

Face Emotion Recognition using Convolution Neural Network

Raheena Bagwan¹, Komal Dhapudkar², Sakshi Chintawar³ and Alisha Balamwar⁴

¹⁻⁴ GH Raisoni College Of Engineering & Management/Computer Department, Pune, India

Email: raheena.bagwan.cs@ghrcem.raisoni.net

Email: {komal.dhapudkar,sakshi.chintawar,alisha.balamwar}.cs@ghrcem.raisoni.net

Abstract—Facial Expression has a significant role in every aspect of human life, either it can be for a communication purpose or to improve mental health, or to perceive the state of mind of a person. Even though there are many languages spoken worldwide by humans to communicate with each other which varies with the factors such as different community, different religion or different countries, etc., but sometimes when there are no words left for communication the facial expression is the one which comes into the picture and enhances the communication. The Face Emotion Recognition System has benevolence in the research and development industry. The systems that can detect the face and its features in the form of different emotions gives a rise to the terminology of Human Computer Interaction, in real life where the computers have superpowers to interact with humans. Humans socially interact with each other via emotions when communicating through words is not possible. In this research paper, we have discussed an approach of how does a system works that recognize Facial emotion using a Convolution Neural Network. A convolution Neural Network is said to be a best suited Neural Network for Pattern Recognition and Classification. Convolution Neural Network has an advantage, in which it reduces the dimension for large resolution images, without losing the quality of the image. Convolution Neural Network also captures and predicts features from Non-Frontal Face Images. This system can be useful for Sentimental Analysis, Business Analysis i.e. can be useful for getting a person's review on a certain product, in Automated Driving Cars, for the treatment of Autism Spectrum Disorder and many more. The primary focus is to conceptually understand the working of Convolution Neural Network for Face Emotion Recognition.

Index Terms— Face Detection, Face Emotion Recognition, Convolution Neural Network, Image Processing.

I. INTRODUCTION

In real life when a person becomes speechless due to some uncertainty they may express their sadness/fear/anger/surprise or when a hustle and a bustle is going on in an everyday life at that moment either they would just smile by looking at the person that they would be communicated if they had time to do so or either they might be in a restless situation and display a disgust expression, thus their facial expression communicates many things going on in their life by expressing different emotions particularly, we can say this that facial expressions generally displays their state of mind at that particular moment [2]. Humans always had the innate ability to recognize and distinguish between different faces and along with this they also can recognize the emotion and understand the state of mind of the person in front of them, for example, a child recognizing

their mother even though when they are not able to communicate via words.

Humans are said to be social animals that mean they need to communicate with each other, and they interact socially with the help of emotions, which is considered to be the universal language. Apart from language or a medium of speaking Facial emotions are important because it helps us understand the perspective of the person in the front of us. Therefore, the research of facial emotion has naturally gained a lot of consideration over the past decades with applications not only in perceptual and cognitive science but also in effective computing and computer animations.

In the face emotion recognition system, when we recognize different emotions, there is a rise in Human Computer Interaction terminology i.e. it refers to associate the relationship between human life and computer systems. In recent years, with the rapid growth and progress in Deep Learning, it has become easy to extract the feature and model in the formal structure of data with an increase in prediction accuracy, this is possible with the help of Neural Networks. Among several neural networks, Convolution Neural Network is the best suited and most used neural network for image processing and pattern finding in the form of feature extraction, in our case it will recognize an emotion based on various patterns after being trained [3]. Convolution Neural Network also has an edge of processing non-frontal images i.e. face images at an odd angle, which makes it more versatile. Although using Convolution Neural Network for image processing is time consuming but reasonably worthy as it would give more accuracy when compared to other neural networks.

In this paper, we are going to go through Face Emotion Recognition using Convolution Neural Network for a better result outcome. Hence, building a system that is capable of both face detection and emotion recognition has been a significant area of research.

