Real Time Emotion Classifier

P.V. Sridhar
Information Technology
Velammal College of Engineering and Technology

P. Madasamy
Information Technology
Velammal College of Engineering and Technology

M. Senthil Kumar
Information Technology
Velammal College of Engineering and Technology

S. Jegadeesan
Information Technology
Velammal College of Engineering and Technology

***Abstract*—Emotion recognition is a key element in human communication. Emotions play a vital role in people's everyday life. Analysis and recognition of human facial expression from live video stream forms the basis for understanding the human's state of mind. In this paper, we propose the use of Convolution Neural Network and machine learning to classify human emotions. The developed Emotion Recognition System classifies human expressions into six types - happy, sad, angry, disgust, neutral and surprise. Jaffe and Paul Ekman database is used for training the database. This Emotion recognition system is found to be 75% accurate in analyzing the human emotion.**

Keywords*—*Facial Expression, Emotion, Convolution Neural Network

I. INTRODUCTION

Emotions play an important role in people's life. Emotions are defined as a conscious experience which is primarily comprised of various biological reactions, differing mental states, as well as psychophysiological expressions. Emotions can be influenced by both external and internal factors, such as an outward event or personality. Emotion play critical role in rational and intelligent behavior. It is a mental state that does not arise through free will and is often accompanied by physiological changes. These changes need to be monitored as they contain information about different types of emotions which will assist in understanding behaviors. We make decisions based on whether we are happy, sad, angry, or surprised. Of all the nonverbal behaviors-body movements, postures, gaze, voice, etc- the face is probably the most accessible window into the mechanisms which govern our emotional and social lives. Face is a complex multidimensional visual model and for developing a model for face recognition is a difficult task.

Facial Expression Recognition plays a vital role when it comes to developing multi-cultural visual communication systems for emotion translation. Various researches have been done in this field. Ekman (1969), a renowned supporter of the first viewpoint, classified emotions into six basic emotions: anger, fear, happiness, disgust, sadness and surprise. Under the first viewpoint, these basic emotions are considered to be biological fixed—that is, they are experienced by all groups of people in a similar manner, and they may even be experienced by some animals. These basic emotions are the basis for all emotional complex emotions are generally considered to be a more refined version of the six basic emotions which are more subjective and variable. Agrawal et al. (2010) discussed the method of Eigen faces which are calculated by using Principal Component Analysis(PCA). Pushpaja et al. (2012) reviews various techniques of facial expression recognition system using Neural Network Toolbox It presents coding and decoding methodology for face recognition Two procedures are necessary for an automatic expression analysis system: facial feature extraction and facial expression recognition.

II. TRAINING DATASET

The dataset used for training the model is from a Kaggle Facial Expression Recognition Challenge a few years back (FER2013). It comprises a total of **35887 pre-cropped, 48-by-48-pixel grayscale images** of faces each labeled with one of the 7 emotion classes:

anger, disgust, fear, happiness, sadness, surprise, and neutral.

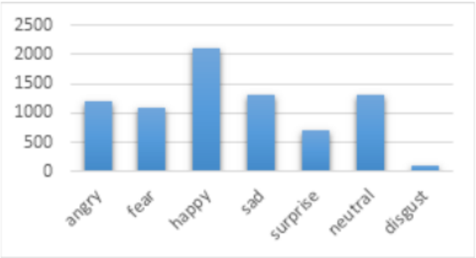

Figure 1. Overview of FER2013 Dataset.

The proposed method uses 28709 labeled faces as the training set and held out the remaining two test sets (3589/set) for after-training validation. The resulting is a **6-class, balanced dataset**, shown in Figure 2, that contains angry, fear, happy, sad, surprise, and neutral. Now we're ready to train.

Figure 2. Training and validation data distribution.

