Journal of Engineering Sciences

Vol 11, Issue 7,July/2020

ISSN NO:0377-9254

FACIAL EXPRESSION RECOGNITION

#Nikitha S, #Ponmadathi V , #Praharshita T, *Narendarn G

#Student, *Assistant Professor Department of Computer Science and Engineering,

Sri Ramakrishna Engineering College, Coimbatore-641 022, India

Abstract— Facial expression recognition (FER) has received significant interest from computer scientists and psychologists over recent decades, as it holds promise to an abundance of applications, such as human-computer interaction, affect analysis, and mental health assessment. Although many facial expression recognition systems have been proposed and implemented, majority of them are built on images captured in controlled environment and other lab-collected datasets. In this paper is Human activities for Facial expression recognition using CNN Alogrithms

Keywords— Image Processing, Convolution Neural Network.

1. INTRODUCTION

Facial expressions are natural and powerful signals to interpret human's emotional states and intentions. Nowadays, everything is getting automated through computers. **Facial Expression Recognition (FER)** has become very popular research subject in the field of computer vision. Recognition of facial expressions can be used in robotics, neuro-marketing, academics, and more significantly in security. Accurately predicting facial expressions of human. The identification of objects in an image. This process would probably start with image processing techniques such as noise removal, followed by (low-level) feature extraction to locate lines, regions and possibly areas with certain textures.

The clever bit is to interpret collections of these shapes as single objects, e.g. cars on a road, boxes on a conveyor belt or cancerous cells on a microscope slide. One reason this is an AI problem is that an object can appear very different when viewed from different angles or under different lighting. Another problem is deciding what features belong to what object and which are background or shadows etc. The human visual system performs these tasks mostly unconsciously but a computer requires skillful programming and lots of processing power to approach human performance. Manipulating data in the form of an image through several possible techniques. An image is usually interpreted as a two-dimensional array of brightness values, and is most familiarly represented by such patterns as those of a photographic print, slide, television screen, or movie screen. An image can be processed optically or digitally with a computer.

2. PROJECT SCOPE

The main goal of this project is to develop a program implementing real time Facial Expression Recognition (FER)

Input Image

⇩

Pre processing

⇩

Future Selection

⇩

Extraction Images

⇩

Commands Display

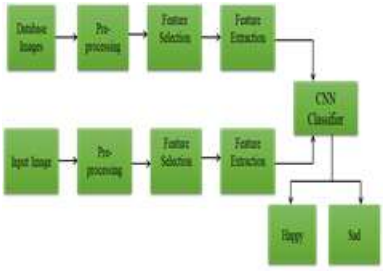
3. PROJECT DESCRIPTION

3.1 PROBLEM DEFINITION

Facial expression recognition (FER) has attracted many researchers in different fields such as human interaction systems, mental disease detection, and affect recognition. Most of the FER systems focus on recognizing six basic expressions. While the majority of existing systems have achieved high

accuracy above 90% in the controlled conditions, real-world applications require more support to improve the accuracy to more than 50% by addressing three main challenges. First, illumination variation, the most common challenge in the wild, has been solved by some researchers. The proposed MBPC could handle pose variation, and slight lighting variation but it is not robust in uncontrollable conditions where various noise levels are added to the images due to illumination variation. This computation time and necessary resources do not fit in real-world applications. Therefore, the FER system area needs to support a fast, robust classifier with an appropriate feature extraction method resistant to the unwanted noises to be applicable in the wild.

3.2 BLOCK DIAGRAM



3.3. PROPOSED SYSTEM

3.3.1 CLASSIFICATION OF IMAGES

There are 3 types of images used in Digital Image Processing. They are

1. Binary Image
2. Gray Scale Image
3. Color Image

1. Binary image

A binary image is a digital image that has only two possible values for each pixel, typically the two colors used for a binary image are black and white through any two colors can be used. The color used for the object(s) in the image is the foreground color while the rest of the image is the background color.

