International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

FACIAL EMOTION RECOGNITION USING DEEP LEARNING

Ms. Harini P[1], Mrs. Sangeetha N[2]

[1] III B.Sc. Computer Technology, Department of Computer Technology, Sri Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu
22bsct217harinip@skacas.ac.in
Cell:9566702457.
[2] Assistant Professor, Department of Computer Technology, Sri Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu

ABSTRACT :

An important aspect of human beings is their ability to display emotions. Emotions form the basis of intimate means of communications between one another. Recognition of emotions displayed via facial expression with the help of a computer could prove to be a very powerful too. In recent years, an increased number of intelligent systems are using facial emotion recognition to improve human interaction. These systems cause constant changes in their operation based on the emotion of humans. In this paper, we propose an architecture based on the convolutional neural network (CNN) for the facial recognition of emotions. In the implementation, we use the Facial Expression Recognition 2013 dataset (FER- 2013). Through behavioral analysis, we show how different emotions seem to be sensitive to different parts of the face.

Keywords - Haar Cascade Classifier, Convolutional Neural Network, Softmax Classification, Real-Time Detection

I. INTRODUCTION :

Recognizing facial emotions has become a major issue in many applications today. The research on facial emotion recognition has gained a lot of momentum over the past few years. The state of human emotions is identified using facial emotion recognition (e.g. neutral, happy, sad, surprise, fear, anger, disgust, contempt) based on the stream of images fed in by a video. There has been growing interest in making machines act as close as possible to actual human beings. To make their actions replicate those of humans and add a touch of human feelings in each of these actions. It has been argued that for there to be a proper human-computer interaction, the computer has to interact in a natural way, similar to that when two humans interact. Other fields where emotion recognition can prove to be useful are online teaching, product marketing, health industry, and several others to determining whether certain things are liked or not and what is the reaction of different people towards different stimuli. Humans express their thoughts and feelings in multiple ways. The most important being speech, followed up by their display of emotions. Emotions typically manifest themselves via multiple means, be it touch, visual or physiological. The most appropriate method of detecting human emotions would be to determine their emotion from the visual cues displayed by a human. It is widely accepted that the slightest of change in emotions become visible on the faces of humans and hence developing a system which detects the emotion from an individual's face could prove to be an invaluable tool.
The work done by us in this project enables proposes a system which allows for real-time emotion detection by using a video stream input from the user's webcam. We initially focus on detecting the face from the video using the Haar cascade classifiers. We then make use of convolutional neural networks for determining emotions. The dataset used by us in this project is the FER2013 dataset. The dataset originally was in the form of raw pixels but we had to convert them into actual images and use those images for training our model. This greatly helped in improving the performance of our model. The software used by us in this project is OpenCV and TensorFlow. OpenCV is used for operating the webcam and for face detection, while TensorFlow was used for training the CNN for emotion detection.

II. LITERATURE REVIEW :

The leading research group on emotion recognition is Affective. They have employed various deep learning methodologies as they have a humongous dataset of over 3 million facial videos. This dataset has been gathered from over a total of 75 countries which ensures that geographical factors don't affect their research [1]. The classifier developed by them can classify the emotions of joy, disgust, contempt, and surprise with a ROC (Receiver operating characteristics) score of 0.8.
Work done by Kotsia et al. [2] threw a spotlight on the effect of occlusion during the classification of emotion expressions. To obtain satisfactory results multiple classification models and feature engineering techniques were used. Gabor filter, which is a linear filter primarily used for texture analysis which analyzes whether there is any specific frequency content in the image in specific directions in a localized region. They used the support vector machine (SVM) and the multi-layer perceptron (MLP) for classifying these features.

Wang and Yin studied how the magnitude of facial expressions changed with the change in distortion of detected face region and eventually the change of outcome in their model. They used Topographic context expression descriptors to perform feature extraction. Every image in this research was treated as a 3-d model of the face and the labelling of pixels was done taking into account the terrain features of each image.

Apart from facial emotion detection research has been done to detect emotion by using biological factors. The research paper by Arturo Nakasone [3] proposes a method for the recognition of emotions from bio-signals. They have used GSR(Galvanic skin response) to measure skin conductance and Electromyography(EMG) to measure muscle activity.

As can be seen from the literature review, the detection of emotions is a very complicated task and there have been several different approaches towards solving this problem.

III. METHODOLOGY

A. Tools and Platforms used

The models are developed using Python-3 with the assistance of its libraries and frameworks. OpenCV-3.3.0 was used in our project for Computer Vision and Image Processing, we leveraged TensorFlow for implementing models of Convolutional Neural Networks, Keras for Machine Learning. Mathematical computations are performed using NumPy and Pandas. The models are built on Google Colab with the use of the GPUs provided.