II. DISCUSSION

A. How Does Image Input Work On A Computer?

In our discussion, we are dealing with the concept of Image Processing rigorously for emotion detection from an image. Image Processing is generally known as a methodology of performing some important kind of operation on an image for a certain output. When an image is provided as an input to the computer system for image processing the available python library without caring what the image is and consists of, it directly converts that image into an array. An Array of pixel values are represented as numbers, where the dimension of the array depends on the resolution of the image. But accordingly, the dimensions may also vary due to the color of the image, it can be grey-scaled or colored. This is how the system manipulates an image based on the pixel value or color of the image.

B. How does the Neural Network work?

Neural Networks consist of three layers *input, hidden, and output* layers. Each having its specific functionalities in the following way:

1. First, we add the input in the input layer i.e. we give an image as an input.
2. After giving input the journey of a neuron ends with the Input Layer and starts with the Hidden layer.
3. Then the activation function makes the specific neural network work based on the value obtained.
4. It gives rise to the propagations in neurons that are the channels for improving the accuracy.
5. After this when we calculate the resultant weight of each position based on the pixel value and input we get certain values.
6. These values can be known as Probability values.
7. After applying the probability upon these values in the hidden layer we finally reach the output layer of the neural network.
8. Thus at the end of the output layer, we get the feature extracted from the input image given to the specific neural network.

This is the workflow of a neural network in Artificial Intelligence from the Input layer to Output layer after traversing many Hidden layers in between.

Let us consider the set of elements represented by the set of input(x) in the form of $x_1, x_2, x_3, \dots, x_n$ are connected to, the connection between the input activation function is drawn by the set of weights($\square$) which are represented by $\square_1, \square_2, \square_3, \dots, \square_n$. Besides this, we have Bayer's letter 'b', where b is another weight multiplied by 'dummy' variable value as 1

After applying the output of the activation function represented by (ϕ) we get neural network as $f(y)$:

$$f(y) = \phi(\sum_i^n i * x_i + b) \quad (1)$$

Where x_i is input, w_i is the weight of x_i , ϕ is the activation function and b is the bias added in this neuron and is negligible. This is where we collect the summation of input as a single unit.

III. SYSTEM ARCHITECTURE

A. Designing of the System

The system architecture for the Convolution Neural Network algorithm can be explained with three main steps. They are as follows:

- (1) Facial Detection – This step involves the detection of the face from any input image given. The result is in the form of a box drawn at the border where a face is detected in an image.
- (2) Facial Recognition – This is an optional step as it can only be used when there is an input image with numerous Facial Detection. This can be skipped by using the image of a single person.
- (3) Emotion Detection – After the training of the model with different facial images of different people following (1) and (2), the emotion of the face is been classified.

Generally, we deal with seven basic emotions that are universally accepted for the face emotion recognition system and mostly used. They can be stated as follows: Happy, Sad, Neutral, Angry, Fearful, Disgust, and Surprised.



