Facial Emotion Recognition in Video using CNN

[1]AartiRamdasAakulwad,[2]Dr.V.E.Chirchi
M.Tech Student,Professor
Computer Science Department, M.B.E. Society's College of Engineering, Ambajogai, India.

Abstract : The emotions of persons used in the non-verbal communication process. Sentiments are the expressions, emotions, opinions, nature which like or dislike. For successful facial recognition system needs to design robust features and more accuracy. A lot of research work has been done for facial recognition. Facial expression presents key mechanism to describe human emotion. From starting to end of the day human changes plenty of emotions, it may be because of their mental or physical circumstances. Although humans are filled with various emotions, modern psychology defines six basic facial expressions: Happiness, Sadness, Surprise, Fear, Disgust and Anger as universal emotions. A facial muscles movement helps to identify human emotions. Basic facial features are eyebrow, mouth, nose & eyes. The facial recognition, facial expression detection and its classification are the fields where Convolution Neural Network (CNN) plays very important role in the facial detection.The proposed module consisting of the terms Convolution Neural Network (CNNs), Rectified Linear Unit(RELU),pooling layer and fully connected layer. The experiment conducted on the extended Cohn-Kanade dataset (CK+): database gives that our approach is robust in dealing with video-based facial emotion recognition problem under lab-controlled environment. The Deep learning concepts i.e. convolution neural networks have achieved significant success in the area of computer vision including the difficult face recognition problems. The features have been calculated for the face model. The classifications of features were performed using CNN classifier. The network is trained using the standard CK+ database. The features have been calculated for the face model.

***IndexTerms* - Facial Expression Recognition, Convolutional Neural Networks, Computer Vision, CK+ database.**

I. INTRODUCTION

A facial expression plays an important role in communicating the state of our mind. Both humans and computer algorithms can greatly benefit from being able to classify facial expressions. Possible applications of automatic face expression recognition embody higher transcription of videos, movie or advertisement recommendations, detection of pain in medicine etc. Emotion being a highly subjective issue, maximum utilization of knowledge and science behind labeled data and extracting the components that constitute it, has been a challenging problem in the industry for many years. With the evolution of deep learning in laptop vision, emotion recognition has become a widely-tackled research problem. The Deep learning concepts i.e. convolution neural networks have achieved significant success in the area of computer vision including the difficult face recognition problems. In this model work, we propose two independent methods for this very task. The first technique uses autoencoders to build a novel illustration of every feeling, The results show that with more fine-tuning and depth, our CNN model can outperform the state-of-the-art methods for emotion recognition. The experiments conducted on the extended Cohn-Kanade dataset (CK+): database shows that our approach is powerful in managing with video-based facial emotion recognition drawback below lab-controlled surroundings. Facial expression recognition systems have attracted a lot of analysis interest among the sphere of computing. Many established facial expression recognition (FER) systems apply standard machine learning to extracted image features and these methods generalize poorly to previously unseen data. This project builds upon recent research to classify images of human faces into discrete emotion categories using convolution neural networks (CNNs).In this project, we have developed convolution neural networks (CNN) for a facial emotion recognition task. The goal is to classify each facial image in video into one of the seven facial emotion categories considered in this study. We trained CNN models with different depth using gray-scale images from the Kaggle website. We developed our models in Torch and exploited Graphics process Unit (GPU) computation so as to expedite the coaching method. We additionally gift the visualization of various layers of a network to indicate what options of a face will be learned by CNN models. Facial expression is the most powerful and natural non-verbal emotional communication method. Facial Expression Recognition (FER) has significance in machine learning tasks. Deep Learning models perform well in FER tasks, but it doesn't provide any justification for its decisions. Based on the hypothesis that facial expression is a combination of facial muscle movements, we find that Facial Action Coding Units (AUs) and Emotion label have a relationship in CK+ Dataset. The CNN model is trained with CK+ Dataset and classifies emotion based on extracted features.

The field of image process and analysis provides resolution for several advanced issues like improvement of degraded images for the aim of clarity, medical image process, identity verification etc. Automatic emotion detection kind the facial image is additionally comes below these classes of complicated problems. For personal right to liberty. Face is a very important a part of person and needs detection for various applications like security, forensic investigation. It needs correct techniques for face detection and recognition with challenges of various facial expressions, create variations. To overcome this challenge we tend to planned the final module for facial feeling detection exploitation CNN with four layers. The model with high accuracy is used anyplace in Medical, Distant education, Clever promoting, Driver police investigation. Overall method of automatic recognition of emotions and highlight key problems or challenges during these fields. Facial expression presents key mechanism to describe human emotion.From starting to end of the day human changes plenty of emotions, it may be because of their mental or physical circumstances. Although humans are filled with various emotions, modern psychology defines six basic facial expressions: Happiness,Sadness, Surprise,Fear, Disgust and Anger as universal emotions.Facial muscles movements help to identify human emotions. Basic facial features shown in Table 1 are eyebrow, mouth, nose & eyes. as happy, sad, surprise, fear, anger, disgust, and neutral is used in this project facial expression recognition is still a challenging problem in computer vision.In this work,we propose a simple model for facial expression recognition that uses Convolutional Neural Networ.Convolutional Neural Networks achieve better accuracy with big data.However,there are no publicly available datasets with sufficient data for facial expression recognition with deep architectures.Therefore, to tackle the problem).A study of the impact of each image pre-processing operation in the accuracy rate is presented.The proposed method: achieves competitive results when compared with other facial expression recognition methods - 96.76% of accuracy in the CK+ database - it is fast to train, and it allows for real time facial expression recognition with standard computers.

1.1 Problem Statement

The current work reflects on problem statement as:

"Facial Emotion Recognition in Video using CNN"

Existing system do not gives more accuracy and it uses Microsoft Kinect for 3D face model for detection. Kinect is a hardware having all inbuilt features.