Binary images are also called bi-level or two-level. This means that each pixel is stored as a single bit (0 or 1).This name black and white, monochrome or monochromatic are often used for this concept, but may also designate any images that have only one sample per pixel, such as grayscale images.

Binary images often arise in digital image processing as masks or as the result of certain operations such as segmentation, thresholding, and dithering. Some input/output devices, such as laser printers, fax machines, and bi-level computer displays, can only handle bi-level images.

2. Gray scale image

A grayscale Image is digital image is an image in which the value of each pixel is a single sample, i.e, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray(0-255), varying from black(0) at the weakest intensity to white(255) at the strongest.

Grayscale images are distinct from one-bit black-and-white images, which in the context of computer imaging are images with only the two colors, black, and white (also called bi-level or binary images). Grayscale images have many shades of gray in between. Grayscale images are also called monochromatic, denoting the absence of any chromatic variation.

Grayscale images are often the result of measuring the intensity of light at each pixel in a single band of the electro magnetic spectrum (e.g. infrared, visible light, ultraviolet, etc.), and in such cases they are monochromatic proper when only a given frequency is captured. But also they can be synthesized from a full color image; see the section about converting to grayscale.

3.Color image

A digital color image is a digital image that includes color information for each pixel. Each pixel has a particular value which determines it's appearing color. This value is qualified by three numbers giving the decomposition of the color in the three primary colors Red, Green and Blue. Any color visible to human eye can be represented this way. The decomposition of a color in the three primary colors is quantified by a number between 0 and 255. For example, white will be coded as R = 255, G = 255, B = 255; black will be known as (R,G,B) = (0,0,0); and say, bright pink will be : (255,0,255).

In other words, an image is an enormous two-dimensional array of color values, pixels, each of them coded on 3 bytes, representing the three primary colors. This allows the image to contain a total of 256x256x256 = 16.8 million different colors. This technique is also known as RGB encoding, and is specifically adapted to human vision. Colors are coded on three bytes representing their decomposition on the three primary colors. It sounds obvious to a mathematician to immediately interpret colors as vectors in a three dimension space where each axis stands for one of the primary colors. Therefore it will benefit of most of the geometric mathematical concepts to deal with our colors, such as norms, scalar product, projection, rotation or distance.

3.3.2 BASIC OF IMAGE PROCESSING

1)Image

An image is a two-dimensional picture, which has a similar appearance to some subject usually a physical object or a person.

Image is a two-dimensional, such as a photograph, screen display, and as well as a three-dimensional, such as a statue. They may be captured by optical devices—such as cameras, mirrors, lenses, telescopes, microscopes, etc. and natural objects and phenomena, such as the human eye or water surfaces.

The word image is also used in the broader sense of any two-dimensional figure such as a map, a graph, a pie chart, or an abstract painting. In this wider sense, images can also be rendered manually, such as by drawing, painting, carving, rendered automatically by printing or computer graphics technology, or developed by a combination of methods, especially in a pseudo-photograph.

An image is a rectangular grid of pixels. It has a definite height and a definite width counted in pixels. Each pixel is square and has a fixed size on a given display. However different computer monitors may use different sized pixels. The pixels that constitute an image are ordered as a grid (columns and rows); each pixel consists of numbers representing magnitudes of brightness and color.

Each pixel has a color. The color is a 32-bit integer. The first eight bits determine the redness of the pixel, the next eight bits the greenness, the next eight bits the blueness, and the remaining eight bits the transparency of the pixel.

2) Image file sizes

Image file size is expressed as the number of bytes that increases with the number of pixels composing an image, and the color depth of the pixels. The greater the number of rows and columns, the greater the image resolution, and the larger the file. Also, each pixel of an image increases in size when its color depth increases, an 8-bit pixel (1 byte) stores 256 colors, a 24-bit pixel (3 bytes) stores 16 million colors, the latter known as true color.