B. Dataset used

There are several datasets available on facial expression. We used the Facial Expression Recognition dataset (FER- 2013). This was published 5 years ago at the International Conference on Machine Learning (ICML). This is an open- source dataset, created by Pierre-Luc Carrier and Aaron Courville and was shared publicly for the Kaggle competition. This dataset contains a training set consisting of 28709 images and a test set consisting of 7178 images. This categorical data consists of 48x48 pixels grayscale images with 7 facial expressions (0: Angry, 1: Disgusted, 2: Fearful, 3:Happy, 4: Neutral, 5:Sad and 6:Surprised).

Image Preprocessing

The aim of pre-processing is to improve the quality of the image. First, we rescale the image to transform every pixel value from the range [0,255] [0,1], which is also referred to as Normalization. This treats high and low pixel images in the same manner (i.e., scales every image to the same range) such that it contributes more evenly to the total loss. The images are then converted into 48x48 pixel gray images to prevent unwanted density within neural networks.

Layers and Activation functions used

2D Convolutional Layer: 2D Convolutional layer is a common type of convolution and is written as conv2d. The most important parameters in this layer are the number and the size of the kernels. It performs feature extraction and transforms a 2D matrix of features into a different 2D matrix of features depending upon the given parameters. Thus, this layer helps in image processing.

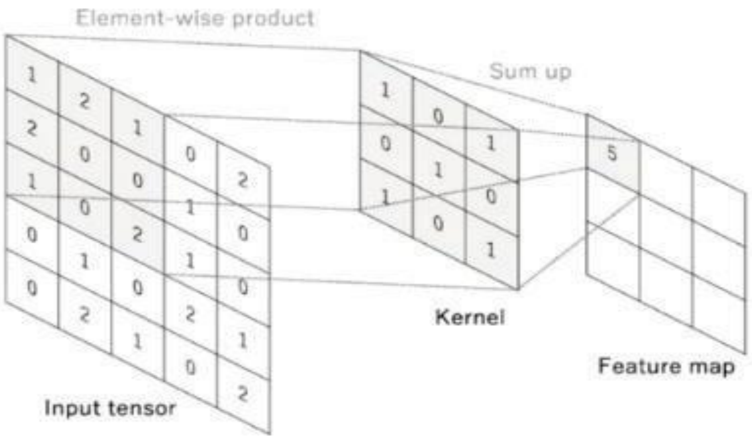

Fig. 2. Example of 2D Convolutional operation

2) Max Pooling Layer: Max Pooling layer operates such that it calculates the maximum value in each patch of each feature map. The layer down samples the input in such a way that it highlights the most present feature in the patch. It reduces the dimensions of the feature map and the computational cost of the network. This works much better for computer vision.

3) Dropout Layer: Dropout means removing data from a neural network to improve the processing and efficiency of results. This mainly removes similar features from the input data in the layer. Thus, the dropout layer is used for a neural network to tackle overfitting. To avoid overfitting, we can also use early stopping in the training phase.

4) Dense Layer: The dense layer is a frequently used layer. Since all the neurons in this layer are connected to the neurons in the next layer, this layer is also called the fully connected layer. It performs matrix-vector multiplications. This is useful for changing the dimensions (like rotating, translation, scaling) of the vector. The below example depicts that a 3-unit input layer is connected to a 5-unit dense layer.

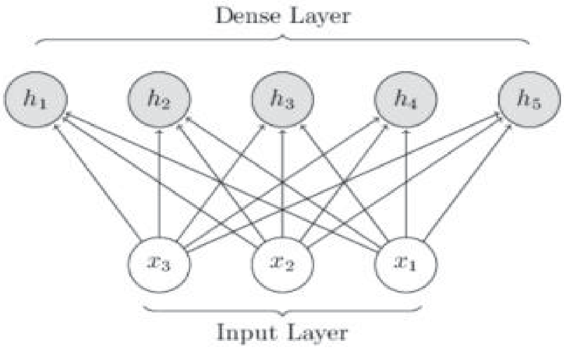

Fig. 5. Example of Dense operation

Flatten Layer: Flatten layer transforms our multi- dimensional input tensor to a 1-D tensor. This means that the pooled feature map is converted into a single column. In flattening, each pixel has its neuron and thus it helps in analyzing every pixel. For example, the previous layer outputs a (4,4,128) tensor. Flattening this tensor gives an input of (2048,1) tensor in the next layer.