- We proposed the general module which can useful for the security, medical and cybercrime's field.
- In our system, the full face is extracted by using Haar cascade algorithm in one frame, which detects and gives more accuracy from the full face expression detection in video.
- We used Open CV that is proposed model is easy to interface with hardware and used in many applications. Such as this model will be used for security purposed in house. Ex, if any stranger in front of door it can detect his emotion and save it in database. As we create SELF-LEARNING mode data i.e. it saves data and create new database for the users which are frequently detected when next time that user come it will take less time as it saved already in new database.

This model provided more accuracy because of CNN having more layer and detecting full face features as it gets more details features for each emotion. The highest accuracy occur 99% and average accuracy is between 89-95 %.

1.2. Aim

As facial expression convey more information and easy to classifies the details. By using CNN we proposed model to detect real time emotion and used it in many such as if Theft or Stalker is near the house who frequently visited the house, then this model make his different data base and detect him in less processing time at next time. Then we can call the police. Not just for security but other area also its help like Medical, Distant education, Clever Marketing, Driver surveillance. As aim for any model it must be help to solve real time problems, less time for processing, maximum Accuracy and this proposed model can help to find solution.

1.3. Objective

We proposed the general module which can useful for the security, medical and cybercrime's field. In our system, the full face is detected and it gives more accuracy from the full face emotion detection. The system uses which consisting of the four layers and it gives the accurate emotion detection.CNN is used from image classification and it can understands the spatial relation (relation between nearby pixels of image) for complicated image and video classification. If any new object is comes then by predicting the unknown object it placed and classified by category wise.The complexity is more by adding more layers in CNN as compare with RNN, MLP.

II. Literature Review

Mittal [1], had explained sentiments are feelings, emotions likes and dislikes or opinions which can be articulate through text, images or videos. They used deep learning techniques for sentiment analysis, as deep learning models have the capability for effectively learning the image behavior or polarity. Image recognition, image prediction, image sentiment analysis and image classification are some of the fields where Neural Network (NN) has performed well implying significant performance of deep learning in image sentiment analysis. In this module, they focused on deep learning as Deep Neural Network (DNN), Convolution Neural Network (CNN), Region-based CNN (R-CNN) and Fast R-CNN along with the suitability of their applications in image sentiment analysis and their limitations. The module study also discussed the challenges and perspectives of this rising field.

Aiman [2], proposed network is composed of a set of elaborately designed CNNs, RELUs and fully connected layers. The training dataset is augmented with synthetically generated samples by applying Gaussian and Poisson noise to each sample of the training set, thus doubling the size of the training set. They proposed demonstrate that the augmented training dataset actually improves the generalization power of CNNs. The network is trained using the standard AT&T face database. Using the proposed approach for limited training data, substantial improvement in recognition rate is achieved.

Majumder [3], this system presented a novel automatic facial expressions recognition system (AFERS) using the deep network framework. The proposed AFERS consists of 4 steps: 1) geometric feature extraction 2) regional local binary pattern (LBP) features extraction. 3) fusion of both the features using auto encoders and 4) classification using Kohenself-organizing map (SOM) based classifier. This model makes three distinct contributions. The proposed deep network consisting of auto encoders and the SOM-based classifier is computationally more efficient and performance wise more accurate. The fusion of geometric features with the use of LBP features auto encoders provides provides higher illustration of facial features. The SOM-based classifier proposed in this model improved by making use of a soft-threshold logic and a better learning algorithm. The performance of the proposed approach is valid on 2 widely used databases (DBs): 1) MMI and 2) extended Cohn–Kanade (CK+). An average recognition accuracy of 97.55% in MMI DB and 98.95% in CK+ DB are obtained using the proposed algorithm. The recognition results obtained from unique features are found to be clearly to both feature extraction using individual features furthermore recognition with a right away of the individual feature vectors.

Soleymani et.al [4], in this system emotions are time varying affective phenomena that are elicited as a result of stimuli. Videos and films in specially created to elicit emotions in their audiences. Detecting the viewers emotions outright will be wont to detect the emotional traces of videos. In this model, they did the approach in instantaneously detecting the emotions of video viewers' emotions from electroencephalogram (EEG) signals and facial expressions. A set of emotion inducing videos were shown to

