Emotion Classification of Human Face Expressions using Transfer Learning

S. Poonkodi[1], D. Alen Benard[2], Sornapudi Aditya Vineeth Raj[3], G. Sowmiya[4], V.Geetha[5]

[1,2,3,4] Final Year B. Tech, Dept. of Information Technology, Pondicherry Engineering College, Puducherry, India
[5] Associate Professor, Dept. of Information Technology, Pondicherry Engineering College, Puducherry, India

---***---

Abstract - *Emotive analytics is an interesting fusion of psychology and technology. The ability to understand a person's emotion from face is quite difficult. Moreover, it plays a vital role in non-verbal communication since 90% of our communication is non-verbal. Frequently, the words that the person speaks does not match with how they feel. Different types of emotions are available which have influence on how an individual live and communicate with others. The decisions, actions, and the perceptions that humans have are all influenced by the emotions that people experience at any given moment. According to psychological research, there are six universally accepted facial emotions. They are happiness, sadness, disgust, anger, fear and surprise. In this paper, the emotion is classified with the given test image starting from processing of image and the emotion category is found using transfer learning in which the already trained CNN model is fed and hence the computational time is reduced [1].*

Key Words: **CNN, Emotion detection, facial features, local binary pattern (LBP), Pre-trained network, Transfer learning.**

1. INTRODUCTION

Emotion detection is one of the most important in software environment. Words may lie but face don't. The only way to understand the emotions is by emotion detection which allows to know the mental or emotional state of the person. The emotion is detected by using transfer learning. Transfer learning is a Deep learning method where a particular trained model developed for a task is reused for some events of another task. Transfer learning [6] is a machine learning method where a model developed for a particular task is reused as the starting point for a model on a another task. It is recently extremely popular and attention grabbing in the field of deep learning as a result of this it makes the model to train advanced deep neural networks with relatively little data. With transfer learning, what has been learned in one task to boost generalization in another is tried to exploit primarily. The learning has the tendency to transfer the weights that a network has learned at task A to a replacement task B. Usually, there is a desired heap of data to train a neural network from scratch however there is no access to enough data and training takes more computational time. Transfer learning, is not a replacement methodology which is extremely specific to deep learning. There is a stern distinction between the traditional approach of building and training machine learning models, and employing a methodology following transfer learning principles. Traditional learning is totally different since it occurs purely based on specific tasks, datasets and training separate isolated models on them. No training knowledge is reserved which might be transferred from one model to another different model. In transfer learning, the knowledge can be swayed from previously trained models for training newer models and even tackle computational issues such as having less knowledge learned by the model for the newer task [6]. Deep learning systems and models are layered architectures that learn different new emotion features at various layers of the model. These layers are then finally connected to a last layer which is called fully connected layer. This efficient and layered architecture allows us to utilize a pre-trained network. This is used by breaking the fully connected layer without its final layer as a fixed feature extractor for emotion detection. The next step is to fine-tune the pre-trained model. This is a more involved technique, where not replacing the final layer for classification, but also selectively retrain some of the previous layers to add any new features while re-training.

2. LITERATURE SURVEY

2.1 Leveraging Unlabeled Data for Emotion Recognition with Enhanced Collaborative Semi-Supervised Learning

Zixing Zhang [9] proposed the system that given a small set of labeled data and a large set of unlabeled data, the classifier at first trains the labeled data and recognizes the unlabeled data with confident samples selected via entropy (Semi-supervised learning). In each iteration, same number of samples per class is permitted. Then it recognizes and train the unlabeled data with the help of labeled data which is called self-learning. After a model finishes self-learning, training is done between the classifier namely SVM and RNN to learn the strength from each other and to avoid weakness. This process is called co-training (Collaborative semi-supervised learning). After co-training for each iterations, merging is done with the recognized confident data

(Enhanced Collaborative SSL). In case of draw, the least entropy value of the classifier is taken for merging.

Limitation: Since this system uses semi-supervised learning, training the data is difficult because the system has to recognize the unlabeled data and also train it.

