International Journal of Future Generation Communication and Networking
Vol. 13, No. 2s, (2020), pp. 134–141

Emotion Recognition Using CNN

Pushkata Petkar[1], K. A. Pujari[2], Srushti Salgare[3], Sumukh Shetty[4]

[1,2,3,4] *Dept. of E & TC Engg., Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune*

[1]*pushkatap@gmail.com*
[2]*kapujari.skncoe@sinhgad.edu*
[3]*srushtisalgare4@gmail.com*
[4]*sumukhshetty123@gmail.com*

Abstract

Humans have the capacity to produce a varied range of facial actions during communication that vary in intensity, complexity and meaning. Machines can be trained to understand human emotions by using processed images converted into data as an input to a neural network. The system will be using cameras to acquire images. The images will be further processed and filtered to prepare the images for data extraction. This data is then provided to the neural network for training and testing purpose. After training our system to recognise emotions, it will be implemented to increase security by recognizing mass fear and notifying the authorities.

Keywords- Emotion recognition, Classification, Neural Network, face recognition, Security system.

I. INTRODUCTION

Metro's are considered to be the key solution to the traffic problem. But even with these provisions the metro's are still prone to accidents and attacks. By making use of the image processing and machine learning we can make our metro safer. This can be done by detecting facial emotion and notifying the authorities. This system will be able to deal with the most unprecedented situations.

In recent last twenty years, human-computer interactions (HCI) filed has been progressed and play an important role in developing computer science by creating wide variety practical application so that incorporate human beings behavior with computer devices. In the robotics research, particularly humanoid robots, there is interesting to apply emotion recognition on the machine to allow a way of communication naturally . These is also interest to use emotion in the HCI to carry out an efficient and intelligent interaction or communication between human and machine like human beings.

Automatic emotion expression recognition include three steps: face image acquision, feature extraction, and facial emotion expression recognition. Feature extraction for emotion recognition can be divided into two approaches: Geometric feature-based method and appearance-based method .In the first method, location and shape of parts of the face such as eyes, months, eyebrow and noise are considered, while in the second method, particular region or whole of face are considered.

ISSN: 2233-7857 IJFGCN

II. LITERATURE SURVEY

This study in paper [1], "Emotion Recognition Based On CNN" attempts to use the EEG signal from the DEAP data set to classify the emotion of the subjects, this data set represents the emotional classification research. Then the principal component analysis is used to reduce the dimension of the preprocessed EEG data, so the main emotional EEG features are obtained. Then the accuracy of the classification of the training samples and the test samples is tested by the CNN algorithm, and the other classification methods are compared to obtain the nerves. The network can be used as a robust classifier for brain signals even better than traditional learning techniques.

The model shown in paper [2], "Human Emotion Recognition using Convolutional neural network in real time " have developed a smart vision for a device that can detect human face and recognize its emotion, in a single integrated module. We have proposed shallow and deep network for identifying human facial emotions accurately. The shallow model consists of single convolution layer attaining accuracy up to 47.94%. This model uses recently introduced swish activation function in the fully connected layer which makes it more unique and efficient in terms of performance.

The paper[3], "Overview on Emotion Recognition System" describes the advances made in the field of emotion recognition and the various approaches used for recognition of emotions.. The main objective of the paper was to analyse the strengths and weaknesses of various techniques adopted in the emotion recognition system. The inter personal communication between a human being and a computer can be increased rapidly by combining all the pros. Most of the real time issues can be improved by using this recognition system.

The system described in paper[4], "Emotion Recognition from 2D Facial Expressions" has developed a paradigm for facial expression recognition using learned feature approach. It consists of a convolutional neural network model that is deep enough to learn features from the facial images. To avoid overfitting the developed CNN model, they have included several techniques such as dropout, batch normalization, and data augmentation. These additions prevent the developed CNN model from overfitting and at the same time decreases the number of parameters in the architecture which reduces the memory usage

In the study shown in paper[5], "Study of CNN in the recognition of image and audio" the performance of Convolution Neural Network (CNN) in emotion recognition in speech and emotion and image recognition was compared and presented. The comparison was carried out between CNN and SVM baseline system. For image recognition, recognition results on CNN are better than them on SVM baseline system with better robustness. For image recognition, recognition results on CNN are better than them on SVM baseline system with better robustness.

