheet music for melodies and are limited to only what is available from other professional transcribers. For professional musicians, it helps to lay the foundations and automates a major part of the creation process. Professionals can further improve the automatically generated sheet music without the burden of starting from scratch.

***Index Terms*— transcription, generation, sheet music, MIDI file, notations**

I. INTRODUCTION

Musicians face a very hard time in converting the music played on the instruments into sheet music. In today's world, the music industry is rapidly growing and more new musicians are dependent on sheet music to play compositions. The system proposes to make the tedious task of converting songs played into sheet music very easy and also saves a lot of time and effort that is put in by the artist. Beginners and lesser experienced musicians also struggle with procuring sheet music for melodies and are limited to only what is available from another professional transcriber This gave rise to the need for a system that can generate sheet music automatically, without the need for significant human intervention. It should also be equally accessible to professionals, beginners, and people with disabilities alike. Beginner aspirants who want to learn music from sheet music don't have the proper knowledge for finding one. Moreover, it is very difficult to identify the correct chords when other noises of guitar, singing, drums, and other sorts of disturbances are in the background. Due to the lack of availability of sheet music, many beginner aspirants leave playing the musical instrument altogether. For professional musicians, it helps to lay the foundations and automates a major part of the creation process. Professionals can further improve the automatically generated sheet music without the burden of starting from scratch.

Our system is applicable to-

- A significant amount of music knowledge is not required to use.
- Help keep track of and transcribe improvised music sessions for accuracy and convenience later
- Helps people with hearing disabilities who would normally be unable to transcribe music by ear, get access to relevant sheet music.
- A fine-tuned ear is not required to be able to produce accurate notational transcriptions of music.
- The system relies solely on audio and does not see hand movements or other visual aspects to accurately transcribe music.
- Can be used as an assistive technology to train your ear for music.
- Beginners newly getting into music will not have to worry about sheet music not being available for pieces they wsh to play
- Can be used alongside existing sheet music to compare and also eliminate errors.

II. PROBLEM STATEMENT

The piano is one of the most famous instruments that a musician can play. One of the requirements to play an instrument like the piano is sheet music which is handwritten or printed form of musical notation that enables musicians to read or play music . Transcribing sheet music is a daunting task and one that takes hours and sometimes even days of

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 4, April 2022

listening and refining. It is a difficult task even for professional musicians with years of experience, moreover, it requires a very sensitive and highly trained ear for music. People suffering from hearing disabilities are incapable of producing sheet music themselves from audio. Newcomers often find themselves at a loss when they want to play a melody, they are fond of but have no sheet music available for it. Beginners are almost incapable of creating sheet music in their initial years and need to rely on the work of others. Creating sheet music is a tedious and time-consuming task even for professionals and requires years of developing one's ear for music. This gave rise to the need for a system that could overcome the aforementioned obstacles. The proposed system will be able to automatically generate sheet music for any given melody. It should be easily usable by beginners while also being complex enough for professionals to integrate it into their workflows.

III. RELATED WORK

"Using beat notation for enhancements of chord sheet music document similarity" by Chaisup Wongsaroj, Nakornthip Prompoon, Athasit Surarerks. In this paper, they presented a notation to specify chord lengths in chord sheet music and approach it by calculation at 3 levels: chord, sequence, and music. Based on chords sequences in a sample file, they can also generate similarity values between two music files based on aggregated chord sequence similarity values.[1]

"Deepsheet: A Sheet music generator based on deep learning" by Yu-Lun Hsu, Chi-Po Lin, Bo-Chen Lin, Hsu-Chan Kuo, Wen-Huang Cheng, and Min-Chun Hu. Their research has created a Deepsheet framework that incorporates voice division of the music record, harmony assessment of ambient sound, and music arrangement techniques. The framework proposed by them distinguishes the stops and harmonies from an example music document and subsequently joins every one of them to create printed music.[2]

"Improved chord recognition by combining duration and harmonic language models" by Filip Korzeniowski and Gerhard Widmer. Chord acknowledgment compromises an acoustic the model that predicts chords for every sound outline and a transient model that projects this expectation into marked chord fragments. The proposed model unravels models fleetingly into a symphonious model which is to be applied to chord succession and a chord length model which abuts the chord-level expectations of the language model to the edge level expectations of the acoustic model.[3]

