Facial Emotion Recognition Using Artificial Intelligence

Dr.S.Brindha[1], Ms.P.Abirami[2], Mr.S.MOHAMMED ASHIK[3], Mr.K.J.SARAVANAN[4], Mr.N.R.HARISH[5]

[1]*Head of Department, Department of Computer Networking, PSG Polytechnic College, Coimbatore, India*
[2]*Lecturer, Department of Computer Networking, PSG Polytechnic College, Coimbatore, India*
[3]*Student, Department of Computer Networking, PSG Polytechnic College, Coimbatore, India*

----------****----------

Abstract *– Facial expression recognition has been an active research area in the past ten years, with growing application areas including avatar animation, neuromarketing and sociable robots. The recognition of facial expressions is not an easy problem for machine learning methods, since people can vary significantly in the way they show their expressions. Hence, facial expression recognition is still a challenging problem in computer vision. In this work, we propose a simple solution for facial expression recognition that uses a combination of Convolutional Neural Network and specific image pre-processing steps. Convolutional Neural Networks achieve better accuracy with big data.*

Keyword *:- Convolutional Neural Networks, image pre-processing, Emotion detection.*

INDRODUCTION

Nowadays, automated facial expression recognition has a large variety of applications, such as data-driven animation, neuromarketing, interactive games, sociable robotics and many other human-computer interaction systems. Facial expression is one of the most important features of human emotion recognition. Expression recognition is a task that humans perform daily and effortlessly, but it is not yet easily performed by computers, despite recent methods have presented accuracies larger than 95% in some conditions (frontal face, controlled environments, high-resolution images). Facial expression recognition systems can be divided into two main categories: those that work with static images [7, 8, 9, 10, 11, 12, 13] and those that work with dynamic image sequences [14, 15, 16, 17]. Static-based methods do not use temporal information, i.e. the feature vector comprises information about the current input image only. Several facial expression recognition approaches were developed in the last decades with an increasing progress in recognition performance. An important part of this recent progress was achieved thanks to the emergence of Deep Learning methods [7, 10, 12] and more specifically with Convolutional Neural Networks [14, 11], which is one of the deep learning approaches. These approaches became feasible due to: the larger amount of data available nowadays to train learning methods and the advances in GPU technology. The former is crucial for training networks with deep architectures, whereas the latter is crucial for the low cost high-performance numerical computations required for the training procedure. Surveys of the facial expression recognition research can be found in [3, 48].

Fig 1. Basic Facial Expressions

RELATED WORK

Research on facial expression has been conducting for years. But there was always a room for improvement for every research. That is why there are many opportunities regarding this topic. In the main goal of their research is to improve accuracy of a particular dataset FER2013. They applied convolutional neural network as the methodology of their proposed model to classify seven basic emotions. This research demonstrated the success of convolutional neural network to improve the accuracy of biometric applications. However, there exists fluctuation in recognition rate of each class as they could not maintain the equal or nearly equal recognition rate for each class. Although overall accuracy has been achieved at 91.12%, recognition rate in classifying disgust and fear only stands at 45% and 41% respectively. In [6] earlier before this, researchers had developed facial expression recognition system based on posed images in static environment. However, [6] introduced a facial expression dataset named RAF-DB that consists of images of different ages and poses in dynamic environment. They applied deep locality preserving CNN method to classify 7 basic emotions. Their proposed model was trained based on RAF-DB and

CK+ datasets. Although 95.78% of accuracy had been achieved, recognition rate in disgust and fear only stands at 62.16% and 51.25% respectively. As their system is based on two datasets, it is biased to those datasets. In addition, they could not classify emotions from the image that carry geometrically displaced faces. In [8] they have compared two types of facial feature extraction methods. One is geometric positions of fiducial points and the other is Gabor-wavelet coefficient fetch method. As a result of this comparison, they have shown that the gaborwavelet coefficient fetch method performs better than the other one. They have succeeded to find out the number of hidden layers required which is five to seven in order to achieve higher recognition rate. Accuracy has been achieved at 90.1%. Although they have not shown individual class recognition rate, admitted having less recognition rate in fear class.

METHODOLOGY

Convolutional Neural Network is considered as the methodology that is used with data augmentation in this research. Dataset that is used in this research has variation as data was collected from different datasets. As a result, the proposed model is not biased to any particular dataset. The event flow chart of this system is illustrated in fig. 1. In figure 1, at first, the model takes an image from the dataset and detects face from the image by Cascade Classifier. If face is found, then it is sent for pre-processing. Data have been augmented by Image Data Generator function offered by the Keras API. At last, the augmented dataset is fed into CNN in order to predict the class.

