JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Facial Expression Recognition Using Convolutional Neural Network

[1]Ch.Sahasra, [2]S.Karthik, [3]E.Deepthi, [4]L.Mankthu

[1,2,3,4] M.Sc (Computer Science), Students

[1,2,3,4] Department of Computer Science,

[1,2,3,4] Chaitanya Deemed to be University, Hanmakonda, India

Abstract: In the modern era, human feeling is extremely important. Human feelings, which can be expressed or not, are the foundation of emotion. Emotion conveys a person's unique behavior, which can take many various shapes. Extraction of emotions reveals a person's unique behavioral state. This study aims to identify emotions and extract facial features from people. Same, to play music in accordance with the emotions found. Modern music services make it simple to access a lot of music. They are continuously trying to enhance music organization and search management, which will address the problem of choice and make discovering new musical works easier. More and more individuals are using recommendation algorithms to choose the right music for any situation. Personalization and recommendations based on emotions are still lacking, though. Humans are greatly influenced by music, which is frequently used to unwind, regulate mood, combat stress and disease, and sustain mental and physical activity. There are many different therapeutic venues and methods used in music therapy to enhance wellbeing. This essay will outline the construction of a system for recommending music based on the feelings, emotions, and activity contexts of the listener. A suggestion system is made to assist people in choosing music for a variety of circumstances and preserve their mental and bodily states by combining artificial intelligence methods and generalized music therapy methodologies.

Index Terms **- Emotion Recognition, Music recommendation, Facial Extraction.**

I. INTRODUCTION

A face detection process involves dividing a photograph into two courses: one with targets (faces), and the other with clutter (background). There are similarities across faces, but they differ in terms of age, skin tone, and facial expression, making this challenging. Different lighting conditions, image qualities, and geometries complicate the issue further. Partial occlusion and disguise are also possibilities. Any face should be detectable by a face detector in just about any background state and under any set of illumination conditions. Two jobs can be separated from the face detection analysis. These machines can be utilised in a variety of settings, including service centres and interactive games. According to Ekman, there are six basic human emotions: fear, disgust, surprise, anger, sadness, and happiness. These expressions can be identified by observing variations in the face. By raising the corners of the mouth and tightening the eyelids, we might say, for instance, that a person is happy. Changes in facial expressions reveal a person's psychological moods, social communication, and intentions. Automatic facial expression identification is widely used in numerous applications, such as human emotion analysis, natural human-computer interaction, picture retrieval, and talking robots. Since humans consider facial expressions to be one of their most natural and effective ways to communicate their intentions and feelings, face recognition using an oriented gradient histogram using CNN recognition has become a significant topic in the technology community. . The system's final step is facial expression recognition. The training process for expression recognition systems mainly consists of three steps: feature learning, classifier development, and feature selection. The first step is the feature learning stage, followed by the feature selection stage and the classifier development stage. After the feature learning stage, only learnt changes in facial expressions between all features are extracted. The best characteristics, as determined through feature extraction, then serve as a representation of facial expression. They should strive to minimize the intra class variances of phrases in addition to maximizing inter class variety. They should maximize expression interclass variation while minimizing expression intraclass variation. Reducing the intra class variance of expressions is difficult because identical expressions of various people in an image are distant from one another in image pixels.

II. LITERATURE SURVEY

The goal of the facial expression-based music player is to scan and evaluate the data before constructing a playlist depending on the given criteria. To develop an emotion-based music player, our proposed system focuses on detecting human emotions. It describes the approaches used by existing mp3 players to detect emotions, the approach our music player adopts to detect human emotions, and how it is preferable to be using our system for sentiment analysis. There is also a quick explanation of how our algorithms operate, playlist creation, and emotion classification. For analysis on this application, we used the PyCharm programme. They first train two Convolutional Neural Networks and use a transfer learning technique to learn how to map human characters to create a common integration. Convolutional layers, namely GoogLeNet, a deep network architecture with 22 layers proposed by

Szegedy et al. The 2014 ImageNet Larger Size Visual Recognition Challenge saw its introduction (ILSVRC 2014). A deep neural architecture that was modelled after Google Net and AlexNet was proposed by Mollahossein et al.

