Facial Emotion Recognition and Detection Using CNN

Boddepalli Kiran Kumar[1]
Associate Professor, Dept. of CSE, Aditya College of Engineering, Srampalem

Affiliated toJawaharlal Nehru Technological University Kakinada,
East Godavari, India

Korla Swaroopa[2]
Associate Professor, Dept. of CSE, Aditya Engineering College, Surampalem

Affiliated toJawaharlal Nehru Technological University Kakinada,
East Godavari, India

Tarakeswara Rao Balaga[3]
Professor,Dept. of CSE**,** KallamHaranadhareddy Institute Of Technology, Guntur,
Affiliated toJawaharlal Nehru Technological University Kakinada,
East Godavari, India

Abstract: People's facial expressions reveal a common set of feelings that we all share. Face-recognition technology has been deployed in a wide range of applications that require additional security or personal information. Facial expressions such as sadness, happiness, surprise, rage, and fear may all be used to determine a person's emotional state using facial emotion detection. Face emotion recognition and detection is critical for marketing purposes. Customers' reactions to all of a company's products and offerings are the lifeblood of the majority of enterprises. It is possible to determine whether or not a consumer is satisfied with a product or service based on their emotional response to an image or video captured by an artificially intelligent system. Using transformed photos, several machine learning approaches, such as Random forest and SVM, were previously utilised to estimate sentiment. Computer vision, for example, has made significant strides in recent years thanks to advancements made possible by deep learning. Facial expressions may be detected using a convolutional neural network (CNN) model. This dataset is used for both training and testing purposes.

Keywords: CNN, facial expressions, intelligent, machine learning, SVM.

Introduction:

Any inter-personal relationship involves emotions. These may be expressed through facial expressions, conversation, gestures, and even attitude. Faces are the most apparent and information-rich options for Emotion Recognition. Faces are also easier to gather and process than other expressions. A facial expression is a complex movement of the face muscles that conveys the subject's feelings to others. Expressions convey a person's inner feelings. For these reasons, researchers in psychology, animation, HCI, linguistics, neurology, medicine, and security are becoming interested in a human-computer interaction system for autonomous face recognition.

Face and expression analysis using computers is a new area. Emotion analysis is matching a face to an emotion. So, the goal is to read a person's feelings from their face. Automated face expression analysis systems facilitate human-machine interaction. But this is not an easy process. Many characteristics of facial expressions can now be retrieved and evaluated for good sentiment analysis using deep learning and convolutional neural networks (CNN) [5]. Our goal is to create a deep learning-based model for face sentiment analysis. Using a convolutional network architecture, face characteristics can classify emotions into Disgust, Fear, Anger, Surprise, Happiness, Sadness, and Neutral.

In this study, a typical neural network with data augmentation is used to recognise face expressions. This method can categorise images into Anger, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Due to their huge number of filters, CNNs are superior for image identification tasks.

2. LITERATURE SURVEY:

An individual's emotions have a significant impact on their ability to learn. Because of their inability to regulate their emotions, kids with high-functioning autism (HFA) often have difficulty focusing and paying attention in class. Once unpleasant emotions have emerged in HFA children, attempts to regulate them are typically unsuccessful since they are difficult to quiet down. In an e-learning environment, students with high-functioning autism (HFA) may benefit from early detection of emotional changes and prompt provision of adaptive emotional regulation tools to manage negative emotions. An emotion identification approach using facial expressions was proposed in this work. For the objective of creating emotion identification classifiers, an experiment was conducted to gather facial-based landmark signals [1]. Classifiers for recognising emotions were constructed using the sliding window approach and support vector machines (SVM). Information Gain (IG) and Chi-square were employed for feature assessments in order to determine strong features for emotion identification. Classifiers with various sliding-window values were also tested for efficacy. The results of the experiments show that the suggested strategy is able to distinguish between different samples.