III. PROPOSED MODEL

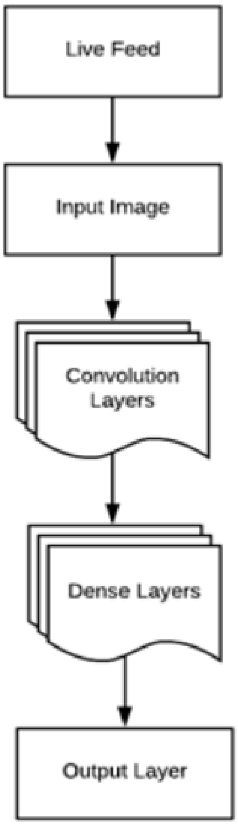

Figure 3. Facial Emotion Recognition CNN Architecture

Deep learning is a popular technique used in computer vision. We chose convolutional neural network (CNN) layers as building blocks to create the model architecture. CNNs are known to imitate how the human brain works when analyzing visuals. A typical architecture of a convolutional neural network will contain an input layer, some convolutional layers, some dense layers (fully-connected layers), and an output layer. These are linearly stacked layers ordered in sequence. In Keras, the model is created as Sequential () and more layers are added to build architecture.

3.1 Input Layer

The input layer has pre-determined, fixed dimensions, so the image must be **pre-processed** before it can be fed into the layer. I used OpenCV, a computer vision library, for face detection in the image. The haar-cascade_frontalface_default.xml in OpenCV contains pre-trained filters and uses Adaboost to quickly find and crop the face. The cropped face is then converted into grayscale using cv2.cvtColor and resized to 48-by-48 pixels with cv2.resize. This step greatly reduces the dimensions compared to the original RGB format with three color dimensions (3, 48, 48). The pipeline ensures every image can be fed into the input layer as a (1, 48, 48) numpy array.

3.2 Convolutional Layers

The numpy array gets passed into the Convolution2D layer where I specify the number of filters as one of the hyper parameters. The **set of filters** (kernel) are unique with randomly generated weights. Each filter, (3, 3) receptive field, slides across the original image with **shared weights** to create a **feature map. Convolution** generates feature maps that represent how pixel values are enhanced, for example, edge and pattern detection. Other filters are applied one after another creating a set of feature maps.

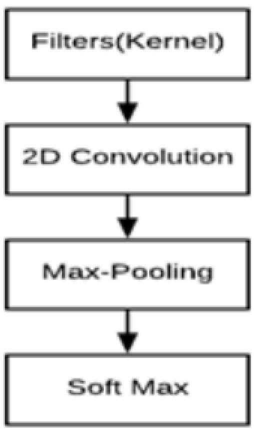

Figure 4. Convolution and 1st max-pooling used in the network

Pooling is a dimension reduction technique usually applied after one or several convolutional layers. It is an important step when building CNNs as adding more convolutional layers can greatly affect computational time. I used a popular pooling method called MaxPooling2D that uses (2, 2) windows across the feature map only keeping the maximum pixel value. The pooled pixels form an image with dimensions reduced by 4.

3.3 Dense Layers

The dense layer (fully connected layers), is inspired by the way neurons transmit signals through the brain. It takes a large number of input features and transform features through layers connected with trainable weights.

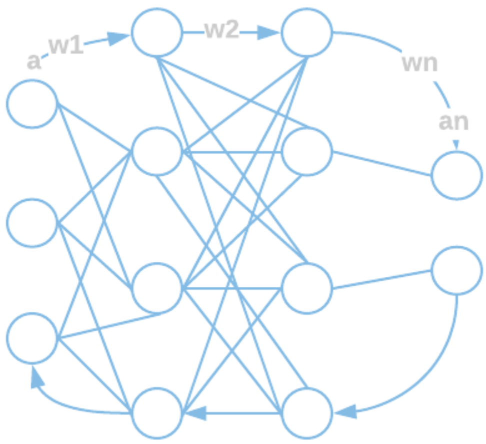

Figure 5. Neural network during training: Forward propagation to Backward propagation.

These weights are trained by forward propagation of training data then backward propagation of its errors. **Back propagation** starts from evaluating the difference between prediction and true value, and back calculates the weight adjustment needed to every layer before. We can control the training speed and the complexity of the architecture by tuning the hyper-parameters, such as **learning rate** and **network density**. As we feed in more data, the network is able to gradually make adjustments until errors are minimized.

