Facial Expression Recognition using Deep Learning

Achuthan Babu V G[#1] , Sureshkumar A[#2], Suresh Babu P[#3]
[1 2] Student, [3]Assistant Professor, Department of Information Technology, Velammal College of Engineering and Technology, Madurai, Tamilnadu, India.

Abstract
Facial Expression Recognition is one of the challenging problem in computer vision. It is a tedious process in Machine learning because each person shows their expression in unique way. The Deep Learning Algorithms are used for recognize the seven basic expressions of the humans Anger, Sad, Happy, Scared, Suprise, Disgust, Neutral. In this paper, we use the Convolutional Neural Networks(CNN). Convolutional Neural Networks are the mostly used method for overcoming the difficulties during the feature extraction of the Facial Expression Recognition. Here Visual Geometry Group(VGG) model is used for the construction of CNN. For the evaluation we use the FER2013 database. KAZE feature parameters are used for the feature extraction from the images.

Keywords - *Facial Expression; Recognition; CNN; Architecture; FER2103; KAZE features.*

I. INTRODUCTION

"70% of communications are done through the emotions"Facial Expression is the powerful and natural way for expressing the feeling of a person. Numerous Research and Applications are done for the recognition of the human expression in the field of medical, finance, online tutoring etc,. There may be various facial recognition system are used in the field of computer vision and machine learning to encode the expression information from facial representation.. The six basic expression of human are Happy, Sad, Anger, Disgust, Fear, Surprise and the Neutral . Recently, advanced research on psychology argued that the model of six basic emotions are culture-specific and not universal.

Facial Expressions have the universal meaning and the epressions can be used for the ten to hundreds of years. So nowadays many databases are created for the facial datasets. The Three major stages of facial expression recognition are face acquisition, facial feature extraction Facial Expression Recognition usually employs a three stage training consist of face Acquisition , facial feature extraction and classifier construction. Mostly feature extraction and the classifier construction are used in many works.

II. LITERATURE SURVEY

A. Facial Expression using Support Vector Machine

SVM[5] is a supervised learning technique with associated learning algorithms that analyze data for classification or regression. The SVM training algorithm builds a model for a given training data, which will assign to one category or another for new data, making it a non-probabilistic binary linear classifier. When SVM was introduced, it used for two category classification only. A multi-class classification is a combination of two or more two-category classifications. With a support vector machine, the gap between classes will be maximized as well as the accuracy of classification is also improved. SVM can solve the problems of an inadequate sample of FER and large variance of capacity between different expressions. FER comes under nonlinear classification problem. SVM may use linear algebra and geometry to separate input data into a high dimensional feature space through a selected nonlinear mapping function. This nonlinear mapping function is nothing but kernel function and a learning algorithm is formed to use the kernel functions. Kernel functions include linear, polynomial, RBF and sigmoid. In this paper, RBF kernel function is used. It uses two threads which are one-to-many and one-to-one. SVM generates n different classifiers for n different classes. One-to-one thread selects two different classes as one SVM classifier. Then it will generate $n \times (n - 1)/2$ SVM sub-classifiers. FER comes to multi-class classification issue.

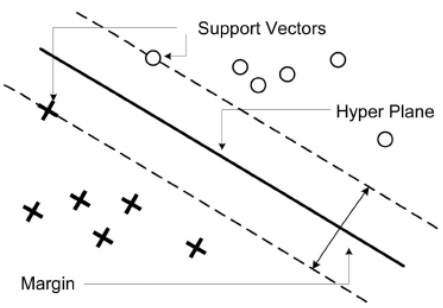

Fig 1: Support Vector Machine Model

B. Facial Expression using Deep Belief Networks

DBNs[6][7] are probabilistic generative models composed of hidden stochastic variables and are similar in structure to neural nets. However, unlike other neural nets, DBNs perform learning one layer at a time, computing generative weight matrices that define connections between the nodes of every two adjacent layers. Once the weighted matrices are calculated, the hidden variables at each layer can be inferred from the visible input by reversing the matrices. The two problems are actually very similar, since both involve training a DBN on a raw image and sorting the training examples into 5-10 classes. We hoped that, as in the digit recognition case, the DBN algorithm's ability to identify complex patterns in the input would yield high accuracy in our classification task.