2.2 Grey Wolf optimization-based feature selection and classification for facial emotion recognition

Ninu preetha Nirmala sreedharan [4] proposed the system that the given input images are initially pre-processed and the facial features from the features are extracted with a help of a feature detection algorithm namely scale-invariant feature transform. Then the extracted features is sent through the grey wolf optimization neural network and the necessary emotion types are classified. The test image is fed into the system necessary facial features are selected using viola jones face object detection algorithm. Those selected features are compared with the extracted trained features.

Limitation: Even it gives efficient overall performance. It attains lesser value in recognizing the features smile and angry compared to conventional methods.

2.3 Facial Expression Recognition Based on Cognitive and Mapped Binary Patterns

Chao qi [3] proposed a system that the facial expression recognition is classified into two steps. First is the process of extraction of features and emotion classification. The experimental dataset used here is Cohn-Kanade Dataset with 150 images foe each expression. The facial contours are extracted using local binary pattern (LBP) operator. Then pseudo 3D model is generated to segment the facial area into six facial expression sub-regions. Then these regions and global images are used in the process of feature extraction and for classification support vector machine and softmax with two type of emotion model.

Limitation: Produce long histograms hence recognition speed is low on large datasets.

2.4 Emotion recognition from facial expressions using hybrid feature descriptors

Tehmina kalsum [8] proposed a system that the emotion detection is divided into training and testing. In Training, the necessary features from the face is extracted and a codebook is constructed based on that facial features. A codebook is a cluster of gathered necessary facial features. The codebook is constructed using spatial bag of features (SBoFs) and spatial scale-invariant feature transform (SSIFT). Then the constructed codebook set is fed into the classifier for classification. During testing, the image given is pre-processed and the facial features are extracted. These extracted facial features are mapper with already trained codebook if a nearest match is detected then the classifier appropriate emotions. For classification support vector machine and K nearest neighbor algorithms are used.

Limitation: The accuracy decreases for the dataset with heavy and noisy images.

2.5 Predicting Personalized Image Emotion Perceptions in Social Networks

Sicheng zhao [5] proposed to predict the personalized emotion perception of images of each individual viewer. The dataset used for examination is IAPS dataset, abstract dataset and emotion dataset. This dataset consists of images depicting complex scenes such as portraits, animals, landscapes etc. from each emotion category 200 images were chosen with associated titles, tags and description. The images taken from the social networks and detected for any sudden change in the emotion. Initially Hybrid hypergraph is constructed and Compound vertex generation takes place. Using this the hyper-edge is constructed and given to the trained rolling Multi-task hypergraph learning to give personalized emotions.

Limitation: The emotion is detected only for social network images and not for facial images. As this only keep tracks of emotion storyline and gives the emotion turning point.

2.6 Facial expression recognition using weighted mixture deep neural network based on double-channel facial images

Biao yang [2] proposed to detect the various categories of image from images using convolution neural network. The dataset used here are Cohn-Kanade Dataset, Japanese Female Facial Expression (JAFFE) and Oulu-CASIA. This undergoes a process of data pre-processing and augmentation process. Then the local binary pattern is calculated. This gives the textural information about the given human face image. Then the image is converted into grayscale and the facial features are extracted from the image. The facial features are also extracted from the LBP also. Then the extracted features are fused to form a fine-tuned trained model. Using the fused model the emotions are detected.

Limitation: The computational time is very high because of the training of convolutional neural networks.

2.7 Automatic facial expression recognition using features of salient facial patches

S L Happy [7] proposed a system to detect the type of human face emotion by using the salient facial features which mainly includes eyes, nose, lips and eyebrow which plays a great role in making the particular emotion. The pre-processing of facial images is made and the facial features are extracted from the respective ROIs. The Location of the above features are mentioned via the coordinates. The extracted features are then combined to form a set of active facial patches. The necessary facial patches for each expression is selected and trained by constructing a trained emotion model. During testing, the input image is tested with the selected facial patches of the image and the facial patches which closely matches with the trained features will be the resultant face expression or emotion.

Limitation: LBP is used to extract the facial feature which results in production of long histograms.

