Face Emotion Detection Using Deep Learning

|  |  |  |
| --- | --- | --- |
| Paras Jain | Dr. M Murali | Amaan Ali |
| Department of Computer Science, | Department of Computer Science, | Department of Computer Science, |
| SRM Institute of Science and | SRM Institute of Science and | SRM Institute of Science and |
| Technology | Technology | Technology |
| Chennai, India | Chennai, India | Chennai, India |

***Abstract*— Facial emotion recognition is an emerging research field in detecting Facial Expression. Deep learning algorithms have gained immense success in different areas of implementation such as classification, recommendation models, object recognition etc. The various types of modules that are brought together in this technique for the betterment of the working of the model is mainly contributed by the progress in the field of Deep Learning. The main focus of this work is to create a Neural Network model which is capable of classifying human emotions in a set of 7 different classes. Image data is used for testing, validation, and training of the model.**

***Keywords—****Convolutional Neural Network, facial expression recognition, deep learning*

INTRODUCTION

Facial emotion recognition is an emerging field in area of computer vision and robotics. Facial expressions are the important identifiers for human emotions, because it corresponds to the human expressions. This problem is becoming a major research area, particularly in effective computing. Verbal and non-verbal information play a major role in obtaining emotions of the human which is mainly collected through diverse sensors such as facial change, tenor of voice and Physiological make over. Many times, the face expression is a nonverbal way of emotional expression, and it can be taken as strong evidence to recognize emotion of a human.

The current approaches mainly focus on face investigation keeping background intact and hence built up a lot of non-essential and misleading features that confound CNN training process. Detection of action unit and detection of facial point are the major first two steps for the proper working of the model. The process begins with the execution of Facial Action Coding System. FACS characterizes facial muscles movement around forty four areas on face.

The given system focuses on five essential facial expression classes reported, which are sad, happy, feared, anger, surprised and neutral. The algorithm presented in this paper aims for examination of various expressions and to distinguish the given image into these five essential emotion classes.

RELATED WORKS

The complication of the following model comes from the inconsistency of expressions made by the face of a normal human, and it can’t easily be modeled by using pattern of face expression. Tian et.al was the first person who made the research in concerned area that projected facial emotion recognition by using FACS. After the research made by Tian et.al, many researchers planned to detect facial emotions by various approaches. FER has become an emerging researched topic in recent years, mainly because of the reason that it has a lot of applications in the fields of robotics, Computer Vision and Human Computer Interaction. Paul Ekman has presented six universal expressions.

Many of the current working FER techniques categorize various facial expressions based on the face features with the help of pattern recognition approach, which mainly uses

geometric-feature based techniques, texture-based approaches, hybrid approaches and appearance-feature based technique. FER systems were mostly implemented using the FACS. However, recently there has been a trend to implement FER using algorithms such as SVM, Fisher face algorithms, and neural networks.

The chief functions of Face Emotion Detection and Analysis are expression classification and extraction of details from face. The result is determined through feature extraction from face which is extremely essential in determining the output. The face emotion recognition consists of two methods: regression and classification methods. In current generation, with the constant implementation of Deep Learning in various areas, including in face emotion recognition. We have included different Deep Learning approaches which has dropout mechanism in order to trim down over fitting.

PROPOSED METHOD

This paper intends to elaborate a method to develop a FER system using CNN. The system will classify the expression of a human face into one of seven expressions – anger, happiness, sadness, surprise, fear, neutral, disgust. The model thus developed can be used to categorize human faces in real time using a webcam. This FER system can be used for analysis of user expressions, to help the system understand human requirements better.

The dataset used for implementation of the FER system was the FER2013 dataset from the Kaggle. The dataset consists of 35,887 labeled images, which are divided into more than three thousand tests and 28709 trained images. The dataset consists of another 3600 private test images, on which the final test was conducted. The images in the FER2013 dataset have size 48x48 resolutions and are black and white images. It has eight classes’ sad, fear, happy, surprise, neutral, anger and disgust dataset is being preprocessed before training phase.