"Music Transformer: Generating music with long term structure" by Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob, Uszkoreit, Noam Shazeer, Ian Simon, Curtis, Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, Douglas Eck. They proposed an algorithm which reduces intermediate memory requirement to linear in the sequence. By using a modified relative attention method, we demonstrate how a transformer can produce a minute-long composition with a cohesive structure, generate continuations that elaborate on a motif coherently, and in a seq2seq setup, generate accompaniments that are conditioned on the melody. We evaluate the Transformer with our relative attention mechanism on two datasets, JSB Chorales and Piano-e-Competition, and obtain state-of-the-art results on the latter.[4]

IV. PROPOSED SYSTEM

It begins with the user's input, the file with the audio. The audio is transformed into a Log Mel spectrum, an acoustic time-frequency representation. The audio-visual model of every audio recording is then generated. From mono stereo, they are resampled to 16 kHz. On the highest note, C8 has a cut-off frequency covering its frequency. On a piano with a 4168kHz frequency, compressed audio snippets are created on audio. The Fourier transform spectrogram is subsequently developed using a Hann window size of 2048. The Log Mel spectrum is calculated on the fly with the help of a collection of TorchLibrosa tools.

Immediately after extracting the Log Mel spectrogram features, normalize the log-melt spectrogram data by adding a batch normalization layer. After that, CRNN-based acoustic models are used to forecast the outcomes of velocity regression, onset regression, framewise classification, and offset regression. The final feature numbers for the four convolutional blocks are 48, 64, 92, and 128 correspondingly. Feature maps are averaged by a factor of two along the frequency axis after each convolutional block to lessen feature map sizes. No pooling is used in combination with the time axis to keep the transcription resolution in the time domain.

Feature maps are aligned with the frequency and channel axes after convolutional layers and placed into a fully connected layer with 768 output units. Two biGRU layers are followed by a fully-connected layer with 88 sigmoid outputs. To prevent overfitting of systems, dropouts with 0.2 and 0.5 are implemented after convolutional blocks and fully linked layers. After merging the velocity and onset regressions, a biGRU layer is created. Following the biGRU layer, the fully-connected layer with 88 sigmoid outputs is applied to predict the regression onsets and frame-wise final production. Likewise, the results of onset regression, offset regression, and frame-wise classification are aggregated and given to the biGRU layer. There are 20,218,778 trainable parameters in the transcribing scheme of note. The note

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 4, April 2022

transcription model was trained using the loss function. The acoustic model design is the same in both the sustain pedal transcription submodule and the note transcription submodule, and there is just one output instead of 88. The loss function is used to train the sustain pedal transcription system.