In Paper[6], "A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks" it is said that Automated Face Expression Recognition (FER) is still continuing to be a challenging and concerning problem in Computer Vision. In spite of all the efforts being made in the evolution of various methods for FER, the present methods lack popularity when it comes to unseen images or pictures captured in wild settings. This paper makes an attempt to design an artificially intelligent system capable of emotion recognition through facial expressions of unknown people. The network in this paper consists of three convolutional layers each followed by max pooling and ReLU. The network was trained on FER2013 dataset and was tested on RaFD dataset thus giving a wide range of training images to the network, so that it can overcome the basic problem of recognition of unknown faces. The pertinence of the final model is depicted in a live video application that can instantaneously return users emotions based on their facial posture. The accuracy obtained by this method was 68%, which is better than the previous state-of-the-arts methods. The results provide an important insight on the significance of using different datasets for training and validation..

ISSN: 2233-7857 IJFGCN

In research paper[7], they developed a face recognition system “Design of Face Detection and Recognition System for Smart Home Security Application” and applied for household security. The design was implemented using MyRIO 1900 and programmed using LabVIEW. The connection between myRIO and computer was a wifi network. The image of a person is acquired through a webcam connected to MyRIO using a USB cable. The face recognition system was based on Principle Component Analysis, while the face detection system is built based on the template matching. The testing is done to examine the performance of the face detection in case of changes in distance, light intensity, light position angles, person’s accessories and shirt colour. The face detection module has good performance in some conditions where the distance between the person and the camera is less than 240 cm, the person doesn’t use any accessory to cover any part of the face, person doesn’t use shirt with colour similar to skin colour, and background colour is difference from skin colour. While the face recognition system has 80% of accuracy when it was tested using realtime images.

In the paper[8], “Multiple face detection based on machine learning” they have used methods based on machine learning that allows a machine to evolve byb means of a learning process, and to undertake and perform tasks that are difficult or impossible to fill by more conventional algorithmic means. According to this context, they have established a comparative study between four methods (Haar-AdaBoost, LBP-AdaBoost, GF-SVM, GFNN). These techniques differ according to the way in which they extract data and the adopt learning algorithms. The initial two methods "Haar-AdaBoost, LBP-AdaBoost" are primarily based on the Boosting algorithm, which is utilised both for selection and for learning a strong classifier with a cascade classification. The last two methods used for classification were "GF-SVM, GF-NN" which used the Gabor filter to extract the characteristics. They found that the detection time varies from one method to another. The study seems to indicate that LBP-AdaBoost and Haar-AdaBoost methods are faster compared to others. But in terms of detection rate and false detection rate, the HaarAdaBoost method has a lead on the other four.

In the paper[9], “Real time multiple face recognition using deep learning on embedded gpu system” the fact that recognizing multiple faces in real time on the embedded system is very challenging due to the complexity computation of the processing is taken into account. In the paper, they have proposed a framework for multiple face recognition which is implemented on the embedded GPU system. The framework contains face detection based on convolutional neural network (CNN) with face tracking and state of the art deep CNN face recognition algorithm. They implemented the proposed framework into the embedded GPU system, i.e., NVIDIA Jetson TX2 board. Results indicated that their system can recognize multiple faces up to 8 faces at the same time in real time with a processing time up to 0.23 seconds and with the minimum recognition rate above 83.67%.

The model described in paper[10], “Facial Expression Recognition Combined with Robust Face Detection in A Convolutional Neural Network” was developed considering that for reliable detection of ordinary facial expressions (e.g., smile) considering the variability among individuals as well as facial appearances is an important step toward the realization of perceptual user interface and building the next generation imaging system with autonomous perception of humans. They have attempted to design a robust facial expression recognition system using the result of face detection by a convolutional neural network and rule-based processing. In this study, they have attempted to resolve the problem of subject independence as well as translation, rotation, and scale invariance in the recognition of facial expression. Their result promises reliable detection of smiles with recognition rate of up to 97.6% for 5600 still images. The proposed algorithm demonstrated the ability to differentiate smiling from talking based on the saliency score in the proposed algorithm..