The process of FER is divided into three stages. The preprocessing stage consists of preparing the dataset into a form which implement in such a way to work on a generalized algorithm and generate efficient results. In the face detection stage, the facial features are detected from the images that are captured in real time. The emotion classification stage consists of implementing the CNN algorithm to classify input image into one of its seven classes.

The process is divided into four stages-

1. Input
2. Preprocessing
3. Face Detection
4. Emotion Detection

Preprocessing: - The picture that FER receives consists of noise and has variance in illumination, resolution, and color shade. A range if preprocessing approaches were implemented in the image. The different techniques that were implemented were normalization, gray scaling, and a new size is given to the image.

1. Normalization- In order to get a clear image of the face the technique of normalization is used which is mainly implemented for the removal of illumination.
2. Grayscaling - The process of changing an image from other color spaces is called grayscaling for example red, green, cyan, etc. to shades of gray. It varies between full black and full white. Grayscaling is implemented due to the reason as colorful images are difficult to be analyzed by an algorithm.
3. Resizing - This is mainly done to improve computational speed for using maximum memory as possible. The process involves removing the parts of the image which are not required by the algorithm to implement.

Face Detection: - Face detection is the first step for a FER system Ha-ar cascades is implemented for the detection of the face. The Ha-ar cascades, which are also called as Viola Jones detectors, are classifier which is capable of detecting an object in an image or a video. Training of these components are done on a set of negative and positive face images. Haar cascade are proved as a capable method of object recognition in images and is regarded to provide a very high accuracy. The Haar detect three dark regions on the face. The computer is taught to detect two dark regions on the face, and the location is determined using fast pixel calculation. Haar cascades effectively remove the non

required background data from the image and perceive the facial area from the image. The face detection process using the Haar classifiers were implemented in OpenCV.

Emotion Detection:-Emotion is an intricate outcome with many factors. Therefore, it is hard to label emotions as parameters. The system classifies the image into one of the seven general expressions. The training was done using CNN, which is a class of neural networks proved to be useful in image processing. The dataset was split into training and test datasets, and then trained on the training set. Feature extraction process was not initiated before feeding it into CNN.

1. Splitting of Data:-The dataset was divided into 2 categories in the FER2013 dataset: Training and Public Test . The Training and Public Test set were used for generation of a model.
2. Training and Generation of model:-The neural network architecture have following layers:

Convolution Layer: - In convolution layer, a arbitrarily initialized learnable filter is slid, or convolved above the input. This operation conducts the dot product of the filter and every local region of the input. The output is the three dimensional volume of numerous filters, also known as feature map.

Max Pooling: - The pooling layer is used to lessen the spatial size of an input layer to lower the size of input and the computational cost.

Fully connected layer: - In this layer, output neuron is assembled with every neuron from the prior layer. The size of final output layer is indistinguishable to the amount of classes in which an input image is to be classified.

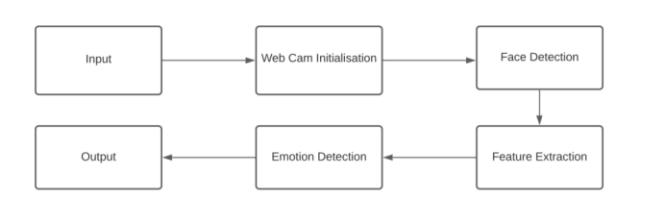
Activation function- The main function of activation function is to reduce the effect of over-fitting. ReLu function is mainly implemented in Classical Neural Network., one of the benefits of this function is that dradent is reduced to the value of one point through which most of the error is passed through back propagation. f(x) = max (0, x) Equation 1: Equation of ReLu Activation Function.

Softmax: - Softmax gives a normalizes the vectorin to various values by taking a vector of n real numbers.

Batch Normalization: The batch normalizer fastens the training process and applies transformations that preserve the standard deviation near to one point and the mean activation near zero point.