Figure 1. Seven Basic Emotions

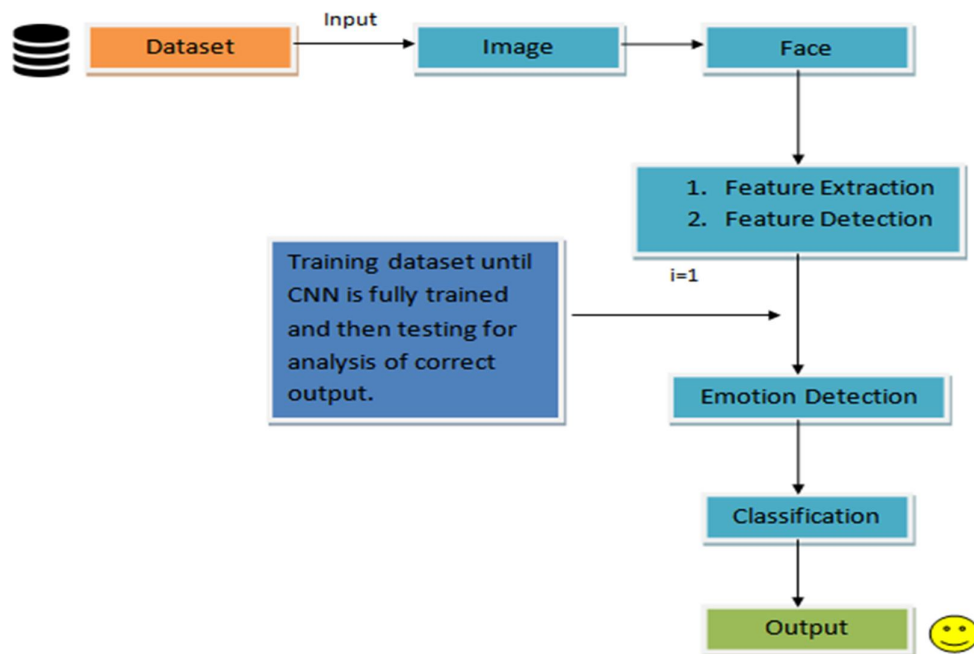


Figure 2. System Architecture

B. Methodology and Workflow of the Algorithm

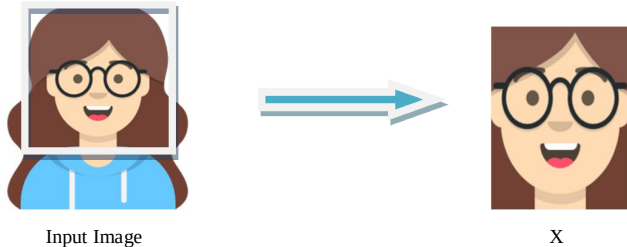
Among several neural networks in deep learning Convolution Neural Network is the most popular and widely used neural network for processing images. Convolution Neural Network is known as Feed-Forward Network due to its property to process and recognize image data with the grid version [6]. As it is a neural network it follows the same methodology of working with three layers namely – *Input Layer*, *Hidden Layers*, and *Output Layer*. *Input Layer* is the one in which the image is given as an input, *Hidden Layers* consist of multiple layers named as *Convolution Layer*, *Re Lu Layer*, *Pooling Layer*, *Fully Connected Network* and *Output Layer* gives the predicted emotion.

Let us further discuss about *Hidden Layers* with their functionalities and mathematical solution for an array of pixels when an input image is given with a feature to be extracted from it, as follows –

Convolution Layer – *Converts images into an array*

- First Layer in *Hidden Layers* of Convolution Neural Network.
- Stores the pixel values of Image into an array.
- Used For extracting the features of the images and reducing their dimensionality.
- The Size of the feature map is controlled by three parameters:
 - Depth-Number of filters used for the convolution operation.
 - Stride- Number of Pixels by which the filter matrix traverses over the input matrix.
 - Padding-It's good to input matrices with zeros around the bordering matrix.

Input Image



Here X is a feature extracted from an image which is converted into an array with numeric values as below:

INPUT WINDOW

4	21	54	92	48
2	22	54	36	22
3	42	37	86	73

filter/feature Map /mask /kernel /template /window detector

0	-1
1	1

In the input window above, all these numeric values represent pixels of the image in that position. Let us consider the above filter value, when we multiply the filter with the first colored array the dimension is reduced and then after performing this step throughout the array we obtain an array with reduced dimensions. Here X represents the position resultant of the particular array when multiplied by a filter, its summation of all weights in that array respectively. Here multiplication of two arrays takes place which results in as follows:

3	22	-2	10
23	25	87	139

$$\begin{aligned}
 X(1,1) &= (4*0)+(21*(-1))+(2*1)+(22*1)=3 \\
 X(1,2) &= (21*0)+(22*(-1))+(54*1)+(54*1)=22 \\
 X(1,3) &= (54*0)+(92*(-1))+(54*1)+(36*1)=(-2) \text{ and so on.}
 \end{aligned}$$

Re Lu Layer (Activation function) or Non-Linearity- Converts negative values into Zero.

Re Lu stands for Rectified Linear Unit.

- Re Lu is half Rectifier
- $f(y)=0$ when $y<0$
- $f(y)=y$ when $y\geq 0$
- Range of Re Lu: [0 to infinity]

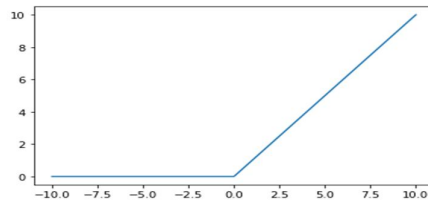


Figure 3. Re Lu function graph

In the above array, the -2 value is replaced by 0, thus after removing the negative value we have our array as

3	22	0	10
23	25	87	139

Pooling Layer or Sub-Sampling - Reduces the spatial size and the numbers of parameters.