Essentially, the more layers/nodes we add to the network the better it can pick up signals. As good as it may sound, the model also becomes increasingly prone to overfitting the training data. One method to prevent overfitting and generalize on unseen data is to apply **dropout**. Dropout randomly selects a portion (usually less than 50%) of nodes to set their weights to zero during training. This method can effectively control the model's sensitivity to noise during training while maintaining the necessary complexity of the architecture.

3.4 Output Layer

Instead of using sigmoid activation function, we used **softmax** at the output layer. This output presents itself as a probability for each emotion class.

Therefore, the model is able to show the detail probability composition of the emotions in the face. As later on, you will see that it is not efficient to classify human facial expression as only a single emotion. Our expressions are usually much complex and contain a mix of emotions that could be used to accurately describe a particular expression.

3.5 Deep Learning

we built a simple CNN with an input, three convolution layers, one dense layer, and an output layer to start with. As it turned out, the simple model performed poorly. The low accuracy of 0.1500 showed that it was merely random guessing one of the six emotions. The simple net architecture failed to pick up the subtle details in facial expressions. This is where deep learning comes in. Given the pattern complexity of facial expressions, it is necessary to build with a deeper architecture in order to identify subtle signals.

IV. RESULT

For the validation of the facial expression recognition system we selected 20 respondents to check weather their response was similar to the results given by the system. As a result, it was observed when the respondents were happy 80% of the times the results were accurate, when the respondents were sad the results were 75% accurate, when the respondents were surprised the results were 80% accurate and when the respondents were angry the results were 70% accurate. So, our facial expression recognition system is overall approx. 76% accurate.

V. CONCLUSION

In this paper, an emotion recognition system was developed which is based on data provided that will recognize the human emotion. The neural network approach is based on face recognition and feature extraction and training is provided to the software to analyses or recognize the emotion. A still image facial extracted from live stream expression recognition technique has been developed. The facial expression recognition system is found to be 75% accurate. For future improvement and development of the system, a real time facial expression recognition system can be developed with accuracy of 80%

VI. REFERENCES

[1] AlMejrad,A.S. (2010).” Human emotions detection using brain wave signals: A challenging”. European Journal of Scientific Research, 44(4), 640-659.

[2] G. Zhao and M. Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” Pattern Analysis and Machine I

[3] Agrawal, N.Jain,M.Kumar and H. Agrawal(2010)”Face Recognition Using Eigen faces and Artificial Neural Network” International Journal of Computer Theory and Engineering.

[4] Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969).” Pan-cultural elements in facial displays of emotion”. Science, 164(3875), 86-88.

[5] Thorsten Joachims. Text categorization with support vector machines: learning with many relevant features. In Claire N´edellec and C´eline Rouveirol, editors, Proceedings of ECML-98, 10th European Conference on Machine Learning, pages 137–142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE.

[6] Pushpaja V. Saudagare and D.S Chaudhari(2012)"Facial Expression Recognition Using Neural network-An Overview" International Journal of Soft Computing and Engineering ISSN:2231-2307,Volume-2,Issue-1.

[7] Sirakaya, E., Petrick, J., & Choi, H. (2004). The role of mood on tourism product evaluations. Annals of Tourism Research, 31, 517– 539.

[8] A. Mehrabian. Communication without words. Psychology Today, 2(4):53–56, 1968.

[9] Saudagare, P. V., & Chaudhari, D. S. (2012). Facial Expression Recognition using Neural Network–An Overview. International Journal of Soft Computing and Engineering (IJSCE), 2(1), 224-227.

[10] J.L.Raheja and U.Kumar(2010)"Human Facial Expression Recognition from Detected in Captured Image Using Back Propagation Neural Network".International Journal of Computer Science and Information Technology.

[11] Savran, A., Gur, R., & Verma, R. (2013, December). Automatic Detection of Emotion Valence on Faces Using Consumer Depth Cameras. In Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on (pp. 75-82). IEEE.