Evaluation of model: - In this model, system generated during the training phase was evaluated on the validation set, which consisted of 3589 images. Using model to classify real time images:-The concept of transfer learning can be used to detect emotion in images captured in real time. The model created during the training process consists of pre trained weights and values, which can be used for implementation of a new face expression detection problem.

As the model created already contains weights, FER becomes faster for real time images.



**General Block Diagram**

EXISTING SYSTEM AND PROPOSED SYSTEM

1. All the existing system were on tensorflow 1.0 but now its tensorflow 2.0
2. Face recognition 1.3 library was updated with many new dataset and variable was using updated dlib which led accuracy to 99.99%.
3. Keras library has updated features for expression detection.

METHODOLOGIES AND REQUIREMENTS

1. CNN (Convulational Neural Networks) works on the concept of a Deep Learning algorithm that may take an image as an input, giving preference to diverse components present in the image and have the ability to distinguish one image from the other image

When compared to other algorithms the pre-processing which is needed by the ConvNet is very low. Whereas in primordial strategies filters are mainly hand-engineered, where required amount of training is done, CNN is capable of analyzing those properties.

The pattern in which the neurons in the human brain is connected is very similar to the architecture of ConvNet.The organization of the Virtual Cortex gave a great deal of importance to this idea. Individual neurons respond to stimulus solely in an exceedingly limited region of the visual view referred to as the Receptive Field. A group of such fields overlap to hide the complete visual area.

1. TENSORFLOW 2.0- TensorFlow can be accessed freely by any one as it is an open-source library. It is used for different types of functions however incorporates precise concentration on the working of the model and reasoning of neural networks.
2. PYTHON- It is a high-level structure and used for comman-purpose programming language.
3. KERAS - Keras is an open-source software program library which offers a Python interface for an artificial neural network. Keras incudes one of the most important library that is the TensorFlow library.
4. KAGLE FER 2013- The dataset have 48x48 picture element grayscale images of faces. The faces were habitually registered in order that the face is centered and occupy around the equal quantity of area in every image.

The training set have of 28,709 cases and the public test set consists of 3,589 cases.

.

