International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 902-907 www.ijaem.net ISSN: 2395-5252

Facial Expression Recognition Using Lstm Framework in Deep Learning

Mrs.M.Shobana,[1] M.E., Assistant Professor, S.Dhanushiya,[2] S.Dharshini,[3] M.Elakkiya,[4] P.Kuralarasi[5],
Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur.

Submitted: 05-05-2021 Revised: 17-05-2021 Accepted: 20-05-2021

ABSTRACT: Facial emotion recognition is the process of detecting human emotions from facial expressions. The human brain recognizes emotions automatically, and software has now been developed that can recognize emotions as well. This technology is becoming more accurate all the time, and will eventually be able to read emotions as well as our brains do. AI can detect emotions by learning what each facial expression means and applying that knowledge to the new information presented to it. Emotional artificial intelligence, or emotion AI, is a technology that is capable of reading, imitating, interpreting, and responding to human facial expressions and emotions. Facial expression is an effective way for humans to communicate since it contains critical and necessary information regarding human affective states. It is a critical part of affective computing systems that aim to recognize and therefore better respond to human emotions. Automatic recognition of facial expressions can be an important component in human-machine interfaces, human emotion analysis, and decision making. As a result, facial expression recognition has become a prominent research topic in human-computer interaction, as well as in the fields of image processing, pattern recognition, machine learning, and human recognition. In this project, we will implement the techniques to automatically detect facial parts using HAAR CASCADES algorithm and classify the emotions using Long Short Term Memory algorithm,. To recognize emotion using the correlation of the facial feature sequence, a deep neural network for emotion recognition based on LSTM is proposed. The first layer of the deep neural network is the LSTM layer, which is used to mine the context correlation in the input facial feature sequence. The second layer is the full-connect layer, which is used to integrate information and act as the major role of the classifier. And present playlist of songs which is suitable for his current mood using K-Nearest Neighbor classification algorithm. In testing side, would supply a test image whose expression it desires to recognize. This test image would be matched with facial databases to play music based on recognized emotions. Finally provide emotion based music player with improved recognition rate.

I. INTRODUCTION

ARTIFICIAL INTELLIGENCE

AI (artificial intelligence) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction. Particular applications of AI include expert systems, speech recognition and machine vision. AI can be categorized in any number of ways, but here are two examples. The first classifies AI systems as either weak AI or strong AI. Weak AI, also known as narrow AI, is an AI system that is designed and trained for a particular task. Virtual personal assistants, such as Apple's Siri, are a form of weak AI. Strong AI, also known as artificial general intelligence, is an AI system with generalized human cognitive abilities so that when presented with an unfamiliar task, it has enough intelligence to find a solution. The Turing Test, developed by mathematician Alan Turing in 1950, is a method used to determine if a computer can actually think like a human, although the method is controversial.

IMAGE PROCESSING:

In imaging science, image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 902-907 www.ijaem.net ISSN: 2395-5252

standard signal-processing techniques to it. Images are also processed as three-dimensional signals with the third-dimension being time or the z-axis. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging. Closely related to image processing are computer graphics and computer vision.

DEEP LEARNING

Deep learning is an artificial intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

- Deep learning is an AI function that mimics the workings of the human brain in processing data for use in detecting objects, recognizing speech, translating languages, and making decisions.
- Deep learning AI is able to learn without human supervision, drawing from data that is both unstructured and unlabeled.
- Deep learning, a form of machine learning, can be used to help detect fraud or money laundering, among other functions.

II. LITERATURE SURVEY

TITLE: MULTI-REGION ENSEMBLE CONVOLUTIONAL NEURAL NETWORK FOR FACIAL EXPRESSION RECOGNITION, AUTHOR: YINGRUO FAN

Facial expression recognition (FER) has many practical applications such as treatment of depression, customer satisfaction measurement, fatigue surveillance and Human Robot Interaction (HRI) system. A well-designed CNN trained on millions of images can parameterize a hierarchy of filters, which capture both low-level generic features and high-level semantic features. Moreover, current Graphics Processing Units (GPUs) expedite the training process of deep neural networks to tackle big-data problems. However, unlike large scale visual object recognition databases such as ImageNet , existing facial expression recognition databases do not have sufficient training data, resulting in overfitting problems. The goal of automatic FER is to classify faces in static images or dynamic image sequences as one of the six basic emotions. However, it is still a challenging problem due to head pose, image resolution, deformations, and illumination variations. This paper is the first attempt to exploit the local characteristics of different parts of the face by constructing different sub-networks. Despite the training size of RAF-DB, it is still insufficient for training a designed deep network. Therefore, we utilize both offline data augmentation and on-the-fly data augmentation techniques. The number of training samples increases fifteen-fold after introducing methods including image rotation, image flips and Gaussian distribution random perturbation. Generally, our method explicitly inherits the advantage of information gathered from multiple local regions from face images, acting as a deep feature ensemble with two single CNN architectures, and hence it naturally improves the final predication accuracy.