In this article presented in paper[11], “Convolutional Neural Network (CNN) for Image Detection and Recognition” they have attempted to build CNN models to analyze its performance on image recognition and detection datasets. The algorithm is implemented on MNIST and CIFAR10 dataset

ISSN: 2233-7857 IJFGCN

and its performance are evaluated. The accuracy of models on MNIST is 99.6 %, CIFAR-10 is using real-time data augmentation and dropout on CPU unit. The article discusses various aspects of deep learning, CNN in particular and performs image recognition and detection on MNIST and CIFAR -10 datasets using CPU unit only. The accuracy of MNIST is good but the accuracy of CIFAR-10 can be improved by training with larger epochs and on a GPU unit. The calculated accuracy on MNIST is 99.6% and on CIFAR-10 is 80.17%. The

The proposed face recognition system[12], "Face Recognition and Detection using Neural Networks" carries out face verification and face recognition task. In verification task, the system known a priori the identity of the user and has to check this identity i.e. the system has to decide whether the a priori user is an imposter or not. They have made use of MATLAB 2013 and neural network for face recognition. Training accuracy on CIFAR-10 is 76.57 percent after 50 epochs. There is a further chance of improving the accuracy further by introducing more hidden layers. And this system can be implemented as a assistance system for machine vision for detecting nature language symbols.

The paper[13], "Face Recognition Using Cloud Hopfield Neural network" presents an approach of using cloud Hopfield neural network (CHNN) for the recognition of low resolution grey scale facial images. Here firstly, we transform the grey scale facial images into binary facial images using Otsu's method, then, to store binary faces in the weight matrix of the network the Hebb rule is employed and finally using CHNN retrieval algorithm correct face is retrieved from distorted face. Their results show that even when the distortion in the presented face is 45%, the CHNN is able to give at least 82.8% successful retrieval as opposed to only 63% by conventional HNN for the same amount of distortion. This is much greater than previous reported claims to the best of our knowledge.

In this paper[14], "Real-Time Implementation Of Face Recognition System" the major task of the research is to develop face recognition system with improved accuracy and improved recognition time of a face recognition system. This paper proposes a hybrid face recognition algorithm by combining two face recognition techniques by integrating (PCA) principle Component Analysis, (LDA) Linear Discriminant Analysis. Jacobi method is used to compute Eigenvector that are necessary for PCA and LDA algorithms. Face Recognition system will be implemented on Embedded system based Raspberry pi 3 board. In this paper, he has proposed an efficient Face recognition system based on PCA and LDA. Using these two combination of methods have given me accuracy of 97% by using raspberry pi 3 module. This project on Face Recognition has given me an opportunity to study many face recognition algorithms that were used and being currently used. This project has also provided with the knowledge that combining two or more methods increase the accuracy of Face recognition system. In future this Face recognition system could be incorporated on a Robot to make it more Human like.

III. PROPOSED SYSTEM

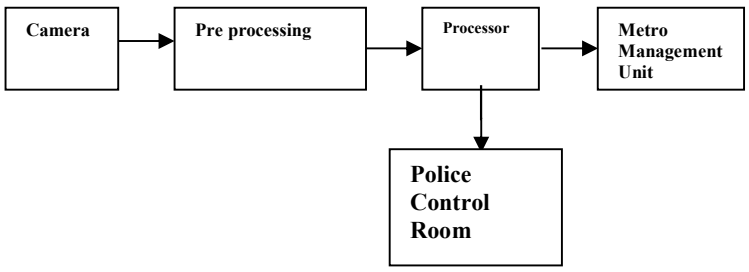

Figure 1. Block Diagram of Proposed System

A. Software Description

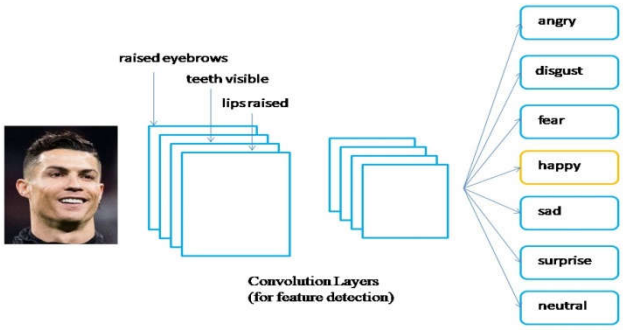

Figure 2. CNN representation

CNN comes back to 1998 which has been shown is very effective for learning feature and modeling high level of abstraction . CNN includes six components: Convolutional layer, Sub-sampling layers, Rectified linear unit (ReLU), Fully connected layer, Output layer and Softmax layer

1) *Convolution Layer:* Convolutional layers are determined by number of generated maps and kernel's size. The kernel is moved over the valid area of the given image (perform a convolution) for generating the map. If f_k be a filter with a kernel size n × m and is supposed io applied into the given image x, output of the layers can be calculated as follows:

$$C(x_{u,v}) = \sum_{i=-\frac{n}{2}}^{\frac{n}{2}} \sum_{i=-\frac{m}{2}}^{\frac{m}{2}} f_k(i,j) x_{u-i,v-j}$$

Where each CNN neuron has n × m number of input connections.