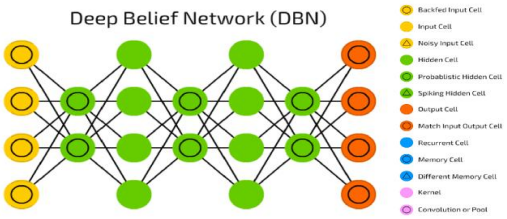

Fig 2: Deep Belief Network Model

III. PROPOSED SYSTEM

While using the SVM and Deep Belief Networks the efficiency of the result is lower. So , we use the Convolutional Neural Networks for the Facial Expression Recognition System.

A. Convolutional Neural Networks

A simple ConvNet is a sequence of layers, and every layer of a ConvNet transforms one volume of activations to another through a differentiable function. We use three main types of layers to build ConvNet architectures [1]: Convolutional Layer, Pooling Layer, and Fully-Connected Layer (exactly as seen in regular Neural Networks). We will stack these layers to form a full ConvNet architecture.

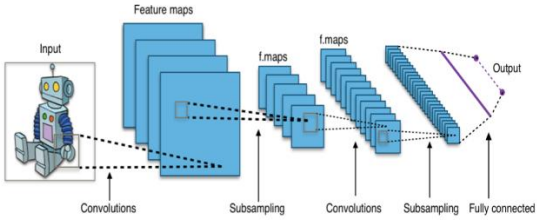

Fig 3: Convolutional Neural Network Model

B. Architecture: VGG-16.

VGG-16[2] represents one of the state of the art architectures for convolutional neural networks, with 16 CNV/FC layers and with an extremely homogenous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end . The VGG model uses more memory and parameters and these parameters are located in first fully connnected layer. It is more expensive for the evaluation. Like a linear classifier, convolutional neural networks have learnable weights and biases; however, in a CNN it is not possible for all the images model at a once, it contains many convolutional layers of weights and biases, and between the convolutional layers there are combination of nonlinear functions that allow the model to approximate much more complicated functions than a linear classifier.

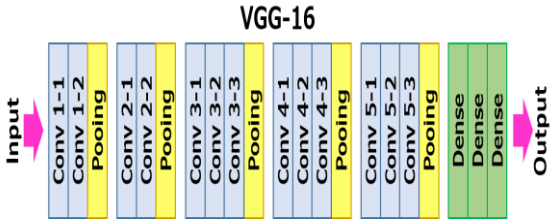

Fig 4 : VGG-16 architecture diagram.

The 48 x 48 image is the input to the VGG model. Subtracting the mean RGB value from the pixel is the only pre-processing in this model. The images is passed through a series of convolutional layers. We use 3x3 filters for each convolutional layers. For linear transformation of input channel we use the 1x1 convolutional filters . The convolution stride is fixed to 1 pixel; The various spatial resolutions is preserved after the convolutional process (i.e. the padding is 1 pixel for 3 × 3 conv. layers). Spatial pooling is done by five sextion of max-pooling layers, which follow some of the convolutional layers (not all the convolutional layers

are followed by max-pooling). Max-pooling is performed over a 2 × 2 pixel window, with stride 2.

A stack of convolutional layers is followed by three Fully-Connected (FC) layers: the first two have 4096 channels each, the third performs 7-way ILSVRC classification and thus contains seven channels (one for each class). The final layer is the softmax layer. The configuration of the fully connected layers is the same in all networks. All hidden layers are equipped with the rectification (ReLU) nonlinearity. To conclude, VGG-16 consists of 16 weight layers that include 13 convolutional layers with filter size of 3x3 and 3 fully-connected layers. The stride and padding of all convolutional layers are fixed to 1 pixel. All convolutional layers are divided into 5 groups and each group is followed by a max-pooling layer. Max-pooling is carried out over a 2x2 window with stride 2. The number of filters of convolutional layer group starts from 64 in the first group and then increases by a factor of 2 after each max-pooling layer, until it reaches 512. We leveraged the keras implementation of VGG-16.

