Original Article

AI Music Generator

Saransh Gupta[1], Sparsh Marwah[2], J. Briskilal[3]

[1]Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Indore, Madhya Pradesh, India.
E-mail: sp1228@srmist,edu.in

[2]Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Shimla, Himachal Pradesh, India.
E-mail: sv5751@srmist.edu.in

[3]Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.
E-mail: briskilj@srmist.edu.in

Abstract

Artificial intelligence is the study and creation of machines that can perform tasks that would ordinarily require human intelligence such as face detection, speech recognition, and decision-making. We will be using deep learning as a source to generate music through a computer using a database of music tunes without any expertise of any trained music artist. The music is generated through sentimental analysis, which helps in mood detection through a sample image of a human self. The music is stored as a piano-roll which is divided in several types of emotions that is generated through the testing process. The use of deep learning is that the machine trains itself through iteration of inputs that is provided in the database which will help in generating music from the input of a face of the user and will make music generation easier. The various models of deep learning are LSTM, CNN, GAN and many more that help in music generation and simultaneously help in enhancing the exposure to sentimental analysis. This project will be using some of these models to demonstrate music generation through sentimental analysis.

Keywords: Deep learning, Sentimental Analysis, LSTM, CNN, GAN.

DOI: 10.47750/pnr.2022.13.S03.012

INTRODUCTION

AI study is known to have had an impact on diagnosis, stock investing, and other areas. Artificial intelligence and music (AIM) has become a typical subject at a variety of conferences and seminars, including the International Joint Conference on computing (IJCAI). Deep learning may also be a machine learning sub-set in AI that has unchecked or unlabeled networks that can learn from results. Deep neural study or deep neural network is also indicated. AI has evolved in full agreement with the fashion boom, culminating inside an explosion of information in all formats from all regions of the earth.Data mining comes from a variety of websites, including social networking sites, internet search engines, e-commerce hubs, and online cinemas, among others.

LITERATURE SURVEY

Long Short-Term Memory was used to process whole data sequences along with single data sets. Each of the units in an LSTM contains three gates and a cell state, allowing it to learn, unlearn, or retain data. The cell state in LSTM permits data to travel across the units without being changed because there are only a few linear interactions. [1]

Recognition of human emotion from a detected human face was undertaken using an artificial neural network in which images of various emotions were combined to achieve more accurate output from these merged images. The Artificial Neural Network is used to examine human emotions in this case. [2]

Convolutional Neural Networks (CNN) are used to detect emotion in order to leverage massive facial detection for sentiment analysis. CNN needs much less well before than most classification methods. CNNs need little or no pre-processing. This ensures that the system learns to refine the filtering or convolution kernels which are hand-engineered in existing methods. This absence of reliance on previous experience or human intervention in extracting features is a substantial advantage. It uses 6 types of emotions such as happy, sad, fearful, disgusted, surprised, angry. This deep learning model can pertain these emotions into different datasets for its classification.[3]

In this paper, a style specific music is generated in which the music notes are refined into more precise dataset. It employs the piano roll description of music, which is a complex

description of MIDI music employed by the Biaxial LSTM architecture. A piano roll is a binary vector that determines the keys played at each time point, with 1 meaning that the note corresponding to its count is played and 0 indicating that it is not. [4]

Facial recognition technology operates by pinpointing and calculating facial expressions from a specific image in order to align a real face derived from a visual image or a video window against a dataset of human faces.[5]

MIDI files are used as raw input to get more information of the data for music generation. In this Convolutional Neural Network and GAN. GANs are a type of generative modelling that uses CNNs, for example, are machine learning systems. Generative modelling is a deep learning problem in machine learning that entails continuously describing and learning sequences in raw data such that the system can generate or output new examples from the test dataset. [6]

In this paper, facial recognition was done through a deep learning model that analyses through various emotions. It uses OpenCV, Haarcascades to help in facial recognition through a real time camera feed which detects different emotions. The following classifiers are used to detect faces:

haarcascade eye

haarcascade frontalface default is the config haarcascade frontalface.[7]

The neural network mode was used to recognise emotions. A RNN is a type of nodes in a neural network form a graph structure which comes after a temporal series. It induces to operate in a temporally challenging way. RNNs, which are developed from feed - forward networks, are responsible for processing variable length sequences of inputs by utilising their internal state (memory). It operates with sequential records / knowledge. It enables the use of a dataset of six different emotions, namely joyful, unhappy, frightened, disgusted, surprised, and furious. [8]

