Facial Emotion Detection Using Convolutional Neural Networks

S. Lalitha Priya[1], Dr. K. Sekhar[2], Dr.K.Sirish Kumar[3]
[1, 2, 3] Dept of Computer Science
[1, 2, 3] SRI VENKATESWARA Engineering College,Tirupati

Abstract- *Recognizing human emotions is crucial to interpersonal relationships. One of the most potent and difficult social communication research tasks is emotion identification from an image. Since the beginning of time, researchers have been interested in the automatic recognition of emotions. Consequently, there have been numerous advancements in this area. Speech, hand and body gestures, as well as facial expressions, all serve as vehicles for the expression of emotions. When it comes to image processing, deep learning (DL)-based emotion recognition outperforms classical methods. This paper proposes a face expression identification approach based on a convolutional neural network (CNN) and image edge detection to avoid the complex process of explicit feature extraction in traditional facial expression recognition. The edge of each layer of the image is retrieved in the convolution process after the facial expression image is normalized. To maintain the texture picture's edge structure information, the retrieved edge information is placed on each feature image. The maximum pooling method is then used to reduce the dimensionality of the extracted implicit features. Finally, a Softmax classifier is used to classify and recognize the expression of the test sample image.*

Keywords- Convolutional Neural Networks, Haar-like Features, Image Edges, Softmax Classifier, Emotion detection, Kirsch Edge operator.

I. INTRODUCTION

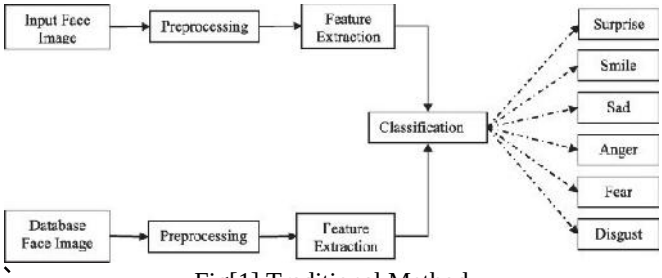
Emotion is a mental state associated with the nervous system associated with feelings, perceptions, behavioral reactions, and a degree of gratification or displeasure. One of the current applications of artificial intelligence (AI) using neural networks is the recognition of faces in images and videos for various applications. The majority of techniques analyze visual data and look for common patterns in human faces in pictures or movies. Law enforcement agencies can employ face detection for crowd control and monitoring. This paper presents a method for identifying seven emotions such as anger, disgust, neutral, fear, happy, sad, and surprise using facial images.

The research of expression recognition in computer field mainly focuses on the feature extraction and feature classification. There are many methods of feature extraction. According to the type of data input, the existing methods of feature extraction can be divided into two categories: one is based on static images and the other is based on a dynamic sequence. Feature extraction methods based on static images include Gabor wavelet transform, Haar wavelet transform, Local Binary Pattern (LBP), and Active Appearance Models (AAM). The dimension of feature is large before and after the completion of feature, and thus the dimension reduction is usually carried out. Commonly used methods of facial expression classification are Hidden Markov Model (HMM), Support Vector Machine (SVM), AdaBoost, and Artificial Neural Networks (ANN). A facial expression recognition method based on CNN and image edge detection is suggested in this study as an alternative to the labor-intensive process of explicit feature extraction used in conventional facial expression recognition. The main innovations of this method are as follows:

1. The edge of each layer of the input image is extracted, and then the extracted edge information is superimposed on each feature image to preserve the edge structure information of the texture image.
2. The maximum pooling method is used to reduce the dimension of the extracted implicit features, which shortens the training time of the convolutional neural network model.

II. EXISTING SYSTEM

The traditional method for emotion detection includes three steps. They are preprocessing, feature extraction and classification.