Due to the wide range of ways in which emotions can be portrayed, automatic affect identification is difficult. In addition to multimedia retrieval and human-computer interaction, applications can be explored. Using deep neural networks to predict emotional states has been a major success in recent years These results inspired us to develop an emotion identification system that makes use of both audio and visual cues. In order to extract the emotional content of distinct speaking styles, substantial traits must be extracted. A Convolutional Neural Network (CNN) is used for voice extraction [6], while a deep residual network (ResNet) of 50 layers is used for the visual modality. A machine learning algorithm must not only be sensitive to outliers but also be able to represent the context [2], which goes hand in hand with the necessity of feature extraction. Long Short-Term Memory (LSTM) networks are used to solve this issue. Using the correlations between each stream, we were able to significantly outperform traditional approaches based on auditory and visual handcrafted features for the prediction of spontaneous and natural emotions on the RECOLA database of the AVEC 2016 research challenge on emotion recognition when training the system from end to end.

Feature learning for multi-view facial expression identification using an unique deep neural network (DNN) approach is presented and tested in this study (FER). To begin, SIFT features corresponding to a collection of landmark points in each face picture are retrieved. Then, these features are combined to create a composite image. It is then utilised as input data for a well-designed DNN model to learn the most discriminative features of expression categorization from the generated SIFT feature vectors SIFT feature vectors and their related high-level semantic information are characterised by multiple levels in the DNN model. We can discover a set of optimum features for categorising face emotions across various facial perspectives by training the DNN model [3]. It was determined that our approach outperformed existing techniques by comparing it to the BU-3DFE and Multi-PIE non-frontal facial expression datasets, which were each utilised to try out our method.

An expression transfer method from humans to many stylised characters is proposed as Deep Expr (DEEP). Two Convolutional Neural Networks (CNN) are trained initially to recognise the expressions of both people and cartoon characters. The mapping from humans to characters is then learned using a transfer learning approach, resulting in a shared embedding feature space [7]. Using this embedding, it is possible to get images based on facial expressions and on the expressions of fictional characters. Character expressions based on people can be found using our perceptual model. We put our approach to the test using a variety of retrieval tasks and a stylized character dataset [4]. A facial expression expert and a series of Mechanical Turk trials have also shown that the suggested characteristics' projected ranking order is strongly associated with the actual ranking order.

3. PROPOSED SYSTEM:

With the goal to improve the process of facial sentiment analysis systems, a classification mechanism is proposed using a CNN architecture. Due to the need of large data required for training of deep networks, FER2013 dataset which is available publically is utilized here. In the subsequent section, the features of our chosen dataset are listed out, followed by the description of our network architecture and finally the performance measures used for evaluation.

Proposed system advantages: High picture quality improves the effectiveness of facial recognition; even low-resolution photographs are usable with the suggested method; and Higher accuracy while while being more computationally efficient.

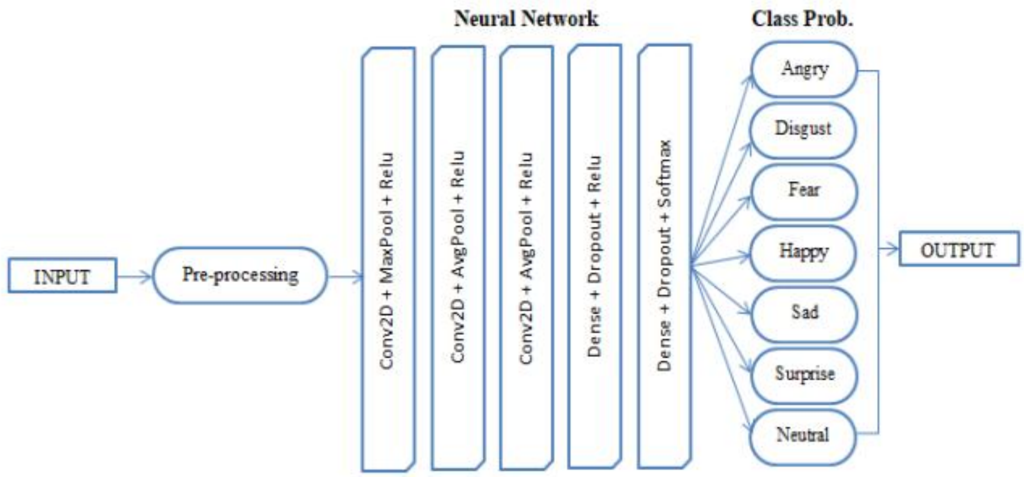

FIG 1. Block diagram of facial emotion recognition and detection using cnn

Figure 1 describes the proposed model for facial emotion detection using CNN machine learning model

Applications: It is widely used in robotics, where Human Computer Interactions (HCI) are extremely crucial. The ability to predict emotions based on facial expressions has a number of scientific advantages.