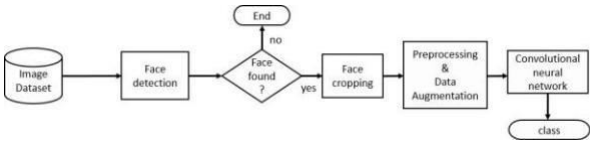

Fig. 2. System Flow Chart

The model that is used to classify the facial expression contains 3 convolution layers with 32, 64 and 128 filters respectively and the kernel size is 3x3. Convolution over an image f(x, y) using a filter w(x, y) is defined in equation (1):

$$w(x,y) * f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x-s, y-t) \quad (1)$$

The activation function that has been used in convolution layer is Relu activation function. Relu is applied to introduce the non-linearity of a model [9] and it is shown in equation (2):

$$f(x) = max(0, x) \quad (2)$$

Model has been provided with 48X48 sized images as model input. The input shape of the model is (48,48,1), where 1 refers to the number of channels exists in input images. Images have been converted into grayscale that is why the number of channels is 1. After convolution layer, the model has 2*2 pool size pooling layer and max pooling has been chosen. Next, there are four fully connected layers which consist of 750, 850, 850 and 750 nodes respectively. Like convolution layer,

TABLE I

SYSTEM ARCHITECTURE

Model Content	Details
First Convolution Layer	32 filters of size 3x3,ReLU, input size 48x48
First Max Pooling Layer	Pooling Size 2x2
Second Convolution Layer	64 filters of size 3x3, ReLU
Second Max Pooling Layer	Pooling size 2x2
Third Convolution Layer	128 filters of size 3x3, ReLU
Third Max Pooling Layer	Pooling size 2x2
First Fully Connected Layer	750 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Second Fully Connected Layer	850 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Third Fully Connected Layer	850 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Forth Fully Connected Layer	750 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Output Layer	7 nodes for 7 classes, SoftMax
Optimization Function	Stochastic Gradient Descent (SGD)
Learning Rate	0.01
Callback	EarlyStopping,ReduceLROnPlateau, ModelCheckpoint,TensorBoard

Relu activation function has been applied in hidden layers. Right after each hidden layer, a dropout layer has been inserted and the value of dropout has been set to 0.5. It randomly deactivates 50% nodes from the hidden layer to avoid overfitting. At last, the output layer of the model consists of 7 nodes as it has 7 classes. Softmax has been used as activation function in the output layer.

$$Softmax(x) = \frac{e^j}{\sum_i e^i} \quad (3)$$

The overview of the convolutional neural network architecture that has been designed for this research is given in Table I.

THEORETICAL BACKGROUND

Three different types of models are used in this paper. This section details the theoretical background for each of these models.

A. Decision Tree

Decision trees are a supervised learning technique that predicts a value given a set of inputs by ”learning” rules based on a set of training data. To put it simply, it is a massive tree of if-then-else rules. The decision making process starts off at the root of the tree and descends by answering a series of yes/no questions. At the end of this if-then-else chain, it arrives at a single predicted label. This is the output of a decision tree.

B. Feedforward Neural Network

A neural network is a system of algorithms that attempts to identify underlying relationships in a set of data by using a method that mimics the way in which a human brain operates. Neural networks consist of nodes connected to each other through edges. Each connection has a weight and a bias. A weight is the strength of the connection. The greater the weight, the greater impact it will have on the final output. A bias is a minimum threshold which the sum of all the weighted inputs must cross. Neural networks are primarily employed for classification tasks. Neural networks consist of three layers - input, output and hidden layers. Hidden layers are sets of features based on the previous layer. They are intermediate layers in the network. A neural network recognises objects based on the concept of learning. Learning consists of six steps. Initially weights are initialised and a batch of data is fetched. This data is known as training data. A forward propagation is done on the data by passing through the network. A metric of difference between expected output and actual output is computed by the use of activation functions which perform computations on the data based on standard mathematical distributions such as Hyperbolic tangent and Sigmoid. This is known as cost. The goal is to minimize or reduce the cost. For this purpose, gradients of cost and weight are backpropagated to know how to adjust the weights to reduce the cost. Backpropagation refers to a backward pass of the network. Later, the weights are updated and the whole process is repeated. Feed-forward networks are also termed as multi-layer perceptron’s.