IV. FEATURE EXTRACTION

Feature detection is the process of selecting certain features from the images and used for further processing.. The main objective of feature extraction is to determine the characteristics of the image. Image feature is a simple image pattern, based on which we can describe what we see on the image. Facial feature extraction is the process of extracting face component features[8] like eyes, nose, mouth, etc from human face image. Facial feature extraction is very much important for the initialization of processing techniques like face tracking, facial expression recognition or face recognition. Among all facial features, eye localization and detection is essential, from which locations of all other facial features are identified. The transformation of visual information into the vector space is the major role of features in computer vision. This helpful to perform mathematical operations on them, for example finding similar vector(which lead us to similar image or object on the image). KAZE Features[10] is a novel 2D feature detection and description method that operates completely in a nonlinear scale space. Previous methods such as SIFT or SURF find features in the Gaussian scale space (particular instance of linear diffusion). However, Gaussian blurring does not respect the natural boundaries of objects and smoothes in the same degree details and noise when evolving the original image through the scale space. So in feature extraction we figure out what parts of an image are distinctive, like lines, corners, special patches that can uniquely describe the image. The characteristics are entropy, contrast, co-relation, homogeneity. Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image.

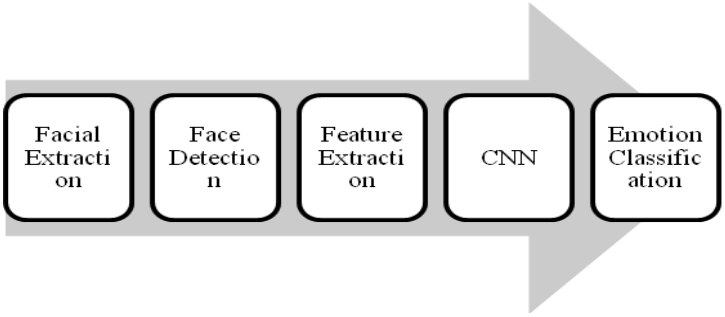

Fig 5: Various Process in Proposed System

V.DATASET

A. FER2013:

FER2013[9] is an open-source .This dataset consists of 35,887 grayscale, 48x48 sized face images with various emotions -7 emotions, all labeled-.Emotion labels in the dataset:0: -4593 image- Angry, 1: -547 images- Disgust, 2: -5121images- Fear, 3: -8989 images- Happy, 4: -6077 images Sad, 5: -4002 images- Surprise, 6: -6198 images- Neutral .

The Dataset is given in CSV file format, the column emotion is used for denoting the type of emotion, second column is denoting the pixel values of images. Image Format: 48 x 48 pixels (8-bit grayscale). Various individuals across the entire spectrum of: ethnicity, race, gender and race, with all these images being taken at various angles. Contains the seven key emotions.

Fig 6: Example images of the seven emotions in the Kaggle dataset.

VI. ANALYSIS

A. Experiments:

For the purpose of this project, we use the simple CNN. This network has the 5 Convolution layers and 3 Pooling layers. It also contains the Normalisation process for the every convolution process. The output of each convolutional layer is normalised and then given to the next convolutional. The trained model will be loaded and the that loaded

model can be checked with the face in video stream from the web camera. Initially the pixel values of the various images from the dataset can be obtained from the csv format. In that the images can be divided into 48 X 48 pixel values.. We will use the various open source libraries for the extraction of the various parameters of the images. Keras library will be used in this project for the implementation of the various algorithms . Keras will be worked with the help of Tensorflow backend for the better optimisation of the classification techniques.