Proposed method MODULES:

Modules used in our model design are:

i.Face capturing module
ii. Pre-processing module
iii. Training module
iv. Face recognition module
v . Expression recognition module

Face Capturing Module: • During this phase, we are taking pictures of people's faces for further processing. We are utilising a webcam or an external web camera for this purpose [10]. There is no way to complete the procedure without first taking the image, and there is no way to identify the emotions without first capturing the image.

Preprocessing Module: Following the capture of photos, we will do image processing on the captured images. The grey scale photos will be created by converting the colour photographs to grey scale.

Training module: This step will involve the preparation of a dataset, which will consist of a binary array of all the photographs that have been taken. The collected photographs will be saved in a.YML file, which will contain all of the face data that was obtained. The YML file allows us to process the collected photos more quickly because of its compressed nature.

Face Recognition Module: The first phase in the face recognition process is to train the host system on the facial data that has been collected. The face is photographed using the web camera on the computer system, which captures 60 different photos of the subject's face [8]. In this session, we will learn how to detect people's faces using the LBP algorithm. The abbreviation LBPH stands for local binary pattern histogram. With the face ID and NAME that were previously stored, it will recognise the faces in the database.

Face Expression Recognition Module: Facial expression recognition software is a system that detects emotions in human faces by using biometric indicators [9]. Because it collects and analyses information from images, it is possible to offer an unfiltered, unbiased emotional reaction or data that is unfiltered and impartial.

SYSTEM DIAGRAM:

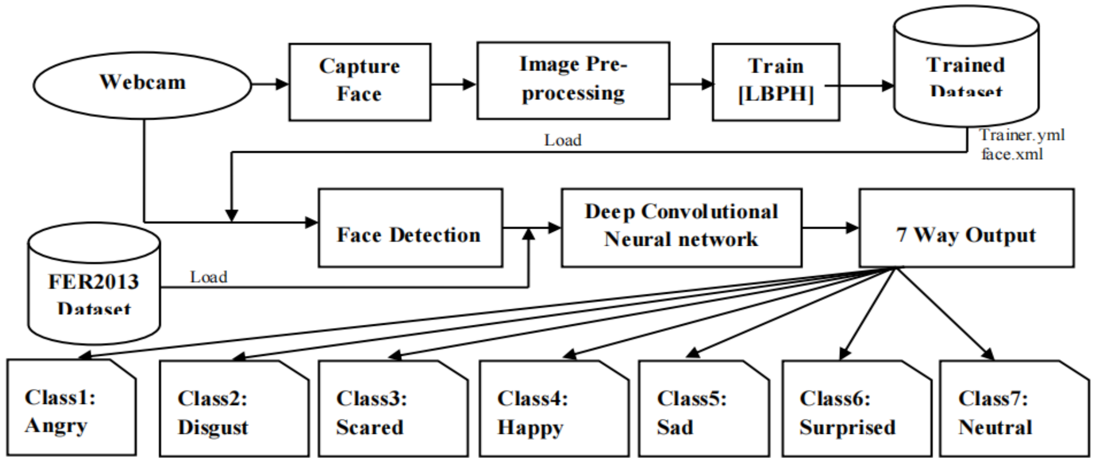

FIG 2. System diagram of facial emotion detection

A webcam is used to capture, identify, and recognise the facial expressions of a person, which is done through the use of software. In the camera, a rectangular frame on the face area is obtained; this identification of the face region from a non-facial region is accomplished by the employment of the Viola Jones method, the LBPH Face Recognizer algorithm, and the Haarcascade frontal face dataset, among other techniques. Captured person faces are preprocessed before being saved in a folder labelled with the subject's ID and name. These photos are trained using the LBPH method, and the resulting trained dataset is saved as Trainer.yml in the Trainer folder. During the Face Detection process: A trained dataset is used to match the face in a video camera with the face in the dataset. If a person's face matches that in the trained dataset, his or her ID and name will be displayed on the screen. In order to classify the obtained face, convolutional neural networks are used in conjunction with the FER2013 database to do the classification [11]. The facial expression represents the chance of acquiring the maximum level of expression based on the characteristics of the individual. One of seven possible facial expressions is presented in conjunction with the recognised picture of the subject.