The input facial pictures are categorized using this method into one of six emotions: anger, disgust, sad, happy, surprise, and neutral. A Trawl (Scale Invariant Feature Transform) feature learning approach was proposed by Zhang et al. and is based on an unique deep neural network. A deep network of neurons that was built on two distinct models was proposed by Jung et al. Some other deep network collected time geometry information from temporal facial feature points the first neural net collected temporal beauty features from the photographs. The results of the tests showed that combining these two models improved facial expression detection task performance. A method combining a convolutional network and particular preprocessing steps was put forth by Lopes et al.

The results of the trials showed that combining the normalizing techniques considerably increased accuracy. Majumder et al. presented a FER system called AFERS that included the following four steps: The steps are as follows: (1) separation of geometric features; (2) collection of regional local binary patterns (LBP); (3) merging of both features using auto-encoders; and (4) categorization using a marking on a Kohonen self-organizing map. A large neural network structure with 5 layers and 75K neurons total was proposed by Hamester et al. The suggested network employs Dropout and data enrichment together with Convolution layers, grouping and regional filter, and Highly linked Layers (also known as Thick Layers) to prevent overfitting, which becomes a severe problem as the network size increases. Three steps make up the FER system that Garca-Ramrez et al. proposed: The preprocessing phase is in charge of segmenting the lips and eyebrows. The pre-processing phase is in charge of segmenting the lips and eyelids. Polynomials are used as features in the second stage of the process, feature extraction done.

III. METHODOLOGY

3.1 DATASET:

CK+ Dataset: The Advanced Cohn-Kanade Dataset (CK+), a typical FER reference dataset, includes 230 people ages 18 and 60. The collection includes a wide range of images of people of both sexes of various ages and origins. The images in this collection are 640 x 520 and 740 x 490 pixels in size. The collection consists of 329 sequences from 124 subjects and includes both coloured and monochrome images. Each sequence is assigned to one of the following six facial emotions, which are: disgust, fear, happy, sad, rage, and neutral. As each participant executes the goal face expression changes from the neutral phase, the sequence length can range from ten to sixty frames.

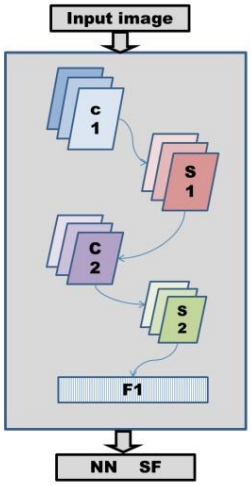
Helen Dataset: About 200 photos from the dataset are utilised to train the classifier. A.txt file containing 175 landmark coordinates is also included with each image in the collection along with the accompanying photos. The system creates the.xml file using these coordinates that are derived from the text files. The classifier can be further trained using this XML file. To ascertain where the landmarks will be located in the second set of unknown photos, a trained classifier is used.

3.2 METHOD:

In investigations on the cat visual cortex in 1958, Laureates in Neurobiology and Medical Hubel and Wiesel initially put out the idea of vision neural cells and made major contributions to the evolution of vision neural networks. On the basis of his forebears, the Japanese scientist Kunihiko Fukushima and Neocognitron created a neural network cognitive model in 1980, which established the groundwork for CNN's later evolution. CNNs were first presented by LeCun and Bengio in 1995, and they were successfully used to recognise handwritten numbers. The CNN can automatically recognise the connection between input and output without the use of artificial adjustment parameters. The processing unit, a fully connected, a subsampling layer, a convolution, and an output layer are the basic building blocks of CNNs. Unprocessed raw picture data can be entered into the input layer, and the forward propagating algorithm and the graded method are primarily responsible for parameterizing the overall network structure. Figure depicts the CNN's organisational structure.