Image compression uses algorithms to decrease the size of a file. High resolution cameras produce large image files, ranging from hundreds of kilobytes to megabytes, per the camera's resolution and the image-storage format capacity. High resolution digital cameras record 12 megapixel (1MP = 1,000,000 pixels / 1 million) images, or more, in true color. For example, an image recorded by a 12 MP camera; since each pixel uses 3 bytes to record true color, the uncompressed image would occupy 36,000,000 bytes of memory, a great amount of digital storage for one image, given that cameras must record and store many images to be practical. Faced with large file sizes, both within the camera and a storage disc, image file formats were developed to store such large images.

3)Image file formats

Image file formats are standardized means of organizing and storing images. This entry is about digital image formats used to store photographic and other images.

Image files are composed of either pixel or vector (geometric) data that are rasterized to pixels when displayed (with few exceptions) in a vector graphic display. Including proprietary types, there are hundreds of image file types. The PNG, JPEG, and GIF formats are most often used to display images on the Internet.

In addition to straight image formats, Metafile formats are portable formats which can include both raster and vector information. The metafile format is an intermediate format. Most Windows applications open metafiles and then save them in their own native format.

4)Image processing

Digital image processing, the manipulation of images by computer, is relatively recent development in terms of man's ancient fascination with visual stimuli. In its short history, it has been applied to practically every type of images with varying degree of success.

The inherent subjective appeal of pictorial displays attracts perhaps a disproportionate amount of attention from the scientists and also from the layman. Digital image processing like other glamour fields, suffers from myths, mis-connect ions, mis-understandings and mis-information. It is vast umbrella under which fall diverse aspect of optics, electronics, mathematics, photography graphics and computer technology. It is truly multidisciplinary endeavor ploughed with imprecise jargon.

Several factor combine to indicate a lively future for digital image processing. A major factor is the declining cost of computer equipment. Several new technological trends promise to further promote digital image processing. These include parallel processing mode practical by low cost microprocessors, and the use of charge coupled devices (CCDs) for digitizing, storage during processing and display and large low cost of image storage arrays.

3.3.3 FUNDAMENTAL STEPS IN DIGITAL IMAGE PROCESSING

1) Image Acquisition

Image Acquisition is to acquire a digital image. To do so requires an image sensor and the capability to digitize the signal produced by the sensor. The sensor could be monochrome or color TV camera that produces an entire image of the problem domain every 1/30 sec. the image sensor could also be line scan camera that produces a single image line at a time. In this case, the objects motion past the line.

Scanner produces a two-dimensional image. If the output of the camera or other imaging sensor is not in digital form, an analog to digital converter digitizes it. The nature of the sensor and the image it produces are determined by the application.

2) Image Enhancement

Image enhancementis among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interesting an image. A familiar example of enhancement is when we increase the contrast of an image because "it looks better." It is important to keep in mind that enhancement is a very subjective area of image processing.

3)Image restoration

Image restorationis an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation.

Enhancement, on the other hand, is based on human subjective preferences regarding what constitutes a "good" enhancement result. For example, contrast stretching is considered an enhancement technique because it is based primarily on the pleasing aspects it might present to the viewer, whereas removal of image blur by applying a deblurring function is considered a restoration technique.

4) Color image processing

The use of color in image processing is motivated by two principal factors. First, color is a powerful descriptor that often simplifies object identification and extraction from a scene. Second, humans can discern thousands of color shades and intensities, compared to about only two dozen shades of gray. This second factor is particularly important in manual image analysis.

5) Segmentation

Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually.

On the other hand, weak or erratic segmentation algorithms almost always guarantee eventual failure.

In general, the more accurate the segmentation, the more likely recognition is to succeed.

Digital image is defined as a two dimensional function f(x, y), where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called intensity or grey level of the image at that point. The field of digital image processing refers to processing digital images by means of a digital computer. The digital image is composed of a finite number of elements, each of which has a particular location and value. The elements are referred to as picture elements, image elements, pels, and pixels. Pixel is the term most widely used.