Fig[1] Traditional Method

1) Preprocessing

Image preprocessing includes different types of processes such as image clarity and scaling, contrast adjustment, and additional enhancement process to improve the expression frames. The face image was cropped and scaled, with the nose of the face acting as the physical midway and the other significant facial features being physically included. The input photos are resized using the Gaussian filter, which gives the image smoothness. The preprocessing technique known as normalization can be used to reduce lighting and fluctuations of the face images with the median filter and to achieve an improved face image. It is also used for the extraction of eye positions which make more robust to personality differences for the FER(Facial Emotion Recognition) system and it provides more clarity to the input images. Localization is a preprocessing method and it uses the Viola-Jones algorithm to detect the facial images from the input image. Adaboost learning algorithm and haar like features are used to detect size and location of facial images. It is mainly used for spotting the size and locations of the face from the image.

2)Feature Extraction

Finding and displaying desirable aspects of interest in an image for subsequent processing is known as feature extraction. Computer vision feature extraction, which marks the transition from graphic to implicit data depiction, is an important step in the processing of images. Then, the data visualization can be used as a classification input. The five different types of feature extraction methods include patch-based methods, global and local feature-based methods, geometric feature-based methods, and methods based on edges and textures.

3)Classification

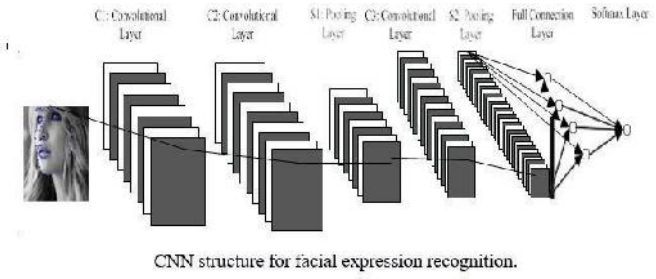
The final stage of the FER system is classification, in which the classifier classifies facial expressions such a grin, sadness, surprise, rage, fear, disgust, and neutral.

One of the categorization methods that uses two different sorts of approaches is Support Vector Machine (SVM). They compete against one other and all other strategies. In one against all classification, one sample is created for each class. SVM is one of the best classification methods for complex dimensionality problems because it uses one versus one classification, which creates a class for each pair of classes. The supervised machine learning technique known as SVM makes use of four different kernel types to improve performance. They are sigmoid, radial basis function (RBF), polynomial, and linear. The high-dimensional data are mapped by the linear kernel, which is linearly separable. The function that converts a single feature into high-dimensional data is used by the RBF kernel. The nonlinear models are learned via the polynomial kernel, which also handles their similarity. The statistical model that classifies phrases into several kinds is called a Hidden Markov Model (HMM) classifier. For classification, Hidden Conditional Random Fields (HCRF) representation is utilised. For better classification performance, it makes use of the full covariance Gaussian distribution.

III. PROPOSED SYSTEM

The difference between the proposed system and the existing system is that the concept of image edge computing is included in this model. The facial emotions are detected and classified using these edge information. Firstly, we locate the face in the image and cut out the face image. Then, we normalize the face image to a specific size. Next, we equalize the histogram of the image to reduce the influence of illumination and other factors. Finally, we extract the edge of each layer of the image in the convolution process.

3.1 PROPOSED ALGORITHM



Fig[2] CNN Structure

There are various algorithms used in this paper like Haar classifier, Softmax Classifier and Kirsch Edge Operator. In addition to these algorithm another important part of this project is the architecture of the Convolutional Neural Networks. The sequence of the layers present in the structure plays the most important role in getting the expected result.

Besides the sequence is what makes this project unique when compared to many of the projects related to Emotion recognition. The sequence of the convolutional architecture implants the process of super imposing the edge information onto the processed facial images repeatedly which minimizes the data loss. The hidden motto that effective results can be drawn with minimal data loss can be achieved using this proposed system.