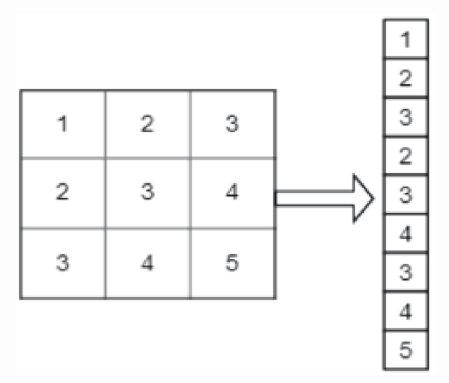

Fig. 6. Example of Flatten operation

ReLU Activation: ReLU (Rectified Linear Unit) activation is commonly used as an activation function in neural networks. This adds non-linearity to the neural network. This basically, returns 0 if it receives a negative value and for a positive value, it returns the same value. Thus, helps in decreasing the time for training.

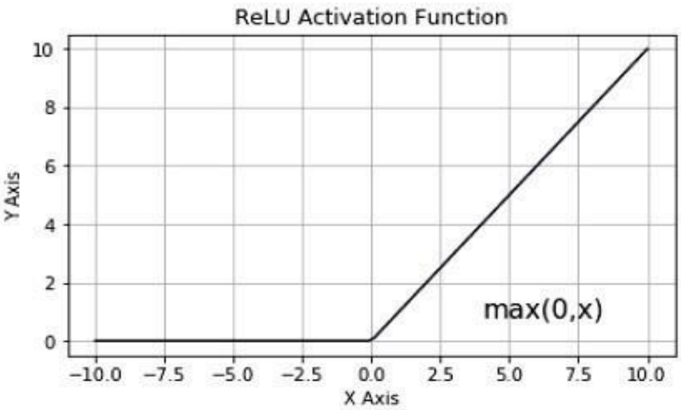

Fig. 7. ReLU Activation function

Softmax Activation: The Softmax function is commonly used with cross-entropy loss in multi-class classification problems. This is used to predict the multinomial probability distribution of the model. This function returns the probability of all inputs such that the sum of all probabilities is equal to 1. So, if one input is negative or small then, the function gives a small probability and if the input is large then, it gives a large probability but these probabilities will always be between 0 and 1.

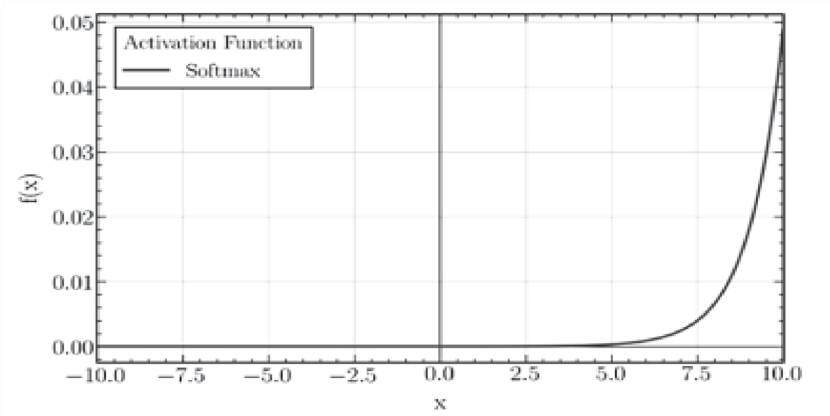

Fig. 8. Softmax Activation function

1. Adam Optimizer: Adam optimizer is a learning rate method, which computes learning rates for various parameters. It is a combination of Stochastic Gradient Descent and RMSprop. In this optimizer for momentum, we make use of the moving average of gradient which, is similar to Stochastic Gradient Descent that uses the gradient. It also uses square gradients to scale the dynamic learning rate which is similar to the RMS prop.
2. Cross-Entropy Loss Function: Cross-Entropy is the most important loss function. It is used in optimizing the binary/multi-class classification models such that it minimizes the loss. It calculates the difference between the probability of network outputs and of labels and then returns the total loss of the network. The smaller is the loss, the better is the model. A model having cross-entropy loss as 0, is said to be perfect. This leads to improving generalization and gives faster training.

A. Convolutional Neural Network Architecture

All Convolutional and Fully Connected layers process 2D data using ReLU (Rectified Linear Unit) and Softmax activation function. ReLU is used due to its faster computation and Softmax is used due to its multi-variate classification. Max Pooling layer is used to highlight the most present features in the image. The dropout layer is a simple way used to prevent overfitting and is capable of better generalization. Since there are 7 classes of emotions, we use categorical cross-entropy loss. Flatten layer is used to change the shape of the data into the correct format for the dense layer to interpret. The dense layer is used for final classification.