C. Convolutional Neural Network

A Convolutional neural network is a neural network comprised of convolution layers which does computational heavy lifting by performing convolution. Convolution is a mathematical operation on two functions to produce a third function. It is to be noted that the image is not represented as pixels, but as numbers representing the pixel value. In terms of what the computer sees, there will simply just be a matrix of numbers. The convolution operation takes place on these numbers. We utilize both fully-connected layers as well as convolutional layers. In a fully-connected layer, every node is connected to every other neuron. They are the layers used in standard feedforward neural networks. Unlike the fully connected layers, convolutional layers are not connected to every neuron. Connections are made across localized regions. A sliding ”window” is moved across the image. The size of this window is known as the kernel or the filter. They help recognise patterns in the data. For each filter, there are two main properties to consider - padding and stride. Stride represents the step of the convolution operation, that is, the number of pixels the window moves across. Padding is the addition of null pixels to increase the size of an image. Null pixels here refers to pixels with value of 0. If we have a 5x5 image and a window with a 3x3 filter, a stride of 1 and no padding, the output of the convolutional layer will be a 3x3 image. This condensation of a feature map is known as pooling. In this case, ”max pooling” is utilized. Here, the maximum value is taken from each sliding window and is placed in the output matrix. Convolution is very effective in image recognition and classification compared to a feed-forward neural network. This is because convolution allows to reduce the number of parameters in a network and take advantage of spatial locality. Further, convolutional neural networks introduce the concept of pooling to reduce the number of parameters by down sampling. Applications of Convolutional neural networks include image recognition, self-driving cars and robotics. CNN is popularly used with videos, 2D images, spectrograms, Synthetic Aperture Radars.

DATA COLLECTION AND PREPROCESSING

Datasets have been collected from different sources so that the output is not biased towards a particular dataset. Various standard facial datasets are available online:

- CK and CK+ [12]
- FER2013 [13]
- The MUG Facial Expression Database [14]
- KDEF & AKDEF [15]
- KinFaceW-I and II [16]

It is worth mentioning that FER2013 dataset has been modified as it contains many wrongly classified images which causes lower accuracy gain by the previous research that is based on this dataset [5]. Dataset samples of the model are shown in fig 2.
For data preprocessing following steps are considered:

Fig 3. Dataset Samples

a. Face detection and crop
b. Grayscale conversion
c. Image normalization
d. Image augmentation

A. Face detection and crop

The process of face detection which is called Face Registration is a process of detecting face location from an image. OpenCV Cascade classifier [17] has been used to detect face from the images. After detecting the face, the face portion has been cropped out to avoid background complexity so that the model training becomes more efficient.

B. Grayscale conversion

Images have been resized into 48*48 pixels having 3 channels red, green and blue. To reduce the complexity in pixel values, dataset images have been converted into grayscale having only one channel. So it has become pretty much easy for the model to learn.

D. Image normalization

Normalization has been applied to model dataset which is a process that modifies the range of pixel intensity values to a certain limit. It is a process by which contrast or histogram of the images can be stretched so that it enables deep network to analyze the images in a better way.

DATA AUGMENTATION

Keras API facilitates data augmentation process by introducing Image Data Generator function which through several operations can be applied on the existing dataset to generate more new data. As the parameters for Image Data Generator function, five operations have been included which are rotation at a certain angle, shearing, zooming, horizontal flip, rescale. The parameters with respective values are shown in Table II.

TABLE II

DATA AUGMENTATION PARAMETERS

Operation Type	Value
Horizontal Flip	True
Rotation	0.30
Rescale	1./255
Shear	0.20
Zoom	0.20

Before data augmentation, the dataset had a total of 12,040 images. Each class contains around 1720 images. As CNN is a data-driven approach, in order to achieve more improved model performance, it had been decided to enrich the existing dataset with more images. So, some operations like zoom, rotation, shear, flip and position shift in certain position have been applied to the existing dataset so that more new data can be generated. After applying data augmentation to the dataset, around 5160 images per class. Image or data augmentation is being used to improve deep learning in image classification problem [20]. Therefore, including data augmentation techniques in facial expression recognition has been chosen for this research.

Fig. 4. Data Augmentation

During learning process, 80% of the images have been selected for model training and the remaining 20% is for system validation. For more experiment, splitting ratio was changed in a more challenging way to test the performance of the proposed model. 65% of the dataset was selected for training purpose and the remaining 35% was chosen for testing so that it could be verified if the model performs well with larger dataset.