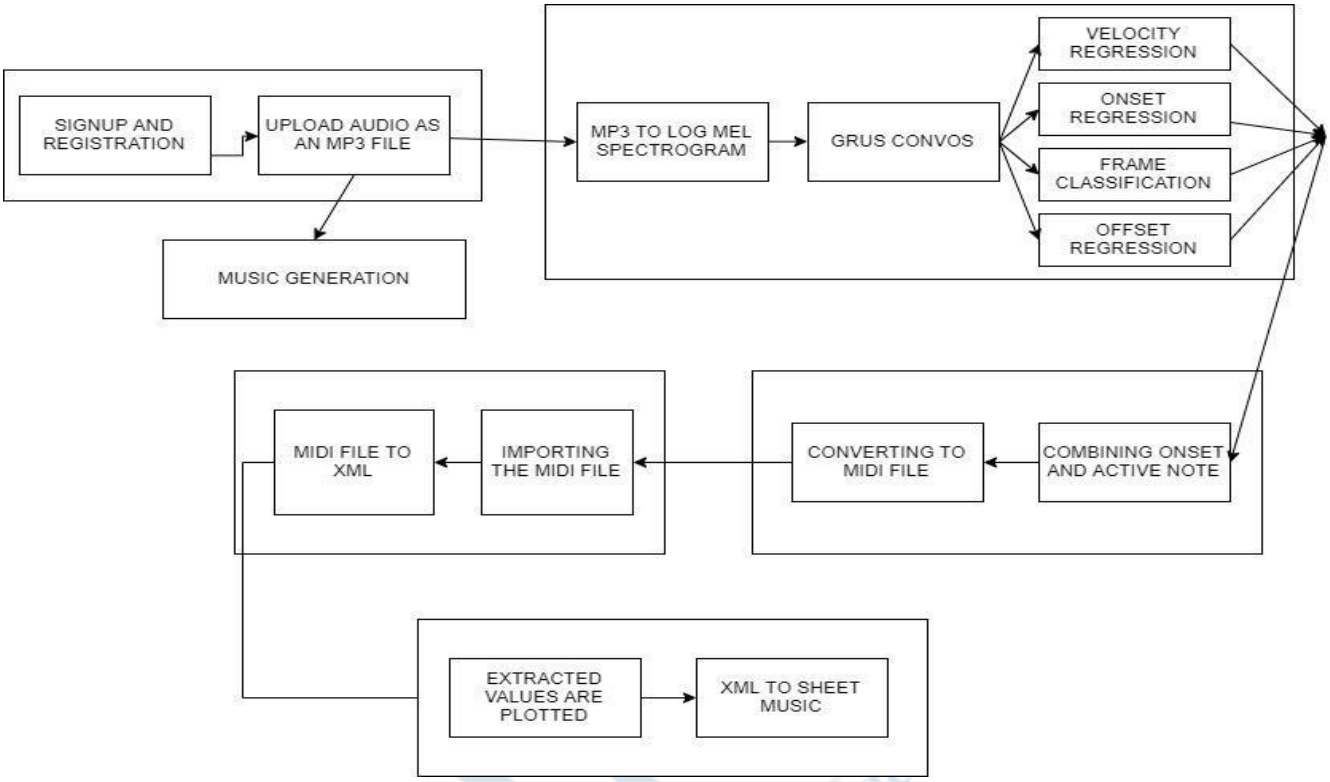

Figure. 1. Architecture Diagram

We all use a set size 12, plus an Adam optimizer using a learning price of 0.0005 for training. The individual learning rate is reduced with the factor of zero. Systems are usually trained for 50k iterations. The learning rate is reduced by a factor of 0.9 every 5 k iterations in training. Systems are trained for 50 k iterations. The movement took one day on a single Tesla GPU card on Google Collab. We all set onset, counter, frame-wise, and your pedal thresholds to zero at inference. 3. All hyper-parameters are tuned upon the validation. The outputs are usually post-processed to MIDI.

The midi files generated are then converted into a music.xml (.xml) file using music21 and then further converted to notes compiled into a pdf file.

A. Music Generation

Pre-trained Transformer models for piano music generation, based on the Music Transformer model introduced by Cheng- Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, Douglas Eck.

The model used here is vigorously trained, transcribed using Onsets and Frames and represented using the event vocabulary from Performance RNN.

Using TensorFlow's Magenta, we can give a priming sequence to the model in the types of a MIDI file, the model then analyses the given primer and generates a continuation sequence.

V. IMPLEMENTATION APPROACH

A. Music Generation

The first model is the music generation in which audio input is given to the system. The pre-trained model decides how many seconds of input needs to be taken from the given file and then predicts the output by running the model. The duration of the output can be decided by the user and the system gives output according to it.

B. MP3 to MIDI

An mp3 audio file is given as an input to the system and then the audio file is split into 10 seconds fragments and then transcribed into the MIDI file. An (.xml) version of the file is also made available in which one can edit and make changes to the existing midi file

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 4, April 2022

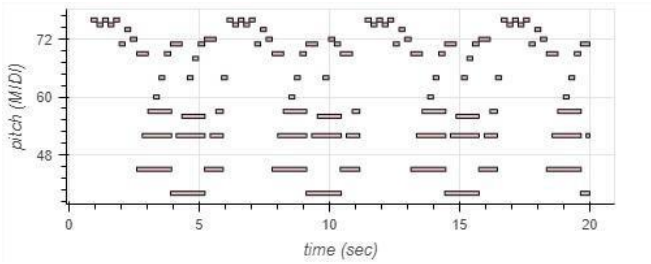

Figure. 2. MIDI file

C. Sheet Music

The final MIDI is then converted into sheet music by parsing the values of the given MIDI file values to their corresponding symbolic notations of the sheet music which is the final output

D. Dataset

MAESTRO dataset V2 is a huge scope dataset containing matched sound recording and MIDI documents to prepare and assess our proposed framework. The dataset comprises 200 hours of independent piano accounts north, resulting in MIDI documents aligned with a 3-ms period goal.

Metadata on every sound recording includes the author, title, and exhibition year. The dataset is then parted into preparing, approving, and testing subsets.