It worked on facial emotion through AI and ML and used a Support Vector Machine (SVM). SVMs was used to solve a classic two-class pattern recognition puzzle. We change the meaning of an SVM classifier's output and devise a description of facial images that is concordant with a two-class problem to fit SVM to face recognition. SVM is a different approach than the other deep learning models that are there.[9]

In this article, a Bidirectional Recurrent Neural Network (BRNN) was used to produce audio. BRNNs were first used to increase the range of input data available to the network. BRNN are especially useful where the context of the input is required. Knowing the letters that come before and after the current letter can boost productivity in handwriting recognition. [10]

Inference from the Survey

Our project uses various deep learning models for music generation through face recognition/Sentimental analysis through emotions. Some of them are LSTM, CNN, GAN and BRNN and many more help in music generation and simultaneously help in enhancing the exposure to sentimental analysis.

A. Recurrent Neural Network (RNN)

A directed graph created by the nodes of a deep neural network that follows a time sequence is known as an RNN. This allows it to act in a temporally complex manner. RNNs, which are descended from feed-forward networks, can track input sequences of variable lengths using their internal statistics (memory). It is a type of neural network that works with sequential data and knowledge. Where a certain collection of features is assigned iteratively to a differential graph-like design, they are referred to as recurrent. Natural Language Acquisition gains immensely from RNN.

B. Long Short-Term Memory (LSTM)

LSTM is a deep learning architecture which provides an RNN architecture. LSTM has feedback relations, unlike traditional feed - forward networks It can process whole data sequences along with single data sets. The cell state in LSTM permits information to travel across the units without being changed because there are only a few basic interactions. Each component has an input, an output, and a gate that adds or subtracts material from the current state of the cell. Using a procedure, the gate evaluates whatever information from the prior cell state should be deleted.

C. Bidirectional Recurrent Neural Network (BRNN)

BRNN connect 2 layers aimed in opposite ways at the same output. The output layer of this method of generative deep learning can accept feedback from both the past (backwards) and potential (forward) states at the same time. BRNNs were initially used to boost the number of input information accessible to the network. BRNN are particularly useful when the context of the input is required. In the field of handwriting recognition, for example, knowing the letters that come before and after the current letter can improve performance.

D. General Adversarial Network (GAN)

GANs, are a form of generative modelling that employs in the machine learning, unmonitored learning tasks such as generative modelling involve automatically exploring and continuing to learn regularities or trends in incoming data can be exploited by the model to produce or output new instances drawn from the training set. By framing the task as supervised learning, GANs are a clever way of teaching a training algorithm.

E. Convolutional Neural Network (CNN)

CNNs are deep neural networks that are used in deep learning to model digital representations. These techniques are also known as shift invariant or space invariant machine learning

techniques since they are based on the cumulative architecture of convolutional layers that test the latent layers and translation separability characteristics. It can also be used for image and video processing, suggestion system schemes, and other purposes.

Proposed Work & Implementation

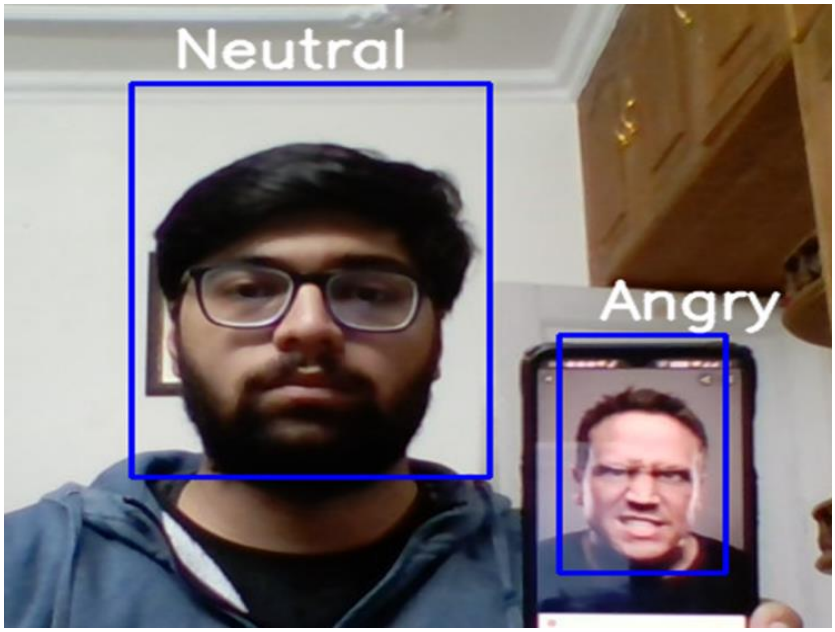
AI is the research and development of computers that can perform tasks that would normally require human intelligence, such as facial detection, voice recognition, and decision-making. Our goal is to use Convolutional Neural Networks to produce music with enhanced facial recognition that senses emotions in real time.