2) *Sub-sampling layers:* the map size of previous layer in order to increase the invariance of the kernels. Sub-sampling includes two types of average pooling and maximum-pooling . By applying maximum function in the Max-Pooling, input value is reduced at the x_i. If m be the size of the kernel, output of max-pooling can be calculated as follows:

3) *Rectified linear unit:* A rectified linear unit is a activation function which it simply thresholded at zero and can be calculated as follows:

$$R(x) = max\ (0, x)$$

ReLU has advantages over tanh/sigmoid function in which it can be implemented by simple thresholding at zero, while in tanh/sigmoid there are expensive operations like exponentials. ReLU is also prevents loosing gradient error, and extremely accelebrate the stochastic gradient descent convergence compared with the lanh/sigmoid functions.

4) *Fully connected layer:* Fully connected layers are similar to neurons in general neural networks which its neurons are fully connected with every neurons in the prior layer. In the be input with size k and the number of neurons representedby l in the fully connected layer, the layer can be calculated as follows:

$$F(x) = \sigma(W * x)$$

Where σ is activation function.

5) *Output Layer*: The output layer represent class of the input image which its size equal to number of classes. Output vector x produce resulting class as follows:

$$C(x) = \{i \mid \exists i \, \forall j \neq i : x_j \leq x_i\}$$

6) *Softmax layer:* The error of the network is propagated back through a softmax layer. If *N* be the size of the input
vector, a mapping can be calculated by softmax such that:
$S(x): R \rightarrow [0,1]^N$.

7) Features of CNNs

i. i) CNNs build their own features from raw signal. On the other hand algorithms that use vector representations where every component tends to make some sense on its own. Outside the context pixels don't make any sense, but when together they may hold more information about the objects on the picture than a bunch of its attributes/features that you feed into SVM.

ii) CNNs have the ability to use infinitely strong priors. This property of max pooling layers is its primary property. Invariance to local fluctuations and good generalization makes them super scalable.

iii) CNNs rely on spatial features. The context of the features can be local or it can be distributed, the former works in favor of the CNN and makes it stronger while the later makes it weaker (a sum of several one-hot encodings is easy for a decision tree to handle but more difficult to handle with CNNs).

iv) The number of parameters stored by CNN is much more compared to other methods. Trees in RFs can only tell you something about the feature importance if you have interpretable features, but that would be it. To use the enormous amounts of parameters convents have a better way out.

IV IMPLEMENTATION

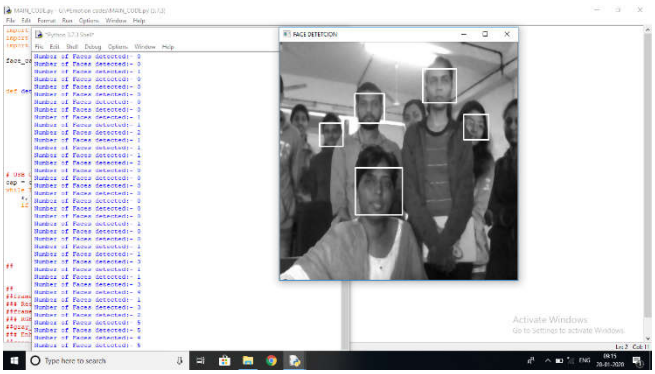

Figure 3. Mutliple Faces detection

In the implementing phase , RGB to Grey conversion is done for image, the image is then passed through the Gabor filter, the features of the image were then extracted using haar cascade feature extraction process. In this phase, attempt was done to recognize multiple images in the same frame.