System flowchart:

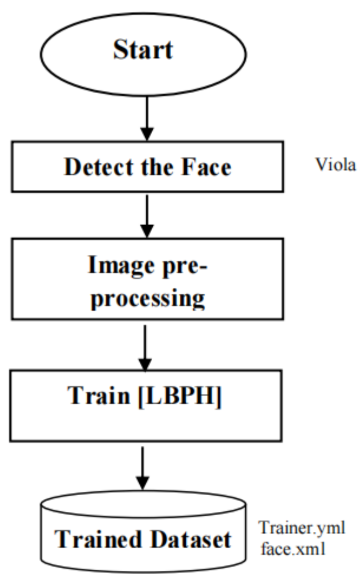

Fig 3.flowchart of training

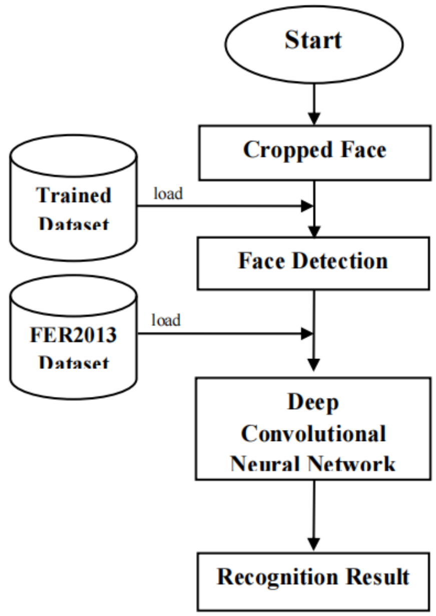

Fig 4. Flowchart of testing

During training Phase, the system received a training data comprising grayscale images of faces with their respective expression label and learns a set of weights for the network. 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. 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 set of trainings performed with samples presented in different orders.

The output of the training step is a set of weights that achieve the best result with the training data. During test, 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 expression.

DATASETS USED:

Several public databases were used in order to assess face expression recognition algorithms:
Frontal face dataset from Haarcascade: HaarCascade Classifier is used to recognise faces in pictures utilising characteristics. The frontal face is detected using the haarcascadefrontalface default.xml [17]. It was created by Viola and Jones in response to a proposal made in 1998[2] by Papa Georgiou et al. To verify that the retrieved faces are all in the same location, we utilised an additional classifier named 'haarcascade eye.xml' from the same OpenCV library[18]. This identifies the region around the eyes and then adjusts the left and right borders of the face window to maintain an equal distance between the eyes and the sides of the face. Thus, superfluous information (such as hair, ears, and background) is removed, and the retrieved faces have their locations adjusted.FER2013 dataset: FER2013[15] is an open-source dataset generated by Pierre-Luc Carrier and Aaron Courville for an ongoing project and later given publicly for a Kaggle competition[15]. The FER2013 database was launched during the 2013 International Conference on Machine Learning's Challenges in Representation Learning. FER2013 is a massive and unrestricted database that was automatically compiled using the Google image search API. After rejecting incorrectly labelled frames and modifying the cropped region, all photos have been registered and resized to 48*48 pixels. This dataset contains 35,887 grayscale, 48x48-pixel pictures of faces displaying a range of emotions -7 emotions, all labeled-.

Emotion labels in the dataset:
0: -4593 images- *Angry*
1: -547 images- *Disgust*

2: -5121 images- *Fear*
3: -8989 images- *Happy*
4: -6077 images- *Sad*
5: -4002 images- *Surprise*
6: -6198 images- *Neutral*

The FER-2013 dataset was created by gathering the results of a Google image search of each emotion and synonyms of the emotions. The images in FER-2013[15]consist of both posed and un-posed headshots.

Fig 5: Example Images of FER2013 dataset

Figureillustrating variability in illumination, age, pose, expression intensity, and occlusions that occur under realistic conditions. Images in the same column depict identical expressions, namely anger, disgust, fear, happiness, sadness, surprise, as well asneutral.

The data file contains 3 columns — Class, Image data, and Usage.

a) Emotion class: is a digit between 0 to 6 and represents the emotion depicted in the corresponding picture. Each emotion is mapped to an integer as shown below.
   0- 'Angry' 1-'Disgust' 2-'Fear' 3-'Happy' 4-'Sad' 5-'Surprise' 6-'Neutral'
b) Image data: is a string of 2,304 numbers and these are the pixel intensity values of our image, we will cover this in detail in a while.
c) Usage: It denotes whether the corresponding data should be used to train the network or test it.