DISADVANTAGES

- Need database for facial expression prediction

TITLE: EXPNET: LANDMARK-FREE, DEEP, 3D FACIAL EXPRESSIONS, AUTHOR: FENG-JU CHANG

Today, however, such attempts are all but abandoned. The reason for turning to other face representations may be due to the real-word imaging conditions typically assumed by modern face recognition systems where even state of the art landmark detection accuracy is insufficient to discriminate between individuals based solely on the locations of their detected facial landmarks. In other applications, however, facial landmarks prevail. This work follows recent attempts, most notably by proposing landmark free alternatives for face understanding tasks. This effort is intended to allow for accurate expression estimation on images which defy landmark detection techniques, in similar spirit to the abandonment of landmarks as a means for representing identities. We propose to estimate facial expression coefficients using a CNN applied directly to image intensities. A chief concern when training such deep networks is the availability of labeled training data. For our purposes, training labels are 29D real-valued vectors of expression coefficients. These labels do not have a natural interpretations that can easily be used by human operators to manually collect and label training data. We next explain how 3D shapes and their expressions are represented and how ample data may be collected to effectively train a deep network for our purpose. Another way of addressing the training data problem is by utilizing

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 902-907 www.ijaem.net ISSN: 2395-5252

a face landmark detection benchmark. That is, taking the face images in existing landmark detection benchmarks and computing their expression coefficients using their ground truth landmark annotations in order to obtain 29D ground truth expression labels.Our approach estimates expressions without the use of facial landmarks, suggesting that facial landmark detection methods may be redundant for this task.

DISADVANTAGES

- Complexity is high

III. SYSTEM ANALYSIS

EXISTING SYSTEM

There are numerous areas in human–computer interaction that could effectively use the capability to understand emotion. The problem of face detection can be viewed as a problem of binary classification of image frame as either containing or not containing a face. In order to be able to learn such a classification model, we first need to describe an image in terms of features, which would be good indicators of face presence or absence on a given image. The existing approach is generally involves two tasks: The first is for extracting ASM motion based a pyramid ASM model fitting method and the second for the projected motion classification obtained by applying Adaboost classifiers. After the segmentation of face candidates, 68 feature points in each face are then extracted using ASM fitting technique. The system then line up three extracted feature points, eyes and nose part, to the mean shape of ASM, and ignore the other portion of the ASM against the mean face shape of ASM to estimate the geometrical dislocation information between current and mean ASM points coordinates. Then, facial expressions recognition is the obtained based on this geometrical motion using Adaboost classifier. And also extracting features using viola jones. The features that Viola and Jones used are based on wavelets. Wavelets are single wavelength square waves (one high interval and one low interval). In two dimensions, a square wave is a pair of adjacent rectangles - one light and one dark

DISADVANTAGES OF THE EXISITNG SYSTEM

- Provide large number of features points from facial images
- Emotions may be wrongly classified
- Complexity is high
- Difficult to implement in real time environments

IV. PROPOSED SYSTEM

In this project, a novel emotion recognition system based on the processing of physiological signals is presented. This system shows a recognition ratio much higher than chance probability, when applied to physiological signal databases obtained from tens to hundreds of subjects. The system consists of characteristic face detection, feature extraction and pattern classification stages. Although the face detection and feature extraction stages were designed carefully, there was a large amount of within-class variation of features and overlap among classes. In order to detect Emotion from an image, used frontal view facial images. If computers can understand more of human emotion, we can make better systems to reduce the gap of human computer interaction .To handle the emotion recognition problem from arbitrary view facial images. The facial region and others part of the body have been segmented from the complex environment based on skin color model. Thus, in this project showed some differences between different color models that are used to implement the system and which color model can be used where. Another aspect is to extract facial parts from the face. And for that used HAAR cascade algorithm to detect the eye and lips region from a face and then by the help of LSTM classification detected emotion from those features. From the positioning of mouth and eyes, tried to detect emotion of a face. The proposed system tries to provide an interactive way for the user to carry out the task of creating a playlist. The working is based on KNN algorithm carrying out their function in a pre-defined order to get the desired output. The classified expression acts as an input and is used to select an appropriate playlist from the initially generated playlists and the songs from the playlists are played. At this stage, the face symmetry is measured and the existence of the different facial features is verified for each face candidate. And draw the bounding box and also calculate distance measurement from web cameras.