A. Methodology

Our model consists of sentimental analysis through emotions using convolutional neural networks and music generation. This model is basically divided into three parts

These are:.

Part1: First Part deals with the data Acquisition and Data Processing. This part uses Semantic analysis or Facial Recognition. In this part we will store the output and then transfer it to another block of the project where a music will be produced as per this output. In this The system of sampling signals that calculates real-world physical conditions is known as data acquisition. It will be used to determine the user's emotion by translating the emotions into digital numeric that the computer will then control.



i. Input Image for Facial Recognition

Part2: This part deals with the Neural Network Models. In this part we will be using a database having various different music/tones.For each output of the previous part a new MIDI file will be generated. Now as per the facial output the new music will be generated using GAN or CNN model. This database consists of various music tones that are distributed into 7 types of emotions : happy, sad, angry, surprised, fearful, disgusted, neutral. This type of database just increases the precision with the type of emotion it will display during real time emotion capturing as shown in the above picture. Through this the emotion detection for every particular emotion a music tone will be generated.

Part3: This Part deals with the final output. The newly generated music/tone will then be overlapped on the selected music/tune from the database and the final output will be stored in the MIDI file.

B. Convolutional Neural Network (CNN)

In this project, we will use Convolutional Neural Networks, which will assign value to different aspects in an input image, helping us to differentiate one from the other. Convolutional layers are the fundamental components of convolutional neural networks.

Convolution is the simplest way to construct an activation by applying a filter to an input. An object's location and strength of a detected element, such as an image, are indicated by a feature map created by repeatedly applying the same filter to the input. The main motive of using CNN is also that it trains the database under the constraints of specific predictive modeling problems, such as image generator, image classification.