participants while their facial expressions and physiological responses were recorded. The expressed valence (negative to positive emotions) within the videos of participants' faces were annotated by 5 annotators. The stimuli videos were also continuously annotated on valence and arousal dimensions. Long-short-term-memory continual neural networks (LSTM-RNN) and Continuous Conditional Random Fields (CCRF) were used in detection emotions automatically and endlessly.
They found that the results from facial expressions to be superior to the results from EEG signals. It analyzed the impact of the contamination of skeletal muscle activities on graphical record signals and located that almost all of the showing emotion valuable content in graphical record options are as a result of this contamination. However, their statistical analysis showed that EEG signals still carry complementary information in presence of facial expressions.
Tarnowski [5], this system presented the results of recognition of seven emotional states (neutral, joy, sadness, surprise, anger, fear, disgust) based on facial expressions. The features are calculated for three-dimensional face model. The classifications of features were performed using k-NN classifier and MLP neural network. They achieved, the classification accuracy of emotions - 96% for random division of data and satisfactory classification accuracy - 73%.
Lucey [6], The CK database has been used for both AU and emotion detection (even though labels for the latter have not been validated), comparison with benchmark algorithms is missing and use of random subsets of the original database makes meta-analyses difficult. To address these and other concerns, they presented the Extended Cohn-Kaneda (CK+) database. The number of sequences is increased by 22% and the number of subjects by 27%. The target expression for every series is totally FACS coded and emotion labels are revised and valid. In addition to this, non-posed sequences for several types of smiles and their associated metadata have been added. They present baseline results using Active Appearance Models (AAMs) and a linear support vector machine (SVM) classifier using a leave one-out subject cross-validation for both AU and emotion detection for the posed data. The emotion and AU labels, along with the extended image information and half tracked landmarks.
Kong [7], in this system face recognition (FR) with additional features achieves better performance than that with single one.
They proposed to fuse the multiple features into a more preferable presentation, which is more compact and more discriminative for better FR performance. As well, it taken the advantage of the dictionary learning framework to derive an effective recognition scheme. We evaluate our model by comparing it with other state-of-theart approaches, and the experimental results demonstrate the effectiveness of our approach. Recently, some methods are developed to deal with multiple types of features through sparse coding techniques. These methods directly use the original training set as multiple dictionaries, and impose some sparsity constraints on the coefficients for FR. They assumed the multiple features have some intrinsic relationships, which bridges all these features for better FR performance. With this assumption, we propose a novel method to generate a more compact and more discriminative dictionary for classification, and to fuse the multiple features into a more preferable representation. Through experimental validation, it show that the method outperforms other state-of-the-art methods on multi-feature face recognition task. Despite the decent performance, there is still large room to improve our proposed method.
Chen [8], this model were Video based facial expression recognition has been a long standing problem and attracted growing attention recently. The key to a prospering face expression recognition system is to take advantage of the potentials of audiovisual modalities and style strong features to effectively characterize the facial look and configuration changes caused by facial motions. We propose an efficient framework to deal with this issue during this paper. It studied, both visual modalities (face images) and audio modalities (speech) are utilized. A new feature descriptor referred to as bar graph of destined Gradients from 3 Orthogonal Planes (HOG-TOP) is planned to extract dynamic textures from video sequences to characterize facial look changes. And a brand new effective geometric feature derived from the warp transformation of facial landmarks is planned to capture facial configuration changes. Moreover, the role of audio modalities on recognition is additionally explored in our study. We applied the multiple feature fusion to tackle the video-based face expression recognition downside underneath lab-controlled setting and within the wild, severally. Experiments conducted on the extended Kohn-Kanada (CK+) information and therefore the Acted face expression in Wild (AFEW) four.0 information show that our approach is powerful in handling video-based facial expression recognition downside underneath lab-controlled setting and within the wild compared with the opposite progressive strategies.
Kahou, Pal [9], had worked on the challenge to classify the emotions expressed by the primary human subject in short video clips extracted from feature length movies. This involves the analysis of video clips of acted scenes lasting close to one-two seconds, together with the audio track which can contain human voices yet as background music. Their approach combines multiple deep neural networks for various knowledge modalities, including: (1) a deep convolutional neural network for the analysis of facial expressions among video frames; (2) a deep belief internet to capture audio data; (3) a deep autoencoder to model the spatiotemporal information created by the human actions delineate among the complete scene; and (4) a shallow network architecture targeted on extracted options of the mouth of the first human subject within the scene. They discussed each of these techniques, their performance characteristics and different strategies to aggregate their predictions. This yielded a test set accuracy of 35.58%. Using best strategy for aggregating our high acting models into one predictor we tend to were able to manufacture an accuracy of forty one.03% on the challenge check set.
These compared favorably to the challenge baseline test set accuracy of 27.56%. Module founded that convolutional network models learned using only their additional static frame training data sets were able to yield higher validation set performance if the labeled video data from the challenge was only used to learn the aggregation model and the static frames of the challenge training set were not used to train the underlying convolutional network. They put their efforts to create both SVM and MLP aggregator models lead to similar observations in that models quickly overfit the training data and no settings of hyperparameters could be found which might yield inflated validation set performance.
This is due to the fact that the activity recognition and bag of mouth models severely overfit the challenge training set and the SVM and MLP aggregation techniques being quite flexible also overfit the data and in such a way that no traditional hyper parameter tuning could yield validation set performance gains. These observation led us to develop the novel technique of aggregating the per model and per class predictions via random search over simple weighted averages. The ensuing aggregation technique is so of extraordinarily low complexness and also the underlying prediction was therefore extremely forced victimisation easy weighted mixtures of complicated deep network models, each of which did reasonably well at this task. As this yielded a marked increase in performance on both the challenge validation and test sets it leads to the

interpretation that given the presence of models that overfit the training data, it may be better practice to search a moderate space of simple combination models compared to more traditional approaches such as searching over the smaller space of SVM hyperparameters or even a moderately sized space of traditional MLP hyperparameters including the number of hidden layers and the number of units per layer.

III. RESEARCH METHDOLOGY

3.1 SYSTEM ARCHITECTURE

In the proposed module, first we want to take images from the video feed and then it saved in the dataset. While detecting the persons facial emotion, firstly it detects and then compare with the training set. After that, the trained dataset the module decides the person's emotion related with their expressions and also the person name is detected from the database. In the following Figure 1 is block diagram of Prediction Model shows the working of our implemented module.

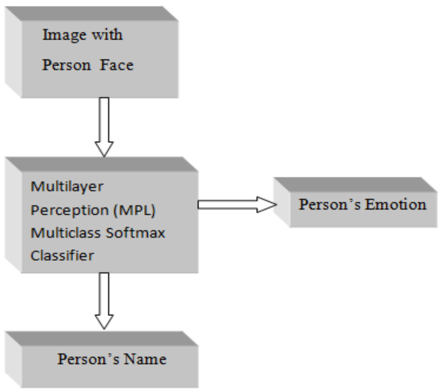


Figure 1 Prediction Model

3.2 CONVOLUTION NEURAL NETWORK

CNN is the feed forward neural network that is generally used to analyze the visual images by processing data. This layer can filter out of features form the data. CNN is better for image and speech classification as well as for complicated image classification. It is suitable for special data i.e. Images and understands the special relation(relation between nearby pixels of image) so, we can used it for video and complicated image classification. It is more powerful than KNN and MLP.

3.2.1 Working Process of CNN

CNN consisting of the total four layers that are as follows

- Convolution layer
- ReLU (Rectified Linear Unit) layer
- Pooling layer
- Fully connected layers

Convolution is the first layer in CNN to extract features from an input image.Convolution stored result between pixels by learning image features using small squares of input data. It is a mathematical operation i.e. (multiplication and summation) that takes two inputs such as image matrix and a filter or kernel. It consists of a convolution mask, bias terms and a function expression. Together, these generate output of the layer.

In the CNN algorithm, each sparse filter is replicated across the entire visual field. These units then form a feature maps, these share weight vector and bias. Output after this convolution passed to non- linear activation function.