The cnn model, the quantization layer, and the dense layer all process the input image before the feature maps classification layer outputs the classification result. The convolutional layer (C1, C2) and the upsampling layer (S1, S2) typically appear in pairs, as shown in Figure. Convolution and the down sampling layer's two sets of processing elements are the only ones shown in Figure. Additionally, the executing project can be changed in accordance with various graphic data. Original image data from the input is convolved with n convolutional layers. The extracted features are then fed to the bottom level for dimension reduction and further processing after the fourier processing, yielding n feature maps. The output layer will output the classification results after the convnet, downsampling layer, and output layer have all been connected.

CNN's core structural component is a network of artificial neurons that are interconnected, similar to a DNN. The neuron sends the signal to other neurons when it is stimulated. CNN learning can be split into two processes: back propagation and feedforward neural networks. The backpropagation neural network algorithm derives the output data layer by layer from the data of the input layer and determines error by comparing the actual output data to the expected output data. The model's parameters are repeatedly changed rearward based on the obtained mistake in back-propagation technique till the functional form reaches a number within a limit. . On just a de facto standard benchmark dataset, the accuracy of a novel CNN model, which was specifically created to really be able to recognize the face expression for a given photo under the tolerated range to latency for practical uses, was estimated to be as good as 96.3%.



3.3 TECHNOLOGIES

PYTHON:

Python is a high-level interpreted scripting language that was developed by Guido van Rossum at the Netherlands' National Research Institute of Computer Science and Mathematics. After the previous version was leaked someplace at the alt. Source newsgroup in 1991, Edition 1.0 was released in 1994.

After Python 2.0's introduction in 2000, the 2.x series of releases dominated the market until December 2008. Version 3.0, which had a few relatively minor but important modifications that was not completely compatible well with 2.x versions, was then released by the development team. Python 2 and Python 3 are fairly similar to one another, and Python 2 has received back ports of certain Python 3 features. But generally speaking, they continue to be incompatible.

IV. RESULTS analysis and discussion

The application allows users to upload or capture photos, after which it extracts video frames. These frames are stored locally on the computer. Typically, frames are 640x480 in size. This paper suggests a recommender system that uses an image attached to the virtual machine to collect the user's image and extract it. In order to increase the effectiveness of the classifiers that is used to recognize the face present in image, the captured frame of the webcam feed image is then transformed to a grayscale image. When the conversion is finished, the image is transmitted to utilising a learning model, which uses feature extraction methods to pull the face out of the picture feed's frame.

Individual facial traits are extracted once the face has been removed, and they are then submitted to the learned network to identify the emotion the user has exhibited. The HELEN dataset is used to create a classification algorithm that is employed to find or extract face region from the user's face. More than 2000 photos can be found in the HELEN dataset. The above pictures will be utilized to train the classifier so that, when a brand-new, unknowable group of images is given to it, it will be capable of extracting the position of facial parts from those images using the knowledge. It had already learned through the test dataset and come back the exact location of the unique face landmarks that is identified.

The system's main purpose is to improve the user's experience and, as a result, reduce their tension or improve their mood. A best tune matching the user's mood is recognised and played instantly by the audio player, saving user time from searching or looking up songs. An user's image is recorded with the aid of a camera. A suitable song from the person's playlist is then played in accordance with the user's mood or emotions after taking the user's photo. System has proved successful in capturing a user's emotions. Additionally, system was able to obtain the user's updated photos and properly upgrade its classification and trained dataset. The system was created utilizing a facial landmarks approach, and it has been put through numerous scenarios to see what kind of results it will produce. For the majority of the test instances, it can be shown that the classification has an efficiency of more than 90%, which is a respectable level of accuracy for classifying emotions. Additionally, it can be demonstrated that the classifier, when put to the test on a live user, can correctly anticipate the user's expression in a real-time scenario.