4. Results:

The training and testing datasets are from a Kaggle Facial Expression Recognition Challenge (FER2013). It comprises of precropped grayscale photos of faces classified as pleased, sad, disgusted, angry, surprised, fearful, or neutral. The webcam image will be used as the input for processing the output. The output labels human facial expressions as pleased, sad, disgusted, angry, surprised, fearful, or neutral.

Neural Evolutionary Network

Convolutional Operation is the first step.

Our strategy of attack begins with a convolution operation. This stage will discuss feature detectors, which act as filters for the neural network. Additionally, we will explore feature maps, including how to learn their parameters, how patterns are recognised, the layers of detection, and how the findings are shown.

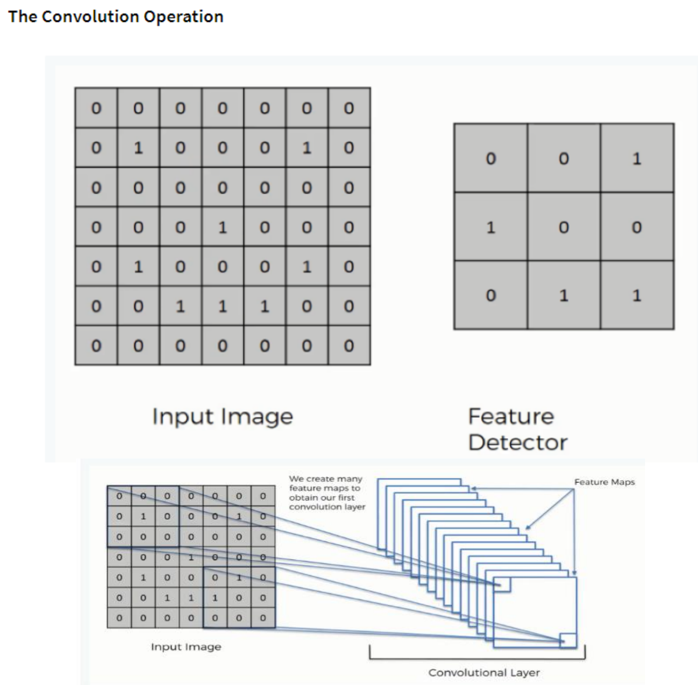

Fig 4.5.1. Convolutional Operation

Step (1b): RELU Layer

The second part of this step will involve the Rectified Linear Unit or Relook. We will cover Relook layers and explore how linearity functions in the context of Convolutional Neural Networks.
Not necessary for understanding CNN's, but there's no harm in a quick lesson to improve your skills.

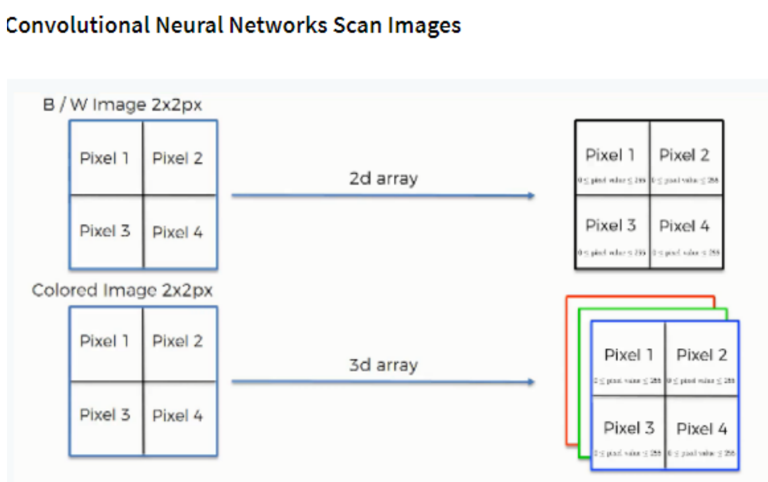

Fig 6: RELU Operation

Step 2: Creating a Pooling Layer
This section will discuss pooling and how it works in general. However, our nexus here will be a specific sort of pooling; maximum pooling. However, we will discuss a variety of ways, including mean (or total) pooling. This section will conclude with a demonstration utilising a visual interactive tool that will undoubtedly clarify the entire subject.
Step 3: Smoothing
This section will summarise the flattening process and demonstrate how we transition from pooling to flattened layers while working with Convolutional Neural Networks.