6) Image Compression

Digital Image compression addresses the problem of reducing the amount of data required to represent a digital image. The underlying basis of the reduction process is removal of redundant data. From the mathematical viewpoint, this amounts to transforming a 2D pixel array into a statically uncorrelated data set. The data redundancy is not an abstract concept but a mathematically quantifiable entity. If n1 and n2 denote the number of information-carrying units in two data sets that represent the same information, the relative data redundancy R_D [2] of the first data set (the one characterized by n1) can be defined as, $R_D = 1 - \frac{1}{C_R}$

Where C_R called as compression ratio [2]. It is defined as

$$C_R = \frac{n1}{n2}$$

In image compression, three basic data redundancies can be identified and exploited: Coding redundancy, interpixel redundancy, and phychovisal redundancy. Image compression is achieved when one or more of these redundancies are reduced or eliminated.

The image compression is mainly used for image transmission and storage. Image transmission applications are in broadcast television; remote sensing via satellite, air-craft, radar, or sonar; teleconferencing; computer communications; and facsimile transmission. Image storage is required most commonly for educational and business documents, medical images that arise in computer tomography (CT), magnetic resonance imaging (MRI) and digital radiology, motion pictures, satellite images, weather maps, geological surveys, and so on.

3.3.4 IMAGE COMPRESSION TYPES

There are two types' image compression techniques

1. Lossy Image compression
2. Lossless Image compression

Compression ratio:

$$compression\ ratio = \frac{B_0}{B_1}$$

B_0 – number of bits before compression

B_1 – number of bits after compression

1) Lossy Image compression

Lossy compression provides higher levels of data reduction but result in a less than perfect reproduction of the original image. It provides high compression ratio. lossy image compression is useful in applications such as broadcast television, videoconferencing, and facsimile transmission, in which a certain amount of error is an acceptable trade-off for increased compression performance.

Originally, PGF has been designed to quickly and progressively decode lossy compressed aerial images. A lossy compression mode has been preferred, because in an application like a terrain explorer texture data (e.g., aerial orthophotos) is usually mid-mapped filtered and therefore lossy mapped onto the terrain surface. In addition, decoding lossy compressed images is usually faster than decoding lossless compressed image.

In the next test series we evaluate the lossy compression efficiency of PGF. One of the best competitors in this area is for sure JPEG 2000. Since JPEG 2000 has two different filters, we used the one with the better trade-off between compression efficiency and runtime. On our machine the 5/3 filter set has a better trade-off than the other. However, JPEG 2000 has in both cases a remarkable good compression efficiency for very high compression ratios but also a very poor encoding and decoding speed.

The other competitor is JPEG. JPEG is one of the most popular image file formats. It is very fast and has a reasonably good compression efficiency for a wide range of compression ratios. The drawbacks of JPEG are the missing lossless compression and the oftenmissing progressive decoding. The average rate-distortion behavior for the images in the Kodak test set when fixed (i.e., nonprogressive) lossy compression is used. The PSNR of PGF is on average 3% smaller than the PSNR of JPEG 2000, but 3% better than JPEG.

3.3.5 IMAGE COMPRESSION STANDARDS

There are many methods available for lossy and lossless, image compression. The efficiency of these coding standardized by some Organizations. The International Standardization Organization (ISO) and Consultative Committee of the International Telephone and Telegraph (CCITT) are defined the image compression standards for both binary and continuous tone (monochrome and Colour) images. Some of the Image Compression Standards are

1. JBIG1

2. JBIG2

3. JPEG-LS

4. DCT based JPEG

5. Wavelet based

JPEG2000

Currently, JPEG2000 is widely used because; the JPEG-2000 standard supports lossy and lossless compression of single-component (e.g., grayscale) and multicomponent (e.g., color) imagery. In addition to this basic compression functionality, however, numerous other features are provided, including:1) progressive recovery of an image by fidelity or resolution; 2) region of interest coding, whereby different parts of an image can be coded with differing fidelity; 3) random access to particular regions of an image without the needed to decode the entire code stream; 4) a flexible file format with provisions for specifying opacity information and image sequences; and 5) good error resilience. Due to its excellent coding performance and many attractive features, JPEG 2000 has a very large potential application base. Some possible application areas include: image archiving, Internet, web browsing, document imaging, digital photography, medical imaging, remote sensing, and desktop publishing.