3. OVERVIEW OF THE PROPOSED SYSTEM

3.1 Problem Definition

There is a need to extract the necessary feature from the available large data. Once it exceeds the limit, it becomes very difficult to handle the data. There is a difficulty of building a trained model from scratch and hence it needs a lot of computational time and the decision boundary might be over-trained. There is a need to leave the training algorithm for a long period of time to run before obtaining a good decision boundary model. The main objective is to reduce the computational time that occurs because of the data training.

3.2 System Model

The objective of the proposed system is to overcome the limitations of the existing system by eliminating the process of training the model classifier from scratch. That means the features that are new to the model are collected and re-trained. After retraining, the pre-trained and retrained models are combined to form a collection of trained features. Hence the dataset can only be used retrain for new features and to test it. The dataset used here is FER 2013.This involves testing the data against Transfer learning technique which involves a pre-trained CNN (residual network) that is already trained with millions of image types. Because of its already trained quality, during the process of retraining it skips the layer that are closely familiar. The architecture diagram is given in the below figure 1

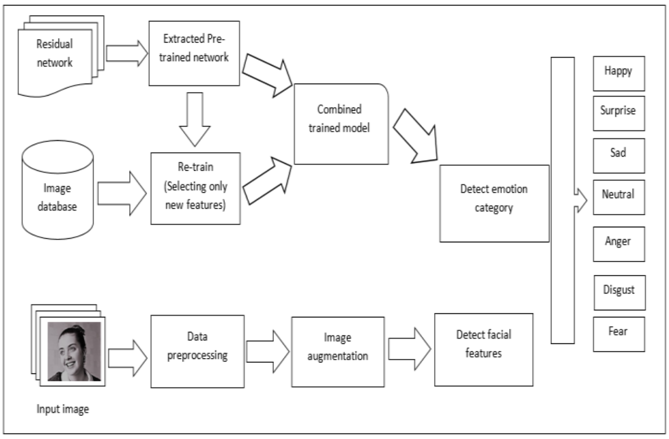

Fig -1 Architecture Diagram for Emotion classification of human face images

With the help of this the computational time because of the training gets reduced and also the emotion of human face images is easily classified.

3.3 Pre-trained network

Residual networks or pre-trained network which is a category of advanced convolution neural networks is used to classify the emotions in the image by passing the input into each chunk of layer straight through to the next piece of data, along with the residual output of the chunk without the input to the chunk that is reintroduced, helped eliminate much of this disappearing signal problem. Thus reducing the computational time in training.

Deep convolutional neural nets have a layered structure and each layers is consisted of convolutional filters [1]. By convolving these filters with the input image, feature vectors for the next layer are produced and through sharing parameters along with the pre-trained model, they can be learnt quite easily. Early layers in convolutional neural networks represent low level local features such as edges and color contrasts while deeper layers try to capture more complex shapes and are more specific. One can improve the classification performance of CNNs by enriching the diversity and specificity of these convolutional filters through deepening the network [1]. Although deep networks can have better performance in classification most of the times, they are harder to train mainly due to two reasons:

- Vanishing / exploding gradients: sometimes a neuron dies during the process of training and based on its activation, the function might never come back. This problem can be solved with initialization techniques that try to start the optimization process with an active set of neurons.

- Harder optimization: when the model introduces more parameters, it becomes more difficult to train the network. This is not simply an overfitting problem, since sometimes adding more layers leads to even more training errors.

Therefore deep CNNs, despite of having better classification performance, are harder to train. One effective way to solve these problems is Residual Networks. The main difference in resnet is that they have shortcut connections parallel to their normal convolutional layers. Unlike convolution layers, these shortcut connections are always alive and the gradients can easily back propagate through them, which results in a faster training.

3.4 Modules Description

3.4.1 Data preprocessing

In this module, each image is segmented to make the image in the correct aspect ratio. Mostly some of the images can have pre-determined aspect ratio. Take the maximum values for each pixel across all training data and normalize the image using the maximum pixel value. Data augmentation process such as scaling and other affine transformations are done. This is made to expose the neural network to a wide variety of variations. This makes the neural network, eliminating the unwanted characteristics in the dataset.