From left to right in the above Figure 2 you can observe that:

- The real time input (capture from video) image that is scanned for extraction. The filter that passes over it is the light rectangle. At first, we will break down picture into a series of overlapping 3*3 pixel tiles.
- After that, we will run each of these tiles via a simple, single-layer neural network by keeping the weights unaltered. This will change the collection of tiles into an array. As we kept every of the images small (3*3 during this case), the neural network required to method them stays manageable and little.
- The output values will be taken and arranged in an array that numerically represents each area's content value in the photograph, with the axes representing color, width and height channels. For each we would have a 3*3*3 representation in this case.
- The next step is the pooling layer which takes 3 or 4 dimensional arrays and applies a down sampling function together with spatial dimensions. The result is a pooled array that contains only the image portions that are important while discarding the rest, which minimizes the computations that are needed to be done while also avoiding the overfitting problem.
- The Activation maps are arranged in a stack on the top of one another, one for each filter which is used in proposed model are larger rectangle is one patch to be downsampled.
- The activation maps condensed via downsampling.
- A new group of activation maps generated by passing the filters over the stack that is downsampled first.
- The second downsampling – which condenses the second group of activation maps.
- The downsampled array is taken and utilized as the regular fully connected neural network's input. Since the input's size has been reduced dramatically using pooling and convolution, we must now have something that a normal network will be

able to handle while still preserving the most significant portions of data. The final step's output will represent how confident the system is that we have the picture of a same image or different.

- A fully connected layer that gives output with one label per node.

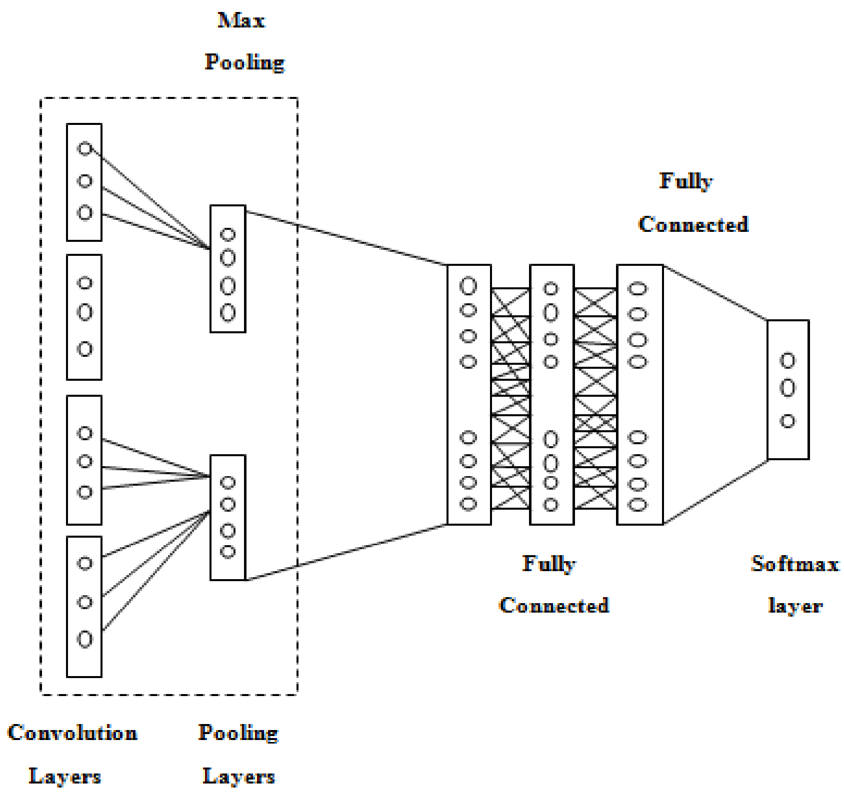


Figure 2 Complete flow of CNN

3.3 SYSTEM FLOWCHART

The first step in the convolution layer which is in turn has several steps in itself. Figure 3 shows flowchart, which represents work flow our proposed system. The proposed methodology of the system is as follow which contains some important points such as algorithm etc.

Step 1

Initialize camera: In first step, the camera is open and the frame for our face will comes. We can use our inbuilt systems camera i.e. webcam or any other usb camera. After that, it goes to the take for video process.

Step 2

Captures image from video feed: The images are captured from video feed in color format. The images of full face with different expression categorized are saved in the dataset.

Step 3

Convert the image into the gray scale image

Then the colored image converts into the gray scale in matrix format image after extraction.

Step4

Face detection: Full face is detected in this stage.

Step 5

Crop and the resize the image: Crop the image from full face. The cropped image resized into the 48*48 matrix format.So, we can easily cut the unwanted background which is not useful for expression detection.

Step-6

Emotion detection from the facial expression: From the cropped face image, the emotion is detected under the given seven category .If this step fails then it starts again with the first step.

Step 7

Resultant Output: If the facial expression is detected then it shows the output in terms of whether the person is under which seven universal expression states.