The main advantage of JPEG2000 over other standards, First, it would addresses a number of weaknesses in the existing JPEG standard. Second, it would provide a number of new features not available in the JPEG standard.

The preceding points led to several key objectives for the new standard, namely that it should: 1) allow efficient lossy and lossless compression within a single unified coding framework, 2) provide superior image quality, both objectively and subjectively, at low bit rates, support additional features such as region of interest coding, and a more flexible file format, 4) avoid excessive computational and memory complexity. Undoubtedly, much of the success of the original JPEG standard can be attributed to its royalty-free nature. Consequently, considerable effort has been made to ensure that minimally-compliant JPEG- 2000 codec can be implemented free of royalties.

3.3.6 CONVOLUTION NEUTRAL NETWORK

Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus. Each convolutional neuron processes data only for its receptive field. Although fully connected feedforward neural networks can be used to learn features as well as classify data, it is not practical to apply this architecture to images. A very high number of neurons would be necessary, even in a shallow (opposite of deep) architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights for *each* neuron in the second layer. The convolution operation brings a solution to this problem as it reduces the number of free parameters, allowing the network to be deeper with fewer parameters. For instance, regardless of image size, tiling regions of size 5 x5, each with the same shared weights, requires only 25 learnable parameters. In this way, it resolves the vanishing or exploding gradients problem in training traditional multi-layer neural networks with many layers by using back propagation.

1) **Pooling**

Convolutional networks may include local or global pooling layers to streamline the underlying computation. Pooling layers reduce the dimensions of the data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, typically 2 x 2. Global pooling acts on all the neurons of the convolutional layer. In addition, pooling may compute a max or an average. Max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Average pooling uses the average value from each of a cluster of neurons at the prior layer.

2) **Fully connected**

Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as thetraditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.

3) **Receptive field**

In neural networks, each neuron receives input from some number of locations in the previous layer. In a fully connected layer, each neuron receives input from every element of the previous layer. In a convolutional layer, neurons receive input from only a restricted subarea of the previous layer. Typically the subarea is of a square shape (e.g., size 5 by 5). The input area of a neuron is called its receptive field. So, in a fully connected layer, the receptive field is the entire previous layer. In a convolutional layer, the receptive area is smaller than the entire previous layer.

4) **Weights**

Each neuron in a neural network computes an output value by applying a specific function to the input values coming from the receptive field in the previous layer. The function that is applied to the input values is determined by a vector of weights and a bias (typically real numbers). Learning, in a neural network, progresses by making iterative adjustments to these biases and weights.

The vector of weights and the bias are called filters and represent particular features of the input (e.g., a particular shape). A distinguishing feature of CNNs is that many neurons can share the same filter. This reduces memory footprint because a single bias and a single vector of weights are used across all receptive fields sharing that filter, as opposed to each receptive field having its own

3.4 OVERVIEW OF THE PROPOSED SYSTEM

Facial Detection, which is the inherent ability of a device to detect presence and location of a face within an input image or frame. Facial Recognition, which is the method by courtesy of which compares multiple faces to identify which belongs to who, and to gather data from it.Emotion Detection, which is the process by virtue of which we classify the emotions on a face according to the input image. It will make use of the python library facedetection to make things easier.This will scan the input image, smoothens it out, reduce the noise and disturbance in the image and then detect the presence of a face within it by checking the arbitrary value of each pixel within the image.