ADVANTAGES

- User friendly applications and reduce the depression
- No trouble of troublesome selection of songs
- Can be used in real time environments with voice alert
- Reduce number of features are extracted

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 902-907 www.ijaem.net ISSN: 2395-5252

V. SYSTEM IMPLEMENTATION

MODULES

(i) FACIAL IMAGE ACQUISITION:

In this module, we capture the face image or upload the datasets. The uploaded datasets contains 2D face images. In face registration we can identify the faces which are captured by web camera. Then web camera images known as 2D images. Admin can be train the face images with multiple emotions. And also train the music player based on languages.

(ii) PREPROCESSING:

In this module, perform the preprocessing steps such as gray scale conversion, invert, and border analysis, detect edges and region identification. The Grayscale images are also called monochromatic, denoting the presence of only one (mono) color (chrome). The edge detection is used to analyze the connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation.

(iii) FACIAL FEATURES EXTRACTION:

In this module implement HAAR cascades which are an algorithm employed the computer technology that determines the locations and sizes of human faces in arbitrary (digital) images. It detects facial features and ignores anything else, such as buildings, trees and bodies. Face detection can be regarded as a more general case of face localization. In face localization, the task is to find the locations and sizes of a known number of faces (usually one).

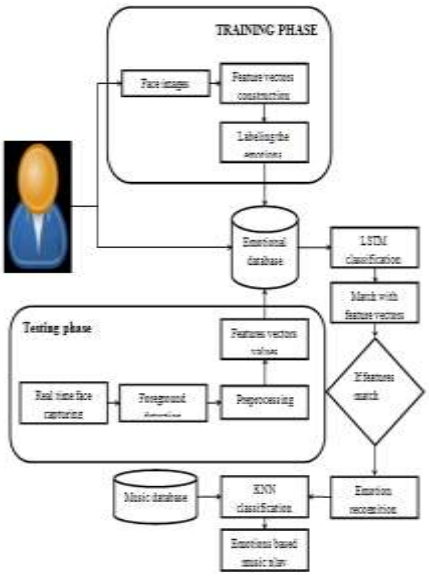
(iv) EMOTION CLASSIFICATION:

In this module analyze on the expression recognition for testing facial images. For a testing facial image, we first extract the facial features and then perform the questionnaire estimation, where LSTM classifier is used for this purpose. After obtaining the question results, we synthesize facial feature vectors based on testing facial feature vector and use them as the model predictors of the positive model. Finally, the model response corresponding to the expression class label vector is calculated and the expression category of the testing facial image can be obtained based on it.

(v) MUSICAL CLASSIFICATION:

If the emotion is positive means, play happiest songs which are stored in database. Using KNN algorithm to classify the music based on emotions classified by previous modules. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms. Both for classification and regression, it can be useful to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbor. The neighbors are taken from a set of objects for which the class (for k-NN classification) or the object property value (for k-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. Based on neighborhood values, music are classified and played in emotional database.

VI. SYSTEM ARCHITECTURE



VII. SOFTWARE DESCRIPTION

FRONT END: .NET FRAMEWORK

The .NET Framework (pronounced dot net) is a software framework developed by Microsoft that runs primarily on Microsoft Windows. It includes a large library and provides language interoperability (each language can use code written in other languages) across several programming languages. Programs written for the .NET Framework execute in a software environment (as contrasted to hardware environment), known as the Common Language Runtime (CLR), an application virtual machine that provides services such as security, memory management, and exception handling. The class

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 902-907 www.ijaem.net ISSN: 2395-5252

library and the CLR together constitute the .NET Framework.