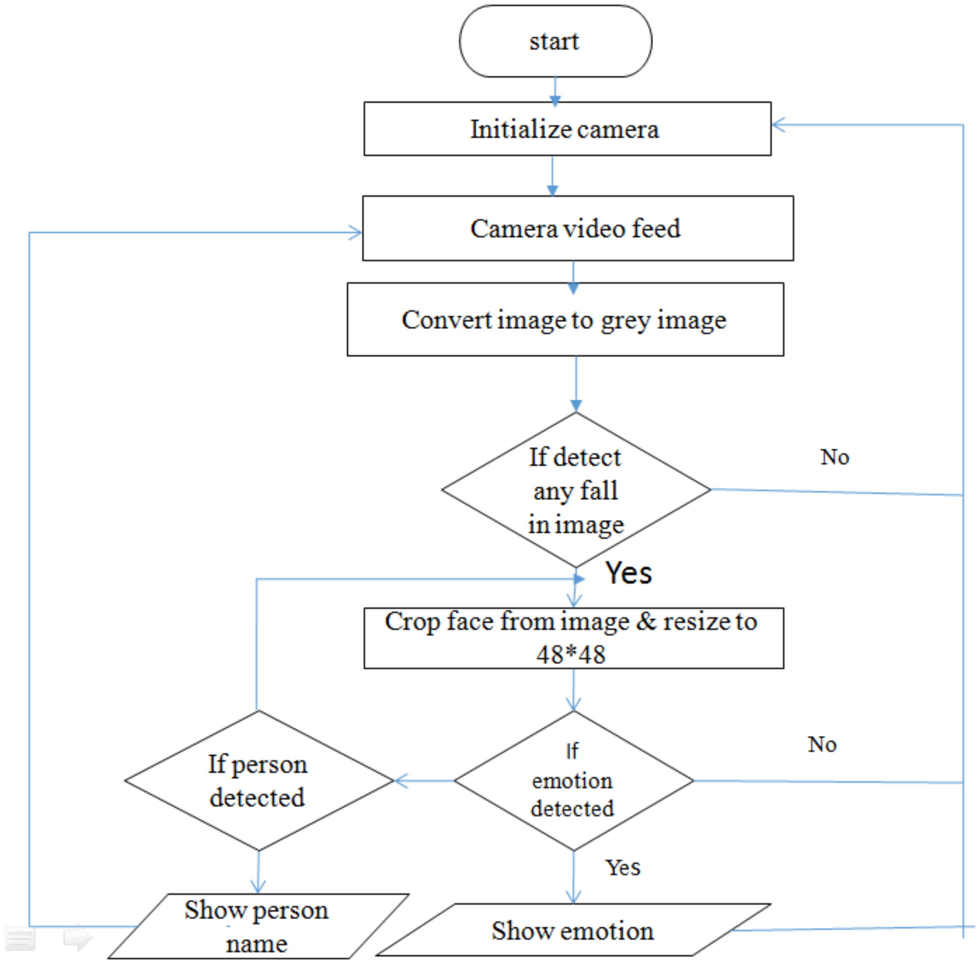


Figure 3 System flow

The resultant image shows in the Neutral, Happy, Sad, Surprise, Anger, Fear and disgust category and also it shows whether the person is unknown or known and if known then it shows person's name verifying from our saved dataset.

3.4 TRAINING AND TESTING

The training step took as input an image with a face. Thereafter, an intensity normalization is applied to the image. The normalized images are used to train the Convolutional Network. During training, Figure 4 shows the system received a training information comprising grayscale images of faces with their various expression label and learns a collection of weights for the network. To ensure that the training performance is not affected by the order of presentation of the examples, validation dataset is used to choose the final best set of weights out of a collection of trainings performed with samples presented in numerous orders. The output of the training step is a set of weights that achieve the best result with the training data.

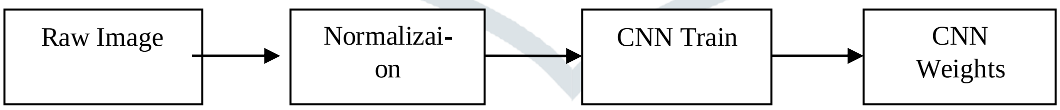


Figure 4 Training Phase

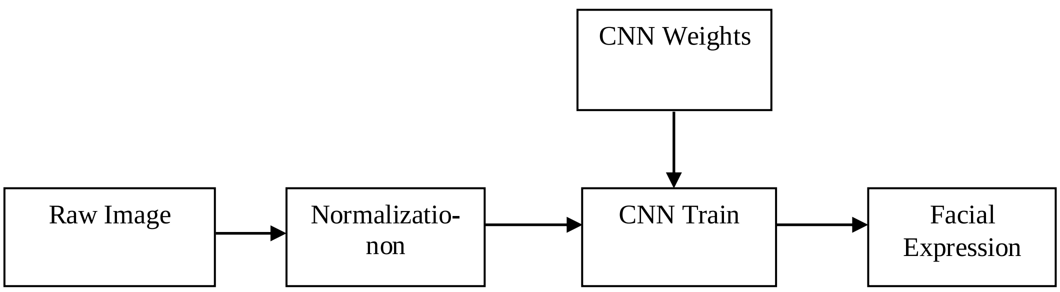


Figure 5 Testing Phase

In above Figure 5 shows the Testig phase in which the system received a grayscale image of a face from test dataset, and output the predicted expression by using the final network weights learned during training. Its output is a single number that represents one of the seven basic expressions.

3.5 DATABASE

The dataset from an Extended Cohn-Kanade (CK+) database is used for training and testing purpose in our proposed model. CK+ database is a open source database, which can easily available and download in our system for testing and training in emotion recognition. The number of sequences is increased by 22% and the number of subjects by 27%. The target expression for each sequence is fully FACS coded and emotion labels have been revised and validated. In addition to this, non-posed sequences for several types of smiles and their associated metadata have been added. It represents baseline results using Active Appearance Models (AAMs) and a linear support vector machine (SVM) classifier using a leave-one-out subject cross-validation for both AU and emotion detection for the posed data. The emotion and AU labels, along with the extended image data and tracked landmarks. The ck+ database contains more than 35000 images from which 28712 images are used for training purpose, 7178 images are used for testing and remaining are used for validation purpose.

The Figure 6 clearly shows some of testing images with all seven emotions i.e. happy, disgust, surprised, sad, angry, neutral and fearful. Which used for testing purpose in our proposed model.

Figure 6 Testing Images from CK+ Database with Different Emotion

Ck+ dataset contains 7178 testing images containing all seven emotions i.e. happy, disgust, surprised, sad, angry, neutral, and fearful from near about 35000 total images. Few of them are shown in Figure 7 which used for testing purpose in our proposed model.

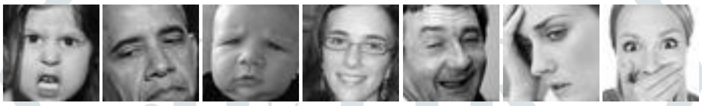


Figure 7 Training Images from CK+ Database with Different Emotion

3.7 SOBEL EDGE DETECTION

The Sobel operator, sometimes called the Sobel–Feldman operator or Sobel filter, is used in image processing and computer vision, particularly within edge detection. In our project we use Sobel Edge Detection for facial emotion detection.