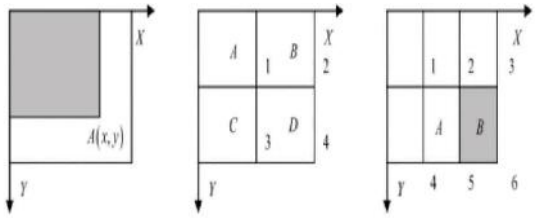
IV. SYSTEM MODULE

1) Face Detection and Location
2) Scale Normalization
3) Gray Level Equalization
4) Image Edge Detection

1)Face Detection and Location

This paper uses a Haar classifier for human detection. The Haar classifier is trained by Haar-like small features (linear, edge, center and diagonal) and an integral graph method combined with the AdaBoost algorithm. This method uses the Haar-like to extract facial features, and uses an integral graph to realize fast calculation of Haar- like features. However, the calculation of eigenvalues is very time-consuming. In order to improve the calculation speed, this paper uses the integral graph method to calculate the Haar like eigenvalues. The integral graph of the coordinate A (x, y) in a graph is defined as sum of all the pixels in its upper left corner.

$$A(x,y)\, ii(x,y) = \sum_{x' \le x, y'} i(x', y')$$



`

Fig[3] Graph Method

Here, ii (x,y) represents the integral image. i (x`,y`) represents the original image for gray image, here represents the gray value and for color image, here represents the color value. The pixel value of an area can be calculated by using the integral graph of the end points of the area, as shown in above image. The pixel value of region D can be calculated by

`

$$S(D) = ii(4) + ii(1) - ii(2) - ii(3)$$

where ii (1) represents the pixel value of region A, ii (2) represents the pixel value of region A + B, ii (3) represents the pixel value of region A + C, ii (4) represents the pixel value of regions A + B + C + D. The eigen values of rectangular features can be calculated by integral graphs of feature endpoints. Taking the edge features as an example, the eigen value calculation can be expressed by Fig 1 (c). The pixel values of point A and point B are:

$$S(A) = ii(5) + ii(1) - ii(2) - ii(4)$$
$$S(B) = ii(6) + ii(2) - ii(5) - ii(3)$$

According to the definition, the eigen value of rectangular feature is the pixel value of region A minus the pixel value of region B. According to formula (3) and formula (4), the formula for calculating eigen value is as follows.

$$T = ii(5) - ii(4) + ii(3) - ii(2) - (ii(2) - ii(1)) - (ii(6) - ii(5))$$

The extracted Haar-like features are used to train the classifier, and the AdaBoost algorithm is used to train the classifier. Finally, the trained classifier is used to extract the face from the image.

2)Scale Normalization

Because the input of the network is a fixed sized picture, the original picture should be normalized to generate a specific size picture. Through normalization, the input image is scaled to 128 X 128 size. Let point (x, y) in the original picture be normalized and mapped to point (x`, y`). The mapping is as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

(a)Before normalization (b)After normalization

Fig[4] Normalization

3)Gray level Equalization

Uneven distribution of light and shade, which will increase the difficulty of feature extraction. Histogram Equalization (HE) method is used to average the gray level to enhance the contrast. It can be simply done by transforming the histogram of the original graph into a uniform distribution form.

If the gray level of the gray image is L, the size is M x N, and the number of pixels in the ri gray level is E, the corresponding probability of gray level occurrence is as follows:

$$P_r(r_i) = \frac{n_i}{M \times N}, \quad i = 0, 1, \cdots, L-1$$

Subsequently, the cumulative distribution function is calculated using the following equation,

$$T(r_i) = \sum_{j=0}^{i} P_r(r_j), \quad i = 0, 1, \cdots, L-1$$

Finally, the image histogram is averaged using the following mapping relations:

$$e_j = INT\left[(e_{\max} - e_{\min})\, T(r) + e_{\min} + 0.5\right] \underset{j=0,1,\cdots,L-1}{}$$