BACK END: SQL SERVER:

Microsoft SQL Server is a relational database management system developed by Microsoft. As a database server, it is a software product with the primary function of storing and retrieving data as requested by other software applications—which may run either on the same computer or on another computer across a network (including the Internet). Microsoft markets at least a dozen different editions of Microsoft SQL Server, aimed at different audiences and for workloads ranging from small single-machine applications to large Internet-facing applications with many concurrent users. Data storage is a database, which is a collection of tables with typed columns. SQL Server supports different data types, including primary types such as Integer, Float, Decimal, Char (including character strings), Varchar (variable length character strings), binary (for unstructured blobs of data), Text (for textual data) among others. The rounding of floats to integers uses either Symmetric Arithmetic Rounding or Symmetric Round Down (fix) depending on arguments: SELECT Round(2.5, 0) gives 3.

VIII. SYSTEM TESTING

Software testing is a method of assessing the functionality of a software program. There are many different types of software testing but the two main categories are dynamic testing and static testing. Dynamic testing is an assessment that is conducted while the program is executed; static testing, on the other hand, is an examination of the program's code and associated documentation. Dynamic and static methods are often used together.

Testing is a set activity that can be planned and conducted systematically. Testing begins at the module level and work towards the integration of entire computers based system. Nothing is complete without testing, as it is vital success of the system.

Testing Objectives:

There are several rules that can serve as testing objectives, they are

1. Testing is a process of executing a program with the intent of finding an error
2. A good test case is one that has high probability of finding an undiscovered error.
3. A successful test is one that uncovers an undiscovered error.

IX. CONCLUSION AND FUTURE ENHANCEMENT

CONCLUSION

In this project proposed convolutional neural network algorithm for emotion recognition. Considering an expressive face as a superposition of a neutral face with expression component, we proposed an algorithm to decompose an expressive test face into its building components. For this purpose, we first generate grids for captured face using HAAR Cascade algorithm. Knowing that the face component of the test face has sparse representation in the face database and the expression part can be sparsely represented using the expression database; we decompose the test face into these feature vectors. The elements of the test face along with the vectors are then used for face and expression recognition. For this purpose, the separated components are sparsely decomposed using vectors while the grouping structures of the vectors are enforced into the sparse decomposition. The experimental results on both databases showed that the proposed method achieves competitive recognition performance compared with the state of the art methods under same experimental settings and same facial feature. Based on their emotions, play the songs to recover from depression. In this project we can be implemented the system to using image processing techniques to detect the faces from camera capturing.

FUTURE ENHANCEMENTS

In future we can extend the system to implement various face detection algorithms to improve the accuracy of the system and implement in different scenarios. We can also implemented in various types of monitors.

REFERENCES

[1]. Fan, Yingruo, Jacqueline CK Lam, and Victor OK Li. "Multi-region ensemble convolutional neural network for facial expression recognition." International Conference on Artificial Neural Networks. Springer, Cham, 2018.

[2]. Chang, Feng-Ju, et al. "Expnet: Landmark-free, deep, 3d facial expressions." 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018.

[3]. Kulkarni, Aishwarya, et al. "Intelligent emotion detection system using facial images." Int. Res. J. Eng. Technol (2015): 2395-0072.

[4]. Tautkute, Ivona, Tomasz Trzcinski, and Adam Bielski. "I know how you feel:

Emotion recognition with facial landmarks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018.

[5]. Mehta, Dhwani, Mohammad Faridul Haque Siddiqui, and Ahmad Y. Javaid. "Facial emotion recognition: A survey and real-world user experiences in mixed reality." Sensors 18.2 (2018): 416.

[6]. Gala, Pratik, et al. "MoodyPlayer: A Music Player Based on Facial Expression Recognition." International Research Journal of Engineering and Technology (IRJET) April (2018).

[7]. Zhang, Zhanpeng, et al. "From facial expression recognition to interpersonal relation prediction." International Journal of Computer Vision 126.5 (2018): 550-569.

[8]. Yang, Biao, et al. "Facial expression recognition using weighted mixture deep neural network based on double-channel facial images." IEEE Access 6 (2017): 4630-4640.

[9]. Vaish, Ashutosh, and Sagar Gupta. "Employing Kaze features for the purpose of emotion recognition." Progress in Computing, Analytics and Networking. Springer, Singapore, 2018. 679-685.

[10]. Ko, Byoung Chul. "A brief review of facial emotion recognition based on visual information." sensors 18.2 (2018): 401.