Pseudo-Codes For Sobel Edge Detection

Input: A Sample Image. Output: Detected Edges.
Step 1: Accept the input image.
Step 2: Apply mask Gx, Gy to the input image.
Step 3: Apply Sobel edge detection algorithm and also the gradient.
Step 4: Masks manipulation of Gx, Gy separately on the input image.
Step 5: Results combined to find absolutely the magnitude of the gradient.
Step 6: the absolute magnitude is that the output edges.

3.8 DETAILS OF IMPLEMENTATION

Based on Methodology we implemented our module by adding 4 convolution layers in Convolution Neural Network (CNN). More layers in CNN gives more accuracy. Keras is a higher level library which operates over either TensorFlow or Theano and is intended to stream-line the process of building deep learning networks. The first required Conv2D parameter is the number of filters that the convolutional layer will learn.Conv2D layers in between will learn more filters.The detail stepwise implementation shown in Figure 8. The convolution layer computes the output of neurons that are connected to local regions or receptive fields in the input, each computing a dot product between their weights and a small receptive field to which they are connected to in the input volume. We implemented a generalized module for facial emotion detection by adding 4 layers of convolutional in cnn.Which means that our neural network will be a linear stack of layers. This netwok will have the following components.

- **Convolutional Layers:** These layers are the building blocks of our network and these compute dot product between their weights and the small regions to which they are linked.Our propsed model added four convoltional layers in CNN.
- **Activation functions**: are those functions which are applied to the output of all layers in the network.In this project,we will resort to the use of two functions-Relu and softmax.
- **Pooling layers:** These layers will downsample the operation along the dimensions.This helps reduce the spatial data and mimimize the processing power that is required.
- **Dense layers:** These layers are present at the end of CNN.They take in all the features data generated by the convolution layers and do the decision making.
- **Dropout Layers:** randomly turns off a few neurons in the network to prevent overfitting.

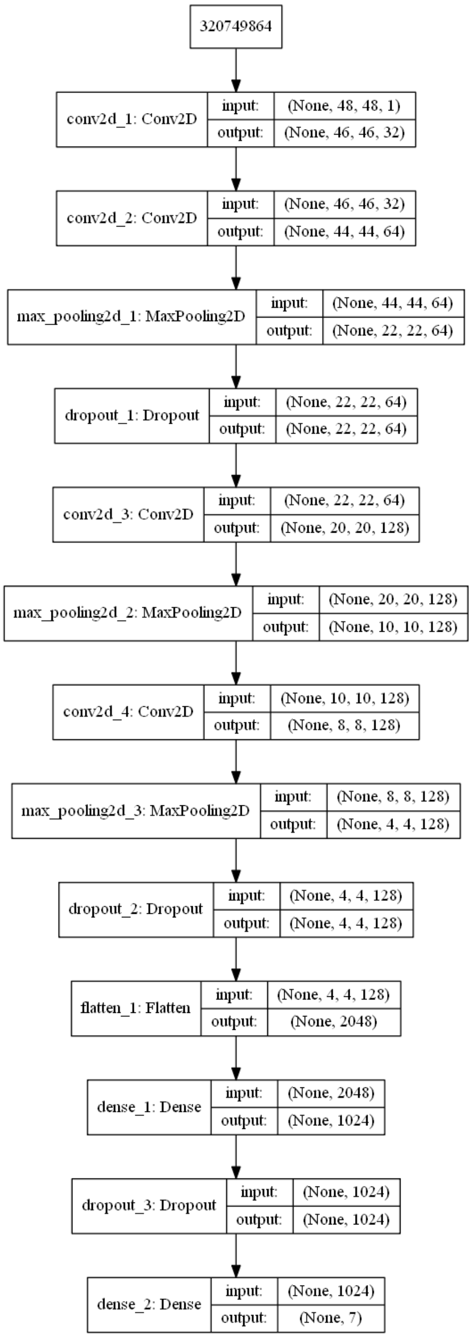
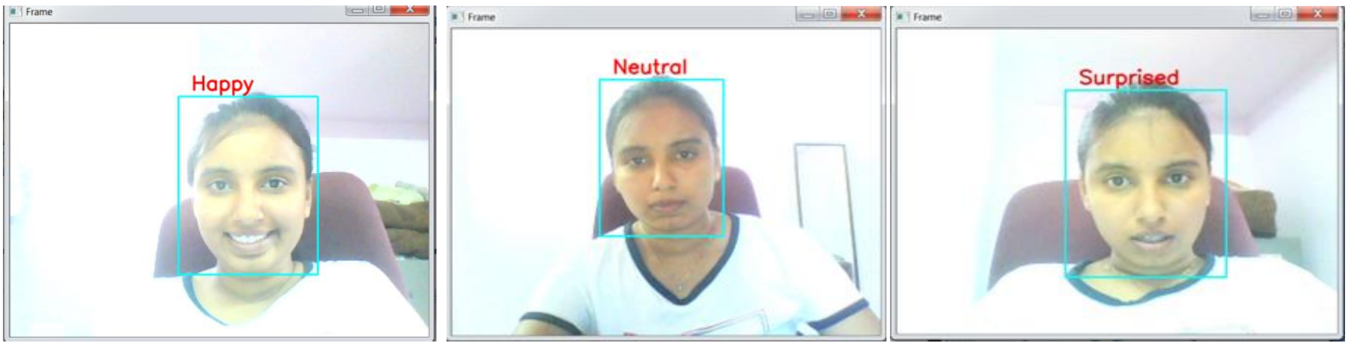