3.4.2 Use bottleneck features of pre-trained CNN

In this module, the fully connected network of pre-trained network is broken into several chunks without any dense layer and select the necessary facial features and save it and set it as the input layer. Along with this set a maximum pooling layer which acts as a building block to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network.

3.4.3 Re-train based on the new featured data

Since the CNN trained model is already fed into this system, there is no need to train the data from the scratch. The training will be done only for adding the new facial features into the system. The bottleneck features that are newly added to the fully connected network are once again trained. Then the trained data is divided into training and validation sets.

3.4.4 Combine both the models

Combine both the pre-trained and re-trained models so that it can have the combination of general and abstract features. Match the image domain which contains the different types of emotions of a human face and detect the emotions based on the fine-tuned learning model.

3.5 Input and Output

Input – The test images to classify the emotion category of an image is shown in the Figure 2 are given as input.

Fig -2 Human face images with various emotions

Output – The output image consists of the input image along with marked facial landmarks and the corresponding emotion that the given image belongs to. The output is given in the figure 3

4. EXPERIMENTAL RESULTS AND EVALUATION

4.1 Simulation Environment

The implementation of the proposed system was carried out using Python version 3.6.5 software running on a personal computer with a 2.07 GHz Intel (R) Core (TM) I3 CPU, 4 GB RAM and Windows 10 as the operating system.

4.2 Emotion Analysis

In this model, the distribution of training data is given in which the data is split into 10 + 10 + 80 for test, validation and training which is given in the Chart 1.

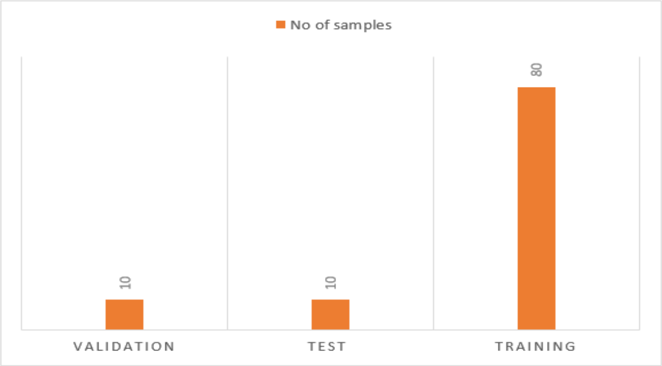

Chart -1 Data distribution

Test is conducted with a variety of available input images which contains a various types of emotion and the output is observed. Since there is no need to spend a large amount of time for training the data the result of the input image is detected with the appropriate emotion with less computational time.

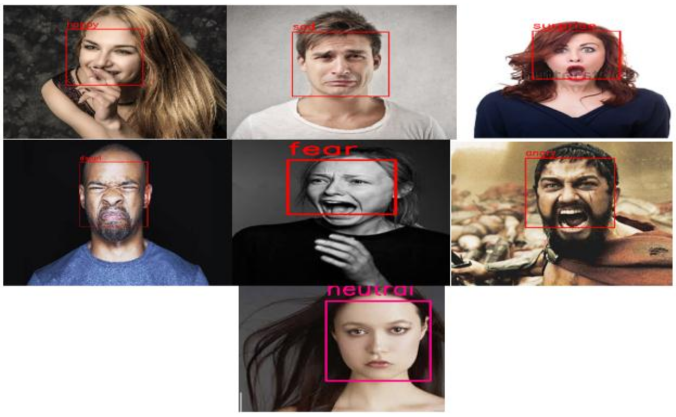

Fig -3 Output of the test image along with the respective detected emotions.

The training history is given in the Chart 2.