B. Network Training

The training is conducted on the CNN model. For training the network, we found the best optimizer to be Adam, using categorical cross-entropy loss. The Adam optimizer used was applied with a dynamic learning rate of $1e^{-4}$ such that it gets trained more efficiently and in less time. Each batch containing 64 images is trained in successive order, taking into account the updated weights coming from the appliance of the previous batch. The total epochs of this training are 50 and the steps of an epoch are 448. The training takes about 6 hours to complete.

C. Real-Time Testing

Following the development of CNN's proposed architecture, the formed model underwent real-time testing. Using the webcam's video and OpenCV implementation of the Haar Cascade library, we detect a face region. We extract, grayscale, and resize the face region to 48x48. Then, we use the Convolutional Neural Network model to predict the probability distribution over emotions and display the respective emotion above the face region.

IV. RESULT AND DISCUSSION :

In this paper, we noted the significant interest of researchers in FER via deep learning over recent years. The automatic FER task goes through different steps like data processing, proposed model architecture, and finally emotion recognition.

This implementation automatically senses emotions on each face in the webcam stream. Using a simple 4-layer CNN with 50 epochs, the train dataset was 85% accurate and the test dataset was 63.2% accurate.

V. Conclusion :

This paper contemplated the development of a low cost and effective solution for the challenging task of recognizing human emotions based on their facial expressions. The seven emotions that the system will be able to detect are Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprised. This was achieved by the implementation of a Convolutional Neural network for training a dataset and achieving high classification accuracy. The insignificant pixels (pixels outside facial expressions) were reduced with the help of the Haar Cascade library present in the OpenCV module. The technology proposed can detect the real-time emotion of the objects present in the video frame. The assessment of facial expressions was based on the perspectives of several scientists in this domain. It is hoped that this paper will have an impact on the way behavioral analysis is carried out. The proper implementation of this technology could change the way we do many things in the present.

REFERENCES :

1. Vandal, T., McDuff, D., & El Kaliouby, R. Event detection: Ultra large-scale clustering of facial expressions. In Automatic Face and Gesture Recognition(FG), 2015 11th IEEE International Conference and Workshops on Vol. 1, pages 1–8, 2015.
2. Irene Kotsia, Ioan Buciu, and Ioannis Pitas. An analysis of facial expression recognition under partial facial image occlusion. Image and Vision Computing, 26(7):1052–1067, 2008.
3. Nakasone, Arturo, Helmut Prendinger, and Mitsuru Ishizuka. "Emotion recognition from electromyography and skin conductance." Proc. of the 5th international workshop on biosignal interpretation. 2005.
4. Li, Chieh-En James and Zhao, Lanqing, "Emotion Recognition using Convolutional Neural Networks" (2019). Purdue Undergraduate Research Conference. 63.
https://docs.lib.purdue.edu/purc/2019/Posters/63
5. M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh and A. Akan, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture," 2019 Medical Technologies Congress (TIPTEKNO), 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895215.
6. Tanner Gilligan, Baris Akis Emotion AI, Real-Time Emotion Detection using CNN " http://web.stanford.edu/class/cs231a/prev_projects_2016/emotion-ai-real.pdf
7. R. Pathar, A. Adivarekar, A. Mishra and A. Deshmukh, "Human Emotion Recognition using Convolutional Neural Network in Real Time," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019, pp. 1-7, doi: 10.1109/ICIICT1.2019.8741491.
8. Huang, Y.; Chen, F.; Lv, S.; Wang, X. Facial Expression Recognition:
9. A Survey. Symmetry 2019, 11, 1189
https://doi.org/10.3390/sym11101189
10. Saravanan, Akash & Perichetla, Gurudutt & K.s, Gayathri. (2019). Facial Emotion Recognition using Convolutional Neural Networks.
11. G. A. R. Kumar, R. K. Kumar and G. Sanyal, "Facial emotion analysis using deep convolution neural network," 2017 International Conference on Signal Processing and Communication (ICSPC), 2017,
pp. 369-374, doi: 10.1109/CSPC.2017.8305872.
12. Dachapally, Prudhvi. (2017). Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units.
13. Helaly, Rabie & Hajjaji, Mohamed & Faouzi, M'Sahli & Abdellatif, Mtibaa. (2020). Deep Convolution Neural Network Implementation for Emotion Recognition System. 261-265. 10.1109/STA50679.2020.9329302.
14. Ghaffar, Faisal. (2020). Facial Emotions Recognition using Convolutional Neural Net.