Figure 8 Implemented Model

IV. RESULTS AND DISCUSSION

4.1 Results

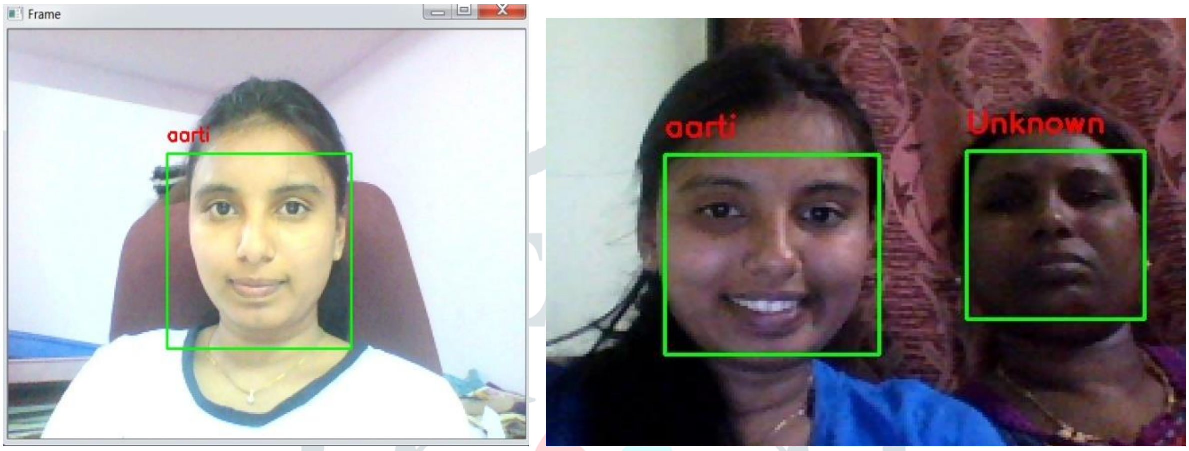
The below images shows screenshots of results, the resultant expression detection from full face detection which comes from the proposed system. The both pictures shows different expression contains one with happy, surprised and other emotions from data by detecting there full face expression. As compare to existing model the average accuracy is more in proposed model.The below Figure 9 shows the facial emotion detecteion with all seven expressions.



(a) (b) (c)

(d) (e) (f)

Figure 9 (a), (b), (c), (d), (e) & (f) Facial Emotion Detecteion with All Expressions



(a) (b)

Figure 10 (a) & (b): Full Face Detection with Name Displayed from Database

As shown in Figure 10(a) & (b) Full face detect with the users name which is feed in database and the unknown person which is not present in the database is shown, if that user suppose myself aarti is is feed in database then easily I can detected by name from our own database.

Self-Learning mode-In proposed model we create a self- learning mode that is this model can create new database of those person who are frequently detected and save them in new database. when next time that person came it's used takes less time to detect that person as it's already create new database for it.

Based on few attributes we will getting following analytical result. From proposed model and previous model. The graph plotted related with accuracy in percentage with reaspect to seven emotions is as shown below in Figure 11

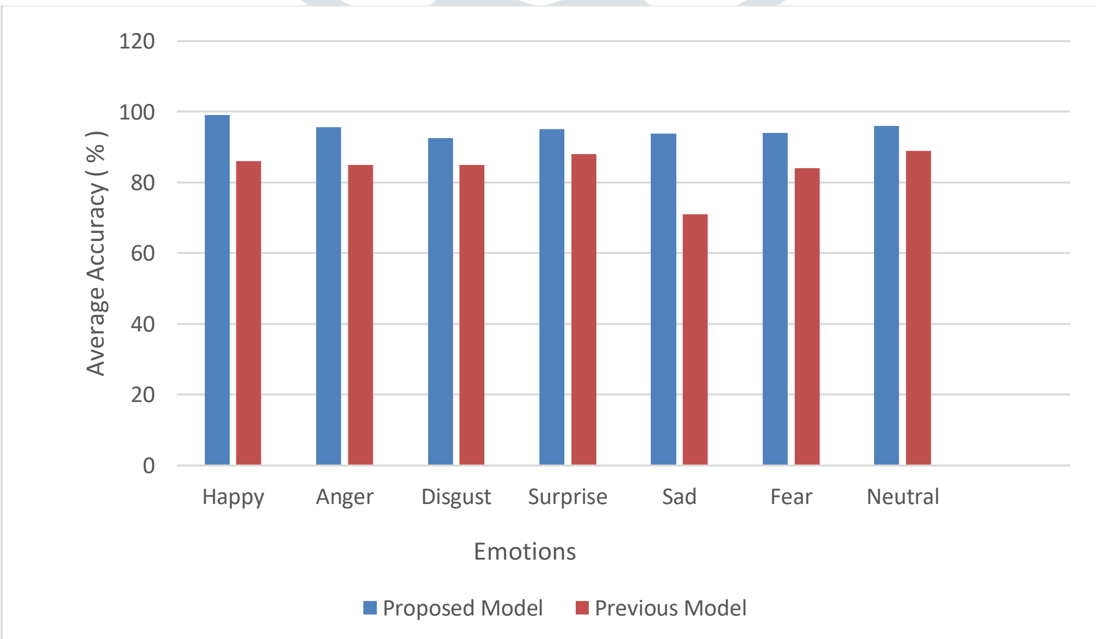


Figure 11 Graph Chart for Accuracy of Proposed Model and Previous Model

The below Figure 12 shows the graph for model accuracy and model loss.From the plot of accuracy we can see that the model could probably be trained a little more as the trend for accuracy on both datasets is still rising for the last few epochs. We can also see that the model has not yet over-learned the training dataset, showing comparable skill on both datasets.