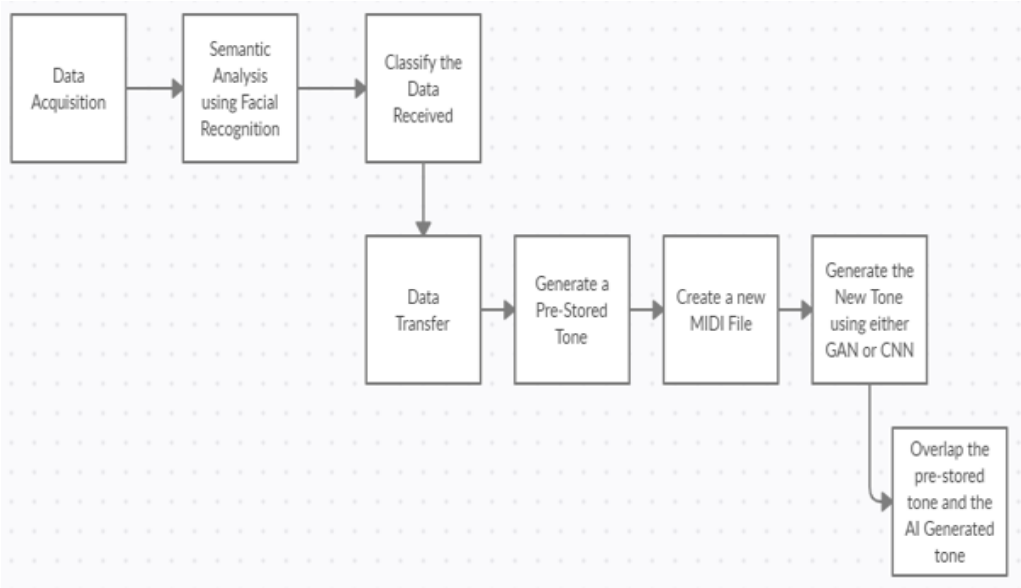


Fig. 1: Architecture Diagram

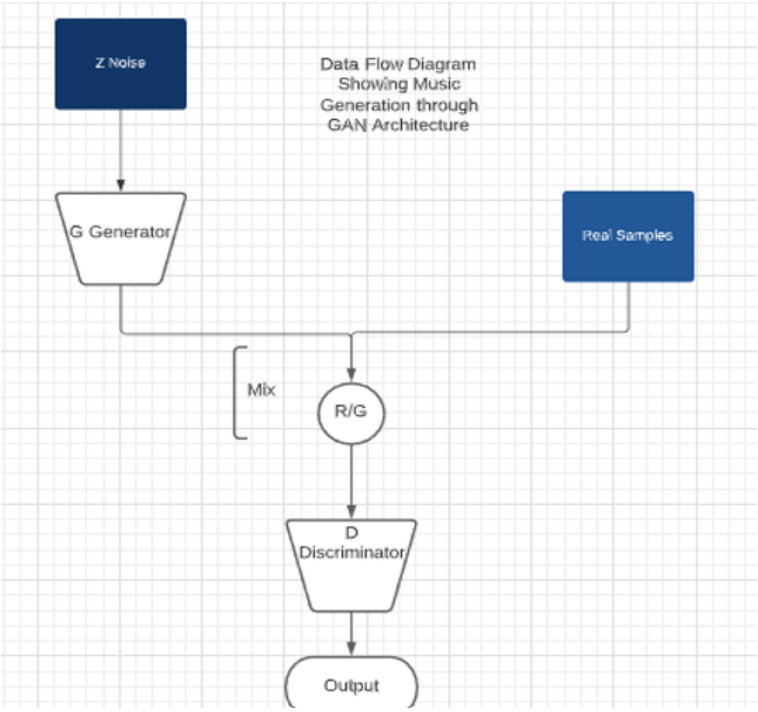


Fig. 2: UML Diagram

C. Dataset

We found data from various mixed resources, the data was in the form of MIDI files which are the music notes that will be distributed among another dataset of emotions consisting seven different emotions such as Happy, sad, angry, neutral, surprised, disgusted, fearful. First the emotions get detected and a screenshot of that emotion is created and then that screenshot is analyzed which gives us a music regarding that emotions in which they are present regarding that emotion. This sort of dataset is only created to get a more accurate result for the output we are expecting. This is possible through the deep learning model that we are using is Convolutional Neural Network.

Results Discussion

The use of CNN led to the generation of music using facial expressions which is used as a dataset in which different music notes are distributed according to the emotions which is giving the music based on the emotion input that has been shown in the real time camera feed.

Conclusion

CNN is used to take an input image, attach value to the image's different aspects/objects and discern one from the other. When opposed to other classification algorithms, CNN has much less pre-processing. Through this deep learning model the music generated is done with facial expressions which is used as a dataset containing innumerable amounts of music notes in every emotion which gives a more precise result.

References

Architecture diagram for AI music Generator:
"https://www.lucidchart.com/pages/"

Boris Knyazev, Roman Shvetsov, Natalia Efremova: "Convolutional recurrent neural network for music classification" 2018 IEEE.

Faizan Ahmad, Aaima Najam and Zeeshan Ahmed: "Face detection & Analysis through various methods" Researcher Gate 2019.

Huanru Henry Mao, Taylor Shin, Garrison Cottrell: "Generation of music through a deep learning model called LSTM" IEEE 2018.

Rishi Madhok, Shivali Goel & Shweta Garg:" Face emotion detection through merged images which helped in getting accurate results of face emotions through artificial neural network" 2018 SCITEPRESS

Tianyu Jiang & Qinyin Xiao: "Bidirectional Recurrent Neural Network as a phenomenon was observed which provided both positive and negative notes of a piece of music" IEEE 2019.

Dr.P Shobha Rani, S.V. Praneeth, Vattigunta Revanth Kumar Reddy, M.J. Sathish: " Music Generation Using Deep learning" IJERT 2020.

Jev, Roman Shvetsov, Natalia Efremova Univ. of Oxford, Oxford, UK. Artem Kuharenko: "Leveraging large face recognition data for emotion classification" IEEE 2018.

Wafa Mellouk & Wahida Handouzi: "Facial emotion recognition using deep learning: review and insights", Science direct 2020.

V.V. Ramalingam, A. Pandian and Lavanya Jayakumar: "Facial Emotion Recognition System – A Machine Learning Approach" IOP Publishing 2018.

Ibrahim, S. (2022). Mathematical Modelling and Computational Analysis of Covid-19 Epidemic in Erbil Kurdistan Using Modified Lagrange Interpolating Polynomial. International Journal of Foundations of Computer Science, 1-17.