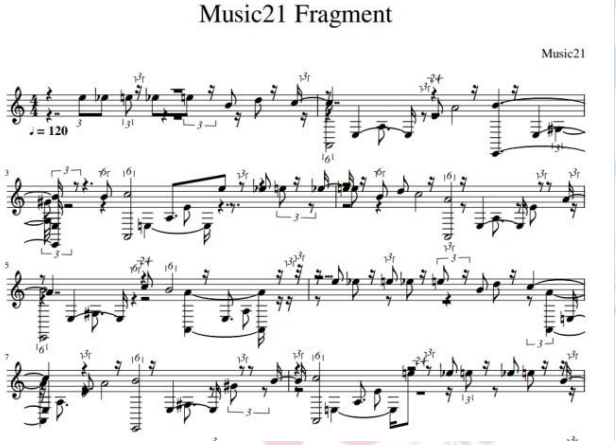

Figure. 3. Sheet Music

VI. CONCLUSION

The proposed system empowers musicians of varied proficiency to generate sheet music of any piano song. It also helps people with hearing disabilities to get relevant sheet music. Transcribing sheet music is a daunting task and one that takes hours and sometimes even days of listening and refining. Due to this the beginner aspirants who want to learn sheet music don't have the proper knowledge for finding one. Due to lack of availability of sheet music, many beginner aspirants leave playing the musical instrument altogether. For professional musicians, it helps to lay the foundations and automates a major part of the creation process. .As it is only using piano for its input ,therefore other music.

Magenta - Magenta is circulated as an open-source Python library fueled by TensorFlow. Among the utilities in this library are controlling source data (mainly music and pictures), using this data to create AI models, and finally creating new content from these models.[5]

TensorFlow - TensorFlow is seen to be one of the most amazing open-source libraries for mathematical calculation. It simply must be acceptable, particularly if goliaths like DeepMind, Uber, Airbnb, or Dropbox have all chosen to use this structure.

PyTorch - PyTorch is primarily used to prepare profound learning models rapidly and viably, so it's the system of decision for an enormous number of scientists. The demonstrating system is basic and straightforward on account of the structure's compositional style.

Music21 - There are a few libraries to control MIDI documents automatically. Music21 is one of them. Citing the definition from their site. Music 21 is a Python-based tool compartment for PC assisted musicology. Individuals use music21 to respond to inquiries from musicology utilizing PCs, to concentrate on huge datasets of music, to create melodic models, to show essentials of music hypothesis, to alter melodic documentation, concentrate on music and the cerebrum, and to make music.

NumPy - NumPy is the principal bundle for logical registering in Python. In this Python library, you will find a multi-layered cluster object, different defined items (like veiled exhibits and networks), and a combination of schedules for the quick procedure on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, etc.[6]

Instruments cannot be taken as an input. Therefore, we built a system which directly converts the audio mp3 played by the user to sheet music for removal of such difficulties faced by the aspirants, thus reducing their time and effort.

VII. FUTURE SCOPE

The proposed system can further be modified by incorporating video file as an input which would further enhance its efficiency and along with it, amalgamation of video and audio system can be incorporated which would benefit the cause of improving efficiency. More musical instruments such as Guitar, Violin, Ukulele, Drums apart from Piano. In a musical medley, multiple instruments are used simultaneously, so this system can select any one instrument's sound and ignore others as noise, this can be further incorporated into the system.

REFERENCES

[1] Chaisup Wongsaroj, Nakornthip Prompoon, Athasit Surarerks, "Using beat notation for enhancements of chord sheet music document similarity", Department of Computer Engineering, Faculty of Engineering Chulalongkorn University, Bangkok, Thailand, 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS).

[2] Yu-Lun Hsu, Chi-Po Lin, Bo-Chen Lin, Hsu-Chan Kuo, Wen-Huang Cheng, and Min-Chun Hu, "Deepsheet: A Sheet music generator based on deep learning", Department of Computer Science and Information Engineering, Institute of Education, National Cheng Kung University, Taiwan Research Center for Information Technology Innovation, Academia Sinica, Taiwan, July 2017, IEEE International Conference on Multimedia and Expo Workshops (ICMEW) 2017.

[3] Filip Korzeniowski and Gerhard Widmer, "Improved chord recognition by combining duration and harmonic language models", Institute of Computational Perception, Johannes Kepler University, Linz, Austria, 19th International Society for Music Information Retrieval Conference, Paris, France, 2018.

[4] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, "Music Transformer: Generating music with long term structure", a conference paper at ICLR 2019.