RESULTS AND OUTPUTS



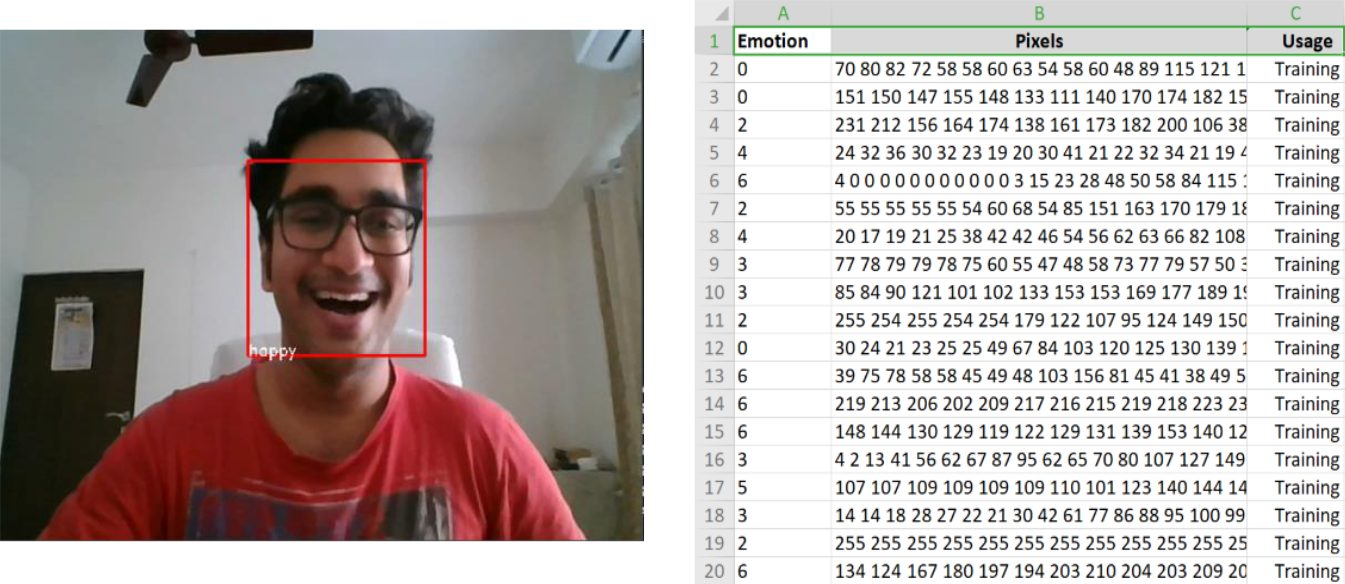
**Before Detection Sample Image 1**



**Before Detection Sample Image 2**

*(This is a sample image we have a tendency to feed to the algorithm and expect our algorithm to observe and determine objects within the image and label them in line with the category appointed to that.)*

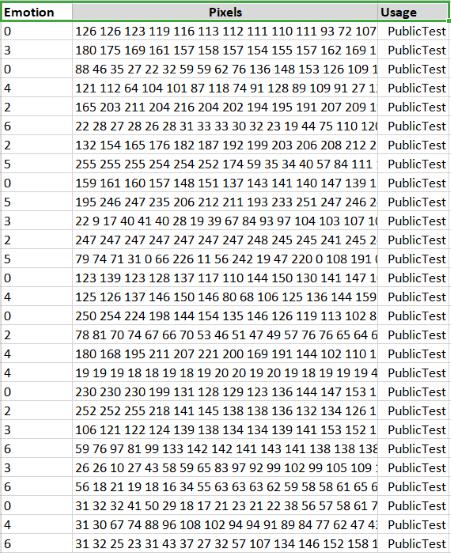
Test and Train labels:



**After Detection Sample Image 1**



**TRAIN (.CSV FILE)**



**After Detection Sample Image 2**

*(As expected our algorithm identifies the objects by its category and assigns each object by its tag.)*

**TEST (.CSV FILE)**

DISCUSSION

We compare the CNN to many high detection frameworks, highlighting key similarities and variations.

**MLP:** Multilayer Perceptron (MLP) is a set of feed forward ANNs. MLP has 3 node levels: hidden level, output level and input level. Each node has a neuron which tends to use a non linear activation function apart from the input node

.MLP uses a supervised learning technique called back propagation learning. Its different layers and their nonlinear activation distinguish MLP from linear perceptrons. You can discover ideas that cannot be shared linearly. MLP is

currently thought of lacking for contemporary advanced pc vision tasks.

Artificial Neural Network: These networks are one of the best varieties of neural networks. They map data in a single direction across a range of input nodes until it reaches the output node. This network may or may not have hidden node layers, which makes its operation much smoother.

CONCLUSION AND FUTURE WORK

In this paper, associate approach for FER using CNN has been discussed. There are seven categories of face we have a tendency to tried to recognize. using the databases we trained using unfamiliar data size and also the result's the mean sq. error decreases because the variety of training data increases.

FER are one amongst the foremost vital ways of giving info regarding the emotional state, however they're perpetually restricted by learning solely the seven-basic emotion. It conflicts by what's there in daily life, wherever the emotions are more advanced. this might push researchers to form larger databases and generate powerful deep learning technique to spot all basic and secondary emotions. these days emotion recognition has modified from uni modal analysis to complex system multimodal.

REFERENCES

1. Y.-L. Tian, T. Kanade, and J. F. Cohn, “Recognizing action units for facial expression analysis,” Proc. IEEE

Conf. Comput. Vis. Pattern Recognit. CVPR 2000 Cat NoPR00662, vol. 1, no. 2, pp. 1–19, 2001.