Also called sub-sampling or down-sampling. The pooling layer does a down-sampling operation along the spatial dimensions (width, height), resulting in volume such as [2X2X2] i.e. reduces the dimensionality of each feature map but retains important information.

- Used to reduce Dimensionality.
- Helps to control Over-fitting i.e. when models start remembering the data while training instead of learning the data or pattern.
- Filters of size 2X2 are commonly used in it.

25	87	139	139
----	----	-----	-----

Here using the Max Pooling and 2*2 filter with stride =1, we have got the above array.

Flatten the Data - Converting the pooled feature map into an array is known as data flattening.

We get the array as follows

25
87
139
139

Fully Connected Layer or Classification - Combines all the features to create a final model.

- Used to reduce the dimensionality
- Helps to control Over-fitting
- Filters of size 2X2

In this layer, we get the node values that are given as an input to the neural networks.

IV. CONCLUSIONS

Thus, we have successfully discussed and learned about the working of Convolution Neural Network when used to recognize facial expression. Compared to traditional methods, the discussed methodology can automatically learn pattern features and reduce the incompleteness caused by artificial design features [1]. We have studied the mathematical model for Convolution Neural Network while working with image processing for recognizing face emotions. After this discussion and study, we can account that facial expressions strongly deals with communication among people.

In the future, researchers can try to develop a more efficient model so that a more standard facial expression recognition system can be delivered. Facial expressions captured in reality may have various noises, such as face posture, occlusion, and blurring. To address this concern, as future work, we will investigate more robust models which satisfy real conditions.

ACKNOWLEDGMENT

We would like to express our deep gratitude to RaddixTech Software Technology Pvt Ltd. who sponsored our project and have become our helping hands in Data Science and Our College Project Mentor Mr. Sandeep Gore, our research supervisors, for their patient guidance, enthusiastic encouragement and useful critiques for this research work.

Having these people as support is good as being a new learner to this technology makes us motivated towards learning it deeply and makes us curious about knowing new things.

REFERENCES

- [1] Cao, Guolu, et al. "Emotion recognition based on CNN." 2019 Chinese Control Conference (CCC). IEEE, 2019.N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)." SN Appl. Sci. 2, 446 (2020). <https://doi.org/10.1007/s42452-020-2234-1>
- [2] Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2.3 (2020): 1-8.
- [3] Khan, Fuzail. "Facial expression recognition using facial landmark detection and feature extraction via neural networks." arXiv preprint arXiv:1812.04510 (2018).
- [4] Zhou, Yue, et al. "Design of Lightweight Convolutional Neural Network Based on Dimensionality Reduction Module." IOP Conference Series: Materials Science and Engineering. Vol. 533, No. 1. IOP Publishing, 2019.
- [5] Liu, Lingling. "Human face expression recognition based on deep learning-deep convolutional neural network." 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA). IEEE, 2019.
- [6] Yang, Guojun, Jordi Saumell Y. Ortoneda, and Jafar Saniie. "Emotion Recognition Using Deep Neural Network with Vectorized Facial Features." 2018 IEEE International Conference on Electro/Information Technology (EIT). IEEE, 2018.
- [7] Naas, Si-Ahmed, and Stephan Sigg. "Real-time Emotion Recognition for Sales." 2020 16th International Conference on Mobility, Sensing and Networking (MSN). IEEE, 2020.
- [8] Pranav, E., et al. "Facial emotion recognition using deep convolutional neural network." 2020 6th International conference on advanced computing and communication Systems (ICACCS). IEEE, 2020.
- [9] Turabzadeh, Saeed, et al. "Facial expression emotion detection for real-time embedded systems." Technologies 6.1 (2018): 17.
- [10] Mishra, Shrija, et al. "Emotion Recognition Through Facial Gestures-A Deep Learning Approach." International Conference on Mining Intelligence and Knowledge Exploration. Springer, Cham, 2017.