It will then return to the main function, the coordinates of the boundary pixels which surround the face.Next, the program will call on the library facerecognition to test and compare the boundary box ithave with other faces. This process requires the extraction and comparison of key features from the boundary box. Face encoding vectors are generated for each image, and the function essentially checks the distance between these vectors, while comparing each feature. Kaggle has a dataset known as Face Emotion Recognition (FER) which has emotions belonging to the following categories- happiness, sadness, fear, disgust, anger, neutral and surprise. Now, this must simulate a test model, and use a multi-layered Convolutional Neural Network to improve the model's performance. Finally, after running the model through enough tests, it will successfully be able to tell the difference between various human facial expressions and categorize it under the apt emotion.

4.MODULES DESCRIPTION

FER mainly comprises of four modules

1. Pre-processing
2. Feature Selection
3. Feature Extraction
4. Library Modules\

4.1.PRE- PROCESSING

The performance of four **pre-processing** methods are compared namely Contrast adjustment, Intensity adjustment, Histogram equalization, Binarization and Morphological operation and resizing the input images.

The original images in a dataset usually have issues with inconsistent size and contain too much redundant information. the expression is mainly reflected by the eyes, nose, and mouth area, whereas the surrounding area is basically useless. Therefore, it is unnecessary to extract features from the whole image, and the processing of redundant information will only increase the workload of the system. Thus, image pre-processing is necessary. Normalization and equalization were performed on the original images. The facial images were detected and all images were normalized to a gray-level image of size 64 × 64 pixels.

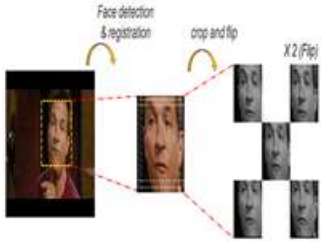


Fig 4.a Gray level image of size 64 × 64 pixels

4.2 FEATURE SELECTION(DWT):

Facial expression recognition consists of three main steps. The first step is face detection in the image. Its effectiveness has a direct influence on the performance of the FER system. The second important step of a FER system is facial features Alternative methods are based on transformation such as Fourier transform (FT), short time FT (ST-FT) and discrete wavelet transform (DWT). Feature extraction based on DWT method is very useful for FER with very low computational cost, which is an ideal tool in image processing and computer vision. The main contributions of the proposed methodology are: first to developing a robust approach of feature extraction, second to improve the performance and speed of FER system and to obtain a high recognition rate.

A model by applying Viola-Jones face detection algorithm, firstly to detect faces, secondly to separate the faces from the rest of the parts, which are considered non-faces. Moreover it is employed to Discrete Wavelet Transform (DWT) on face images to extract features.

4.3 FEATURE EXTRACTION

Feature extracting the facial expression features, training and recognizing the expression feature model. Facial expression feature extraction is the most important part of a FER system. An effective expression feature extraction greatly improves the recognition performance. Many algorithms have been developed for this purpose. Selecting the features.

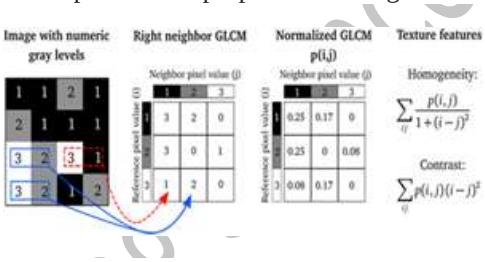


Fig 4.b Selecting the features

4.4 LIBRARAY MODULES

- Numpy
- Matplotlib
- Sklearn
- Tensorflow
- Keras

5. METHODSDESCRIPTION

Convolution layer is the first layer of Convolutional neural network. In this case the input is of 70X70 matrix of pixel values. In the convolutional layer basically, the input matrix is read with a filter. This is it is choosing a sub matrix from this matrix and then this filter is also a matrix of numbers (the numbers are called weights or parameters). The area to be selected by the filter is a square matrix whose area is decided by the Kernel size that is given. The area selected is known as reception field. The Kernel size used is 2. So, a 2X2 reception field will formed. Then, the reception field is changed as the filter starts shifting. This sliding is known as convolution. As the filter is sliding, or convolving, around the input image, it is multiplying the values in the filter with the original pixel values of the image. These multiplications are all summed up and hence give us a single value. This we are having 32 filters. After

sliding the filter to all locations a matrix is formed which is known as activation map or feature map.