Step 4: Complete Connectivity
Everything that we discussed in the previous section will be consolidated in this section. By understanding this, you'll be able to envision a more complete picture of how Convolutional Neural Networks operate and how the resulting "neurons" learn to classify pictures.

Below figures shows detection of emotions using our proposed method

5.Conclussion:

When it comes to communication, facial emotion plays a crucial part, and so finding the appropriate expression is just as important as knowing what is being said. This project provides a method for distinguishing the category of face emotion, which is defined as follows: Achieving good face detection and emotion extraction from facial photos has been accomplished, and this technology is beneficial in a variety of applications, including robots vision, video surveillance, digital cameras, security, and human-computer interface. Face recognition and emotion recognition are the goals of this project, which will use computer vision to accomplish facial recognition and emotion identification while also improving advanced feature extraction and classification in face expression recognition. This research investigates the topic of face emotion analysis, namely the recognition and detection of emotions. A convolution neural network is described for the purpose of classifying face pictures into the seven regular emotions of happiness, fear, sorrow, anger, surprise, disgust, and neutrality. The seven regular emotions are: happiness, fear, sadness, anger, surprise, disgust, and neutral. Because of its extensiveness and resilience, the fer2013 dataset has been used for both training and testing purposes in this study.

References:

[1]Chu, William Wei-Jen Tsai, Hui-Chuan, YuhMin Chen and Min-Ju Liao. Facial expression recognition with transition detection for students with high-functioning autism in adaptive e-learning.” Soft Computing: 1-27, 2017.

[2] Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Panagiotis, Bjrn W. Schuller, and Stefanos Zafeiriou. ”End-to-end multi-modal emotion recousing neural networks.” IEEE Journal of Topics in Signal Processing 11, no. 8: 13011309, 2017

[3] Gampala, V., Kumar, M.S., Sushama, C. and Raj, E.F.I., 2020. Deep learning based image processing approaches for image deblurring. *Materials Today: Proceedings*.

[4] Aneja, Gary Faigin, Deepali, Alex Colburn, Barbara Mones, and Linda Shapiro. ”Modeling stylized character expressions via deep learning.” In Asian Conference on Computer Vision, pp. 136-153. Springer, 2016

[5] Natarajan, V.A., Kumar, M.S., Patan, R., Kallam, S. and Mohamed, M.Y.N., 2020, September. Segmentation of Nuclei in Histopathology images using Fully Convolutional Deep Neural Architecture. In *2020 International Conference on Computing and Information Technology (ICCIT-1441)* (pp. 1-7). IEEE.

[6] Mohammed, M. A., Abdulkareem, K. H., Mostafa, S. A., Khanapi Abd Ghani, M., Maashi, M. S., Garcia-Zapirain, B., ... & Al-Dhief, F. T. (2020). Voice pathology detection and classification using convolutional neural network model. Applied Sciences, 10(11), 3723.

[7] Korla, S.; Chilukuri, S. T-Move: A Light-Weight Protocol for Improved QoS in Content-Centric Networks with Producer Mobility. Future Internet 2019, 11, 28. https://doi.org/10.3390/fi11020028

[8] Kumar, P. M., Gandhi, U., Varatharajan, R., Manogaran, G., Jidhesh, R., & Vadivel, T. (2019). Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things. Cluster Computing, 22(4), 7733-7744.

[9] Liang, L., Lang, C., Li, Y., Feng, S., & Zhao, J. (2020). Fine-grained facial expression recognition in the wild. IEEE Transactions on Information Forensics and Security, 16, 482-494.

[10] IEEE Style Citation: Korla Swaroopa, Sireesha Rodda, Shanti Chilukuri, “Differentiated Caching for Improved QoS in Vehicular Content-centric Networks,” International Journal of Computer Sciences and Engineering, Vol.6, Issue.10, pp.317-322, 2018.

[11] Shao, J., & Qian, Y. (2019). Three convolutional neural network models for facial expression recognition in the wild. Neurocomputing, 355, 82-92.