SYSTEM IMPLEMENTATION

The program has been written in python programming language, using the Spyder IDE. The libraries required in this experiment are Keras, Tensorflow [21], numpy, PIL, OpenCV and matplotlib. Tensorflow was used as system backend whereas keras helped the system by providing builtin functions like activation functions, optimizers, layers etc. OpenCV was mainly used for image pre-processing such as face detection (Cascade Classifier), grayscale conversion, image normalization. Data augmentation was performed by keras API. Matplotlib has been used to generate confusion matrix. The saved

Neural Network model has been hosted on a Python Flask server and it allows a user to choose an image from the local device as input. When the user clicks on the "Predict" button the class of the image is shown after execution. Some sample screenshots of the graphical user interface are given in fig 5: Real-time images have been provided as system input so that it can be shown that the model has the ability to predict unseen images accurately. It should be mentioned that whenever a user provides an image as input in the system, it pre-processes the image in the same way when the model has been trained.

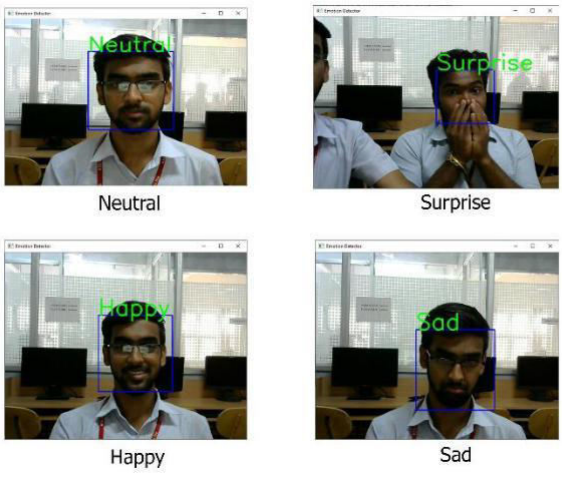

Fig. 5. Real Time Validation

In the future, researchers can try to develop the model more efficiently so that a more standard facial expression recognition system can be delivered. Then with the help of Cascade Classifier [17], the model detects the face from the image. It is mainly the region of interest which is been cropped afterward. As the model has been trained on grayscale images, the system converts the rgb image that contains 3 channels red, green and blue to gray image which consists of only 1 channel. Then to ease the classification task the system has applied image normalization on the image. Then it is sent to the customized Convolutional Neural Network for classification.

CONCLUSION

In this research work, the main agenda was to find out the improvement opportunities for the existing facial expression recognition system. Finding out their limitations and applying probable solutions to overcome the limitations were the objectives of this research. The Convolutional Neural Network method with data augmentation has been proved to be more efficient compared to other machine learning approaches in case of image processing [20]. The proposed model has achieved higher validation accuracy than any other existing model. The Graphical User Interface allows users to do realtime validation of the system. We have considered seven discrete and unique emotion classes (angry, disgust, fear, happy, neutral, sad and surprise) for emotion classification. So, there is no overlapping among classes. However, we are planning to work with compound emotion classes such as surprised with happiness, surprised with anger, sadness with anger, surprised with sadness and so on. In addition, an aggregated view of the facial expression by combining different emotions as well as compound emotions will be determined under uncertainty by using sophisticated methodology like Belief Rule Based Expert Systems (BRBES) in an integrated framework
[22] [23] [24] [25] [26]. As different problems would require different network architectures it is required to figure out which architecture is the best for a particular problem.

REFERENCES

1. performance of automatic facial expression recognition," IT in Business, Industry and Government (CSIBIG), 2014 Conference on, Indore, 2014, pp. 1-7.

2. Mingliang, X., et al. Fully automatic 3D facial expression recognition using local depth features. in Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. 2014.

3. Mohseni, S., N. Zarei, and S. Ramazani. Facial expression recognition using anatomy based facial graph. in Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on. 2014.

4. Ching-Hua, W. and L. Shang-Hong. Online facial expression recognition based on combining texture and geometric information. in Image Processing (ICIP), 2014 IEEE International Conference on. 2014.

5. De, A. and A. Saha. A comparative study on different approaches of real time human emotion recognition based on facial expression detection. in Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in. 2015.

6. Dhavalikar, A.S. and R.K. Kulkarni. Face detection and facial expression recognition system. in Electronics and Communication Systems (ICECS), 2014 International Conference on. 2014.

7. Moeini, A., H. Moeini, and K. Faez. Pose-Invariant Facial Expression Recognition Based on 3D Face Reconstruction and Synthesis from a Single 2D Image. in Pattern Recognition (ICPR), 2014 22nd International Conference on. 2014.

8. Chowdhury, M.I.H. and F.I. Alam. A probabilistic approach to support Self-Organizing Map (SOM) driven facial expression recognition. in Computer and Information Technology (ICCIT), 2014 17th International Conference on. 2014.

9. Jain, S., et al. Significance of facial features in performance of automatic facial expression recognition. in IT in Business, Industry and Government (CSIBIG), 2014 Conference on. 2014.