For train the images , we will need to specify the various parameters such as rotation range, height and width shift range for the images, HOG specifications, etc,. For each time the error value will be decreased. Initially the error will be infinity by doing the training the error rate will be decreased and the accuracy of the prediction can be increased. Finally the type of expression can be appear near the face in the streaming video.

B. Results

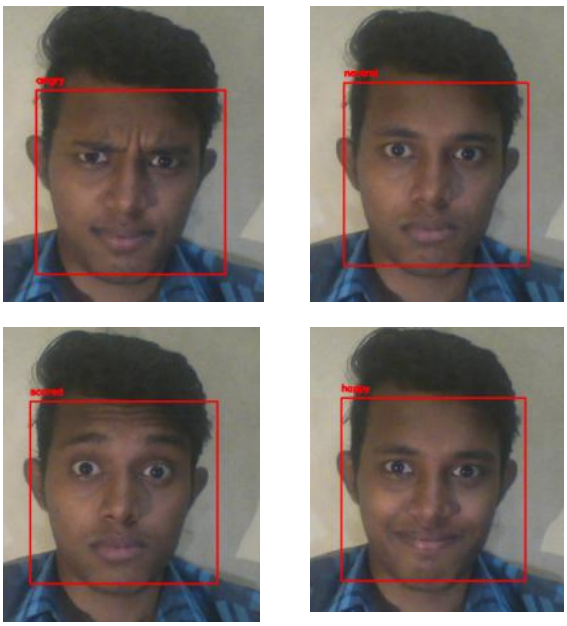

Fig 7: Various Facial Expressions

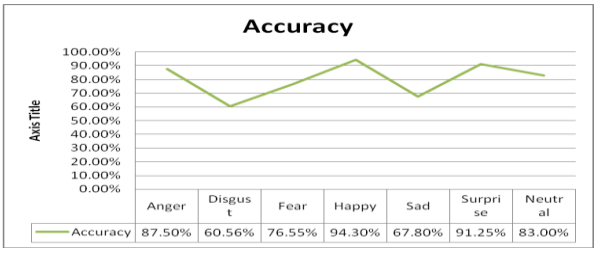

Fig 8: Accuracy for the various Expression

VII. CONCLUSION

We developed CNNs for a facial expression recognition problem and evaluated their performances using different visualization techniques. The results demonstrated that CNNs are capable of learning facial characteristics and improving facial emotion detection. The convolutional networks can intrinsically learn the key facial features by using only raw pixel data. In Future we will improve our project applicable to the color images also.

REFERENCES

[1] Abir Fathallah, Lofti Abdi, Ali Douik., "Facial Expression Recognition via Deep Learning", 2017 IEEE/ACS 14th AICCSA.
[2] Alexandru Savoiu, James Wong, "Recognizing Facial Expression Using Deep Learning. ", Jul 2017.
[3] Raghuvanshi, Vivek Choksi, "Facial Expression Recognition with Convolutional Neural Networks" published on Semantic scholar 2016.
[4] Shan Li , Weihong Deng. "Deep Facial Expression Recognition:Asurvey" published 2018.
[5] Sivaiah Bellamkonda, N.P.Gopalan," A Facial Expression Recognition Model using Support Vector Machines" publsiehd on (IJMSC) April 2018.
[6] Tom McLaughlin, Mai Le, Naran Bayanbat," Emotion Recognition with Deep-Belief Networks", September 2017.
[7] Ping Liu, Shizhong Ha, Zibo Meng, Yan Tong, "Facial Expression Recognition via a Boosted Deep Belief Network" Published in 2014 IEEE Conference on Computer Vision and Pattern Recognition.
[8] Sonali V.Hedaoo , M.D.Katkar , S.P.Khandait, "Feature Tracking and Expression Recognition of Face Using Dynamic Bayesian Network " (IJETT) International Journ al of Engineering Trends and Technology, (ISSN: 2231 -5381) 10 Feb 2014.
[9] Kaggle Dataset. https://www.kaggle.com/deadskull7/fer2013 .