1. F. De la Torre and J. F. Cohn, “Facial expression analysis,” Vis. Anal. Hum., pp. 377–410, 2011.
2. A. Gudi, H. E. Tasli, T. M. Den Uyl, and A. Maroulis,

“Deep Learning based FACS Action Unit Occurrence and Intensity Estimation,” vol. 2013, 2015.

* 1. Z. Ming et al., “Facial Action Units Intensity Estimation by the Fusion of Features with Multi-kernel Support Vector

Machine To cite this version : Facial Action Units Intensity

Estimation by the Fusion of Features with Multi-kernel

Support Vector Machine,” 2015.

1. R. S. Smith and T. Windeatt, “Facial action unit

recognition using multi-class classification,” Neurocomputing, vol. 150, pp. 440–448, 2015. [6] S.

Taheri, Qiang Qiu, and R. Chellappa, “Structure-preserving sparse decomposition for facial expression analysis.,” IEEE

Trans. Image Process. Publ. IEEE Signal Process. Soc., vol. 23, no. 8, pp. 3590–603, 2014.

1. Y. Wu and Q. Ji, “Discriminative Deep Face Shape Model for Facial Point Detection,” Int. J. Computer. Vision., vol. 113, no. 1, pp. 37–53, 2015.
   1. M. F. Valstar and M. Pantic, “Fully automatic recognition of the temporal phases of facial actions,” IEEE

Trans. Syst. Man Cybern. Part B Cybern., vol. 42, no. 1, pp.

28–43, 2012.

1. L. Wang, R. Li, and K. Wang, “A Novel Automatic Facial Expression Recognition Method Based on AAM,” J.

Comput., vol. 9, no. 3, pp. 608–617, 2014.

[10] P. Ekman and W. Friessen, “Facial action coding system.,” Hum. Face, 2002.

1. M. F. Valstar et al., “FERA 2015 - Second Facial Expression Recognition and Analysis Challenge,” 2015. [12] L. Wang, R. Li, and K. Wang, “A novel automatic facial expression recognition method based on AAM,” J.

Comput., vol. 9, no. 3, pp. 608–617, 2014.

[12]E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 317-320, doi: 10.1109/ICACCS48705.2020.9074302.

[13] K. M. Rajesh and M. Naveenkumar, "A robust method for face recognition and face emotion detection system using support vector machines,"

[14] L. Wang, R. Li, and K. Wang, “A novel automatic facial expression recognition method based on AAM,” J.

Comput., vol. 9, no. 3, pp. 608–617, 2014. [15] V. Sudha, et.al., “A fast and robust emotion recognition system for real-world mobile phone data,” in 2015 IEEE International

Conference on Multimedia Expo Workshops (ICMEW), 2015, pp. 1–6.

1. D. Y. Liliana and C. Basaruddin, “A review on conditional random fields as a sequential classifier in machine learning,” in 2017 International Conference on

Electrical Engineering and Computer Science (ICECOS), 2017, pp. 143–148.

* 1. Z. Ming, A. Bugeau, J.-L. Rouas, and T. Shochi,

“Facial Action Units intensity estimation by the fusion of features with multi-kernel Support Vector Machine,” in

Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, 2015, vol. 6, pp. 1–6.

[18]Y. LeCun, K. Kavukcuoglu, and C. Farabet,

“Convolutional networks and applications in vision,”

ISCAS 2010 - 2010 IEEE Int. Symp. Circuits Syst. Nano-Bio Circuit Fabr. Syst., pp. 253–256, 2010.

* 1. J. Zraqou, W. Alkhadour, and A. Al-Nu’aimi, “An efficient approach for recognizing and tracking spontaneous facial expressions,” 2013 Second Int. Conf. E-Learn. E-Technol. Educ. ICEEE, pp. 304–307, 2013.

1. X. Chen, X. Yang, M. Wang and J. Zou, "Convolution neural network for automatic facial expression recognition," 2017 International Conference on Applied System Innovation (ICASI), Sapporo, 2017, pp. 814-817. doi: 10.1109/ICASI.2017.7988558.