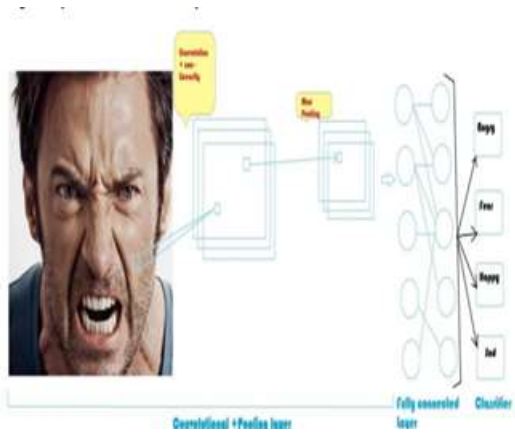


Fig 5.a working of convolution layer

The process is a 2D convolution on the inputs. The "dot products" between weights and inputs are "integrated" across "channels". Filter weights are shared across receptive fields. The filter has same number of layers as input volume channels, and output volume has same "depth" as the number of filters.

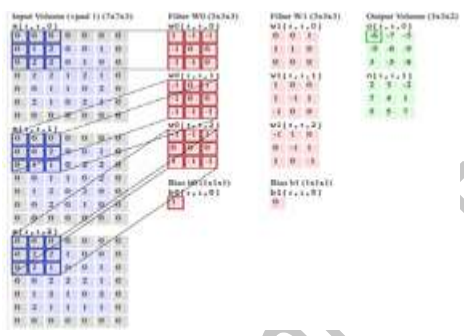


Table 5.1 2D Convolution Working

Accepts a volume of size W1*H1*D1 Requires four hyperparameters:

- Number of filters K,
- Their spatial extent F,
- The stride S,
- The amount of zero padding P,

Produces a volume of size W2*H2*D2

Activation Layer

Used to increase non-linearity of the network without affecting receptive fields of conv layers. Prefer ReLU, results in faster training. Leaky ReLU addresses the vanishing gradient problem

Softmax

A special kind of activation layer, usually at the end of FC layer outputs. Can be viewed as a fancy normalizer (a.k.a. Normalized exponential function). Produce a discrete probability distribution vector. Very convenient when combined with cross-entropy loss.

$$P(y = j \mid \mathbf{x}) = \frac{e^{\mathbf{x}^\top \mathbf{w}_j}}{\sum_{k=1}^{K} e^{\mathbf{x}^\top \mathbf{w}_k}}$$

Given sample vector input **x** and weight vectors {$\mathbf{w}_j$}, the predicted probability of y = j

Fig 5.a Vector input

3. **Pooling Layer**

Convolutional layers provide activation maps. Pooling layer applies non-linear down sampling on activation maps.Pooling is aggressive (discard info); the trend is to use smaller filter size and abandon pooling.

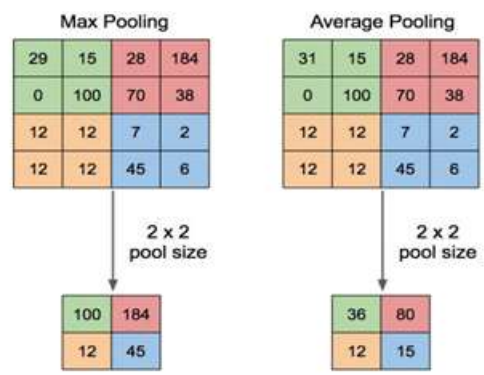


Table 5.2 Pooling filter

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires two hyperparameters:
  - their spatial extent F,
  - the stride S,
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
  - $W_2 = (W_1 - F)/S + 1$
  - $H_2 = (H_1 - F)/S + 1$
  - $D_2 = D_1$
- Introduces zero parameters since it computes a fixed function of the input
- Note that it is not common to use zero-padding for Pooling layers

4. Fully