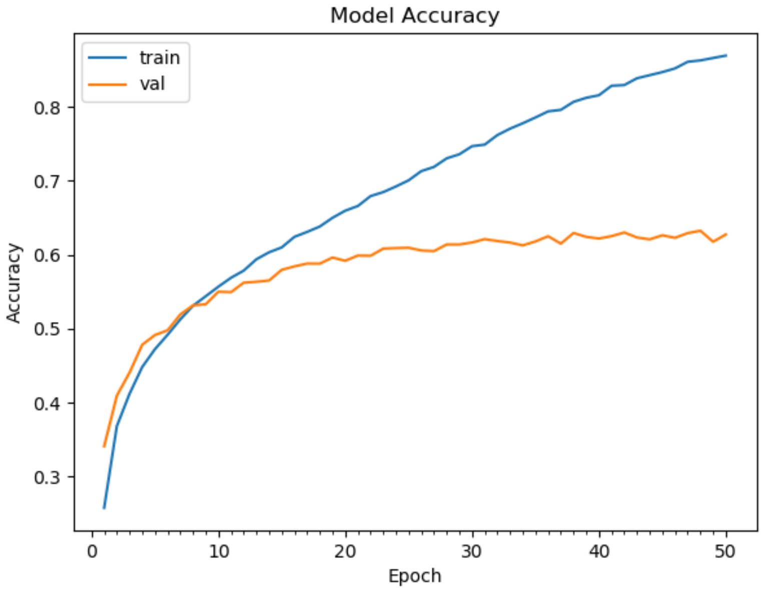


Figure 12 Graph for Facial Emotion Recognition Model Accuracy

The below Figure 13 shows the graph for model loss. From the plot of loss, we can see that the model has comparable performance on both train and validation datasets (labeled test). If these parallel plots start to depart consistently, it might be a sign to stop training at an earlier epoch.

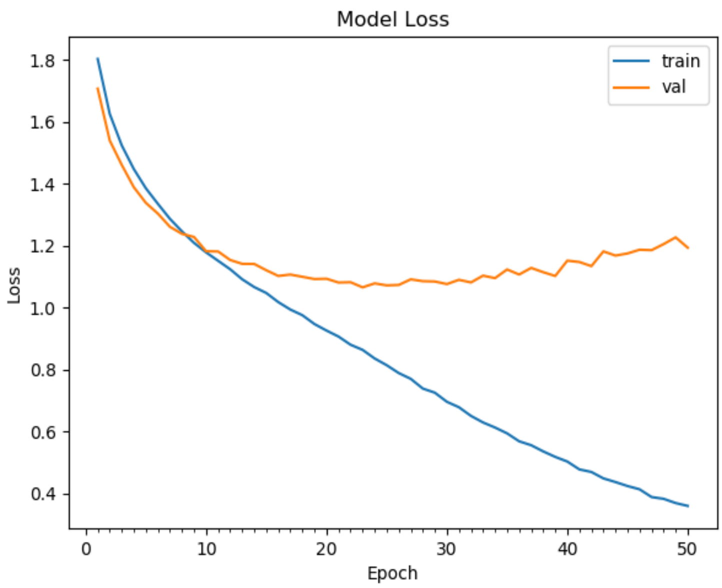


Figure 13 Graph for Facial EmotionRecognition Model Loss

IV. CONCLUSION

In this, we researched a generalized model which will used to recognize facial emotion easily with highest accuracy. Extensive efforts have been made over the past two decades in academia, industry and government to discover more robust methods of assessing truthfulness, deception, and credibility during human interactions. Many Efforts have been made to detect human expressions. Emotions are due to any activity in brain and it is known through face, as face has maximum sense organs. Hence human facial activity is considered. The objective of this proposed model is to give brief introduction towards techniques as we used CNN, forn application such as security, medical etc. It challenges accuracy by using CNN as well as Self-Learning mode which is to find whether the user is know or unknown by showing it’s name, emotions interfacing, in detecting real time detection of emotions. The module detects emotions by full face in single frame through the images captured during video. The effectiveness of the proposed approach is testified by the recognition performance, computational time, and comparison with the state-of-the-art performance.The experimental results also demonstrate significant performance improvements due to taking of facial movement features and promising performance beneth face registration errors.In near

Future, this technique will be more developed and eventually it's tackle the problems and get more efficient result. Future work should attempt to combine our technique with other modalities such as audio modality which will give more accuracyately to find the users emotions, including working with other datasets.

VI. Acknowledgment

I would like to thank Dr. V.E. Chirchi for her valuable discussions and help in implementing this research work.

References

[1] N.Mittal, D.Sharma and M.L.Joshi "Image Sentiment Analysis using Deep Learning" IEEE/WIC/ACM International Conference on Web Intelligence (WI) 2018 .

[2] U.Aiman,V.P.Vishwakarma "Face Recognition Using Modified Deep Learning Neural Network" 28th ICCCNT 2017

[3] A.Majumder,L.Behera and V. K. Subramanian "Automatic Facial Expression Recognition System Using Deep Network-based Data Fusion" pp.2168-2267 2016.

[4] M.Soleymani,S.a.Esfeden and Yun Fu " Analysis of EEG signals and facial expressions for continuous emotion detection" p.p 1949-3045.

[5] P.Tarnowski, M.Kołodziej, A.Majkowski, R.J. Rak "Emotion recognition using facial expressions" in International Conference on Computational Science, ICCS 2017.

[6] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+) A complete dataset for action unit and emotion-specified expression," IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 94-101.

[7] Shu Kong, Xikui Wang, Donghui Wang, Fei Wu "Multiple Feature Fusion for Face Recognition".

[8] J.Chen and Z.Chen "Facial Expression Recognition in Video with Multiple Feature Fusion", pp. 1949-3045.

[9] Samira Ebrahimi Kahou, Christopher Pal, Xavier Bouthillier, Pierre Froumenty, Çaglar Gülçehre, Roland Memisevic, Pascal Vincent , Aaron Courville & Yoshua Bengio "Combining Modality Specific Deep Neural Networks for Emotion Recognition in Video".

[10] L.Zhang,and Dian Tjondronegoro "Facial Expression Recognition Using Facial Movement Features" IEE Transactions on affective computing, p.p 1949-3045,2011.

[11] Wisal Hashim Abdulsalam, Rafah Shihab Alhamdani, and Mohammed Najm Abdullah "Facial Emotion Recognition from Videos Using Deep Convolutional Neural Networks".

[12] Prudhvi Raj Dachapally "Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units".

[13] Haitao Xiong, Qing Liu1 , Shaoyi Song1 and Yuanyuan Cai, "Region-based convolutional neural network using group sparse regularization for image sentiment classification".

[14] Shima Alizadeh, Azar Fazel "Convolutional Neural Networks for Facial Expression Recognition".

[15] S. Z. Li and A. K. Jain, "Handbook of face recognition," springer, 2011.

[16] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886-893.