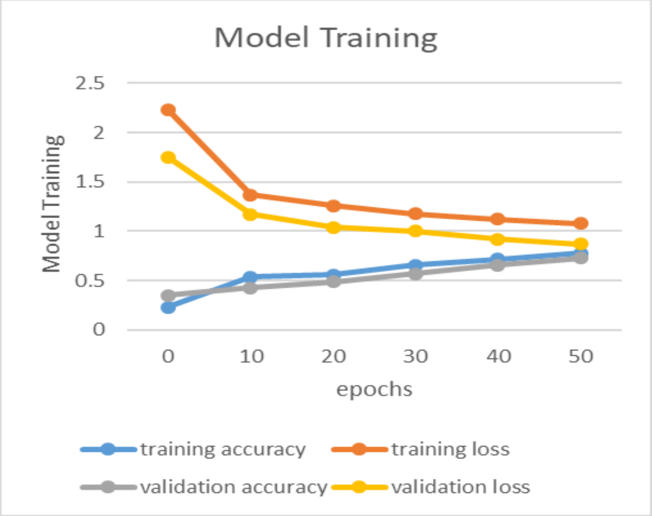

Chart -2 Emotion model training

5. CONCLUSION AND FUTURE ENHANCEMENT

In this paper, emotion classification of human face images has been implemented using transfer learning technique along with the face emotion recognition dataset FER 2013. Using the pre-trained convolution neural networks, features are extracted, re-trained and tested against the test image. The emotions are classified into happiness, sadness, disgust, fear, anger, surprise and neutral which are the universally accepted emotions. This made the emotion detection efficiently with less computational time. This helps in knowing the emotional state of a person.

In the future work, we are looking forward to try applying the proposed method for both video emotion recognition and speech emotion recognition by using the same transfer learning in which the computational cost is reduced by using an already trained emotion detection model.

REFERENCES

[1] Arash Ardakani, Carlo Condo, Mehdi Ahmadi, Warren J. Gross,"An Architecture to Accelerate Convolution in Deep Neural Networks", IEEE Transactions On Circuits and Systems, Volume 5, Issue 4, Page s: 1349 – 1362, October 2017.

[2] Biao Yang , Jinmeng Cao , Rongrong Ni , Yuyu Zhang, "Facial expression recognition using weighted mixture deep neural network based on double channel facial images", IEEE Journals and Magazines, Volume 6, Page s: 4630-4640, February 2018.

[3] Chao Qi, Min Li, Qiushi Wang, Huiquan Zhang, Jinling Xing, Zhifan Gao, Huailing Zhang. "Facial Expression Recognition Based on Cognitive and Mapped Binary Patterns", IEEE Journals and Magazines, Volume 6, Page 18795-18803, April 2018.

[4] NinuPreetha Nirmala Sreedharan, Brammya Ganesan, RamyaRaveendran, Praveena Sarala, Praveena Sarala, RajakumarBoothalingam R. "Grey Wolf optimisation-based feature selection and classification for facial emotion recognition", "IET Biometrics", Volume 7, Issue 5, Page 490-499, March 2018.

[5] Sicheng Zhao, Hongxun Yao, Yue Gao, Guiguang Ding, Tat-Seng Chua. "Predicting Personalized Image Emotion Perceptions in Social Networks", IEEE Journals and Magazines, Volume 9, Issue 4, Page 526-540, October 2018.

[6] SinnoJialin Pan and Qiang Yang, "A Survey on Transfer Learning",IEEE Transactions on Knowledge And Data Engineering, volume 22, Issue 10, Pages: 1345-1359, October 2010.

[7] S L Happy, Aurobinda Routray, "Automatic facial expression recognition using features of salient facial patches", IEEE Journals and Magazines, Volume 6, Issue 1, Page s: 1- 12, March 2015.

[8] Tehmina Kalsum, Syed Muhammad Anwar, Muhammad Majid, Bilal Khan, Sahibzada Muhammad Ali. "Emotion recognition from facial expressions using hybrid feature descriptors", IET Journals and Magazines, Volume 12, Issue 6, Page 1004-1012, February 2018.

[9] Zixing Zhang, Jing Han, Jun Deng, Xinzhou Xu, Fabien Ringeval, Bjorn Schuller. "Leveraging Unlabeled Data for EmotionRecognition With Enhanced CollaborativeSemi-Supervised Learning", IEEE Journals and Magazines, Volume 6, May 2018.