(a)Before gray level equalization (b)After gray level equalization

Fig[4] Grayscale Equalization

4)Image Edge Detection

The edge information of an image is often reflected in the area where the gradient information of the image changes dramatically. The edge of the image gives people a stronger visual sense. Kirsch edge operator is used to extract image edge information whose templates are in eight directions.

$$a_0 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \quad a_1 = \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix},$$

$$a_2 = \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}, \quad a_3 = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix},$$

$$a_4 = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}, \quad a_5 = \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix},$$

$$a_6 = \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}, \quad a_7 = \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

5. LIMITATION OF EXISTING SYSTEM

Variance of Lights

Since each image is taken in the completely different background and lighting conditions, the intra-class noise of lights will distort the model to classify the emotion. As results, the same type of emotions may be classified differently because of the effect of lighting noise.

Variance of Location

Since each image is taken in the completely different background and lighting conditions, the intra-class noise of lights will distort the model to classify the emotion. As results, the same type of emotions may be classified differently because of the effect of lighting noise.

Differences of Mathematical Model

The conventional method is typically performing classification through linear transformation, so they are typically referred as Linear Classifier. In comparison, CNN as well as other Deep Learning algorithm typically combines the linear transformation with nonlinear function such as sigmoid (Logistic Function) and Rectified Linear Unit (ReLU) to distinguish differences in process of classification.

VI. APPLICATIONS

It is used by Law Enforcement Agencies during interrogations to know the mental state of the suspects.
It is used to evaluate the effect of anti-depressants on patients by using Facial Expressions.
It helps to interpret the emotions of autistic children and help doctors understand the psychological changes in them.
It helps to prevent accidents by alerting the drivers during the Fatigue state.

It is also used in the social media and gaming platforms.

VII. CONCLUSION

The suggested method uses pixel values and emotion labels as input to the model. Convolutional Neural Networks encourages autonomous learning that can acquire more abstract picture feature expressions implicitly. The suggested method's training procedure employs proper weight initialization, which has a significant impact on weight updating. To a certain extent, the suggested algorithm can increase the identification rate of face expressions in complex backgrounds. In addition to it, the proposed strategy produces a greater rate of recognition. Noises such as face position, occlusion, and blurring may be present in real-life facial expressions. Pre- processing methods used in this model reduce most of these noises, though some may persist.

VIII. FUTURE ENHANCEMENTS

1) To further finetuning model using grid search, specifically:
   a. Different optimizer such as Adam, RMSprop, Adagrad.
   b. Experimenting dropout with batch-normalization.
   c. Experimenting different dropout rates.

2) To collect more data and train the model with balanced dataset.

REFERENCES

[1] H. Zhang, A. Jolfaei and M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing," in IEEE Access, vol. 7, pp. 159081- 159089, 2019, doi: 10.1109/ACCESS.2019.2949741.

[2] R. M. Mehmood, R. Du, and H. J. Lee, ‘‘Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors,’’ IEEE Access, vol. 5, pp. 14797–14806, 2017.

[3] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, ‘‘MPED: A multi- modal physiological emotion database for discrete emotion recognition,’’ IEEE Access, vol. 7, pp. 12177–12191, 2019.

[4] E. Batbaatar, M. Li, and K. H. Ryu, ‘‘Semantic-emotion neural network for emotion recognition from text,’’ IEEE Access, vol. 7, pp. 111866–111878, 2019.

[5] Y. Zhang, L. Yan, B. Xie, X. Li, and J. Zhu, ‘‘Pupil localization algorithm combining convex area voting and model constraint,’’ Pattern Recognit. Image Anal., vol. 27, no. 4, pp. 846–854, 2017.

[6] H. Meng, N. Bianchi-Berthouze, Y. Deng, J. Cheng, and J. P. Cosmas, ‘‘Time- delay neural network for continuous emotional dimension prediction from facial expression sequences,’’ IEEE Trans. Cybern., vol. 46, no. 4, pp. 916–929, Apr. 2016.