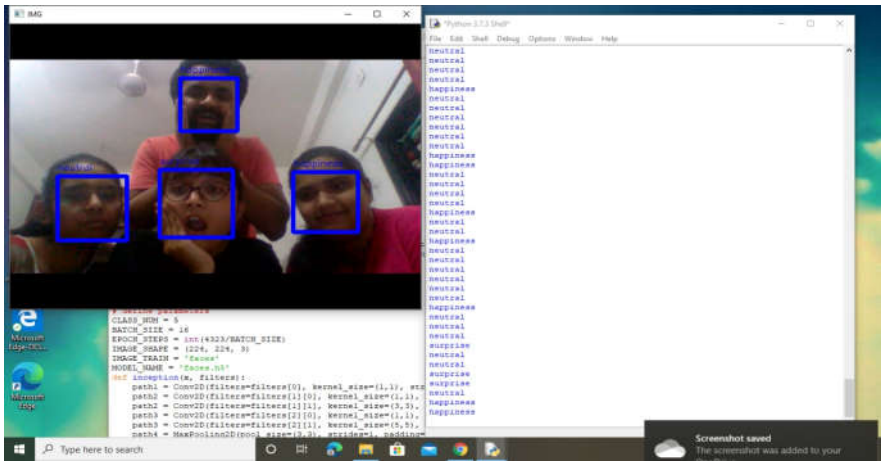

Figure 4. Emotion recognition

In the second phase, recognition of the emotions of the faces detected in the earlier phase is done with the help of Convolutional Neural Network(CNN).

IV. CONCLUSION

The results obtained are very promising, fast and very much applicable for practical purposes. The system is being fed with real time input through a laptop camera. The software continuously scans for Faces after every specified interval and scans for recognizable faces in the image. It is reliably detecting the faces and recognizing the emotions of the recognized faces The emotion of the detected faces are also successfully labelled as anger, sadness, happiness, scared, surprised or neutral in case the face shows no emotions. The system accuracy often declines if the light availability is not enough to illuminate the features of the face. The distance of the faces from the camera also affects the results proportionally with the closest faces giving better results.

REFERENCES

[1] Guolu Cao , Yuliang Ma , Xiaofei Meng, Yunyuan Gao and Ming Meng, “Emotion Recognitio Based On CNN”, Chinese Control Conference 2019 (CCC).

[2] Rohit Pathar, Abhishek Adivarekar, Arti Mishra, Anushree Deshmukh, “Human Emotion Recognition using Convolutional neural network in real time ”1st International Conference on Innovations in Information and Communication Technology 2019 (ICIICT).

[3] Ashwini Ann Varghese, Jacob P Cherian, Jubilant and J Kizhakkethottam, “Overview on Emotion Recognition System”, International Conference on Soft-Computing and Networks Security 2015 (ICSNS).

[4] Bilal Taha, Dimitrios Hatzinakos,“Emotion Recognition from 2D Facial Expressions”, IEEE Canadian Conference of Electrical and Computer Engineering 2019 (CCECE).

[5] Bin Zhang, Changqin Quan, Fuji Ren, “Study of CNN in the recognition of image and audio”, IEEE/ACIS 15th International Conference on Computer and Information Science 2016 (ICIS).

[6] Vibha. V. Salunke, C.G. Patil, “A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks”, International Conference on Computing, Communication, Control and Automation 2017(ICCUBEA).

[7] Dwi Ana Ratna Wati, Dika Abadianto, *"Design of Face Detection and Recognition System for Smart Home Security Application",* 2nd International conferences on Information Technology, Information Systems and Electrical Engineering, 2017 (ICITISEE).

[8] Hajar Filali, Jamal Riffi, Adnane Mohamed Mahraz, Hamid Tairi *"Multiple face detection based on machine learning",* International Conference on Intelligent Systems and Computer Vision ,2018 (ISCV).

[9] Savath Saypadith, Supavadee Aramvith "*Real time multiple face recognition using deep learning on embedded gpu system"* Asia-Pacific Signal and Information Processing Association Annual Summit and Conference , 2018 (APSIPA ASC).

[10] M. Matsugu, K. Mori, Y. Mitari, Y. Keneda, *"Facial Expression Recognition Combined with Robust Face Detection in A Convolutional Neural Network".* Proceedings of the International Joint Conference on Neural Networks, 2003.

[11] Rahul Chauhan, Kamal Kumar Ghanshala & R.C Joshi "*Convolutional Neural Network (CNN) for Image Detection and Recognition",* First International Conference on Secure Cyber Computing and Communication , 2018 (ICSCCC).

Vinita Bhandiwad, Bhanu Tekwani *"Face Recognition and Detection using Neural Networks",* International on Trends in Electronics and Informatics , 2017 ICEI).

[12] Neha Soni ; Narotam Singh ; Amita Kapoor , Enakshi Khular Sharma, *"Face Recognition Using Cloud Hopfield Neural network",* International Conference on Wireless Communications, Signal Processing and Networking, 2016 (WiSPNET).

[13] Neel Ramakant Borkar, Sonia Kuwelakar, "*Real-Time Implementation Of Face Recognition System".* International Conference on Computing Methodologies and Communication, 2017 (ICCMC).