Fig 1: User interface

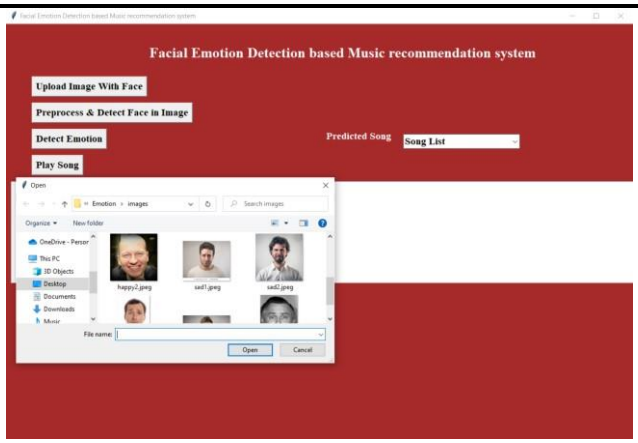


Fig 2: Uploading image

Fig 3:Detected emotion and Song recommanded

V. CONCLUSION PAPER BEFORE STYLING

One of the crucial areas of research is the recognition of emotions from facial expressions, which has previously attracted a lot of interest. It is clear that the difficulty of emotion recognition using algorithms for image processing has been growing daily. By utilizing various feature types and image processing techniques, researchers are actively exploring solutions to this problem. The use of image processing systems to extract the user's emotion and then utilize that feeling to treat the user is constantly leading to the development of new techniques. A strong algorithm that can reliably classify a person's emotions can be created, and this will help the industry advance significantly. Emotion detection has become increasingly important in all facets of life. Users looking for music depending on their emotional state and behavior can benefit greatly from the Emotion Based Music System. It will assist in mini missing the amount of time spent looking for music, consequently reducing the amount of time needed for computation and system's overall accuracy and effectiveness. The device will not only ease bodily tension but will also be beneficial for technologies for music therapy and may also help the music therapist to provide therapy on a client. In addition to its added features listed above, making it a full system for admirers and listeners of music.

.

REFERENCES

[1] H. Gao, W. Huang, and X. Yang, ‘‘Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data,’’ Intell. Automat. Soft Comput., vol. 25, no. 3, pp. 547–559, 2019.

[2] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, ‘‘Research on costdriven services composition in an uncertain environment,’’ J. Internet Technol., vol. 20, no. 3, pp. 755–769, 2019.

[3] H. Gao, Y. Duan, L. Shao, and X. Sun, ‘‘Transformation-based processing of typed resources for multimedia sources in the IoT environment,’’ Wireless Netw., pp. 1–17, Nov. 2019. [Online]. Available: https://doi.org/10.1007/s11276-019-02200-6

[4] H. Gao, Y. Xu, Y. Yin, W. Zhang, R. Li, and X. Wang, ‘‘Contextaware QoS prediction with neural collaborative filtering for Internet-ofThings services,’’ IEEE Internet Things J., early access, Dec. 2, 2019, doi: 10.1109/JIOT.2019.2956827.

[5] X. Ma, H. Gao, H. Xu, M. Bian, ‘‘An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing,’’ EURASIP J. Wireless Commun. Netw., vol. 2019, no. 1, 2019, Art. no. 249. [Online]. Available: https://doi.org/10.1186/s13638-019-1557-3

[6] J. Yu, J. Li, Z. Yu, and Q. Huang, ‘‘Multimodal transformer with multi-view visual representation for image captioning,’’ IEEE Trans. Circuits Syst. Video Technol., early access, Oct. 15, 2019, doi: 10. 1109/TCSVT.2019.2947482.

[7] J. Yu, M. Tan, H. Zhang, D. Tao, and Y. Rui, ‘‘Hierarchical Deep Click Feature Prediction for Fine-grained Image Recognition,’’ IEEE Trans. Pattern Anal. Mach. Intell., early access, Jul. 30, 2019, doi: 10. 1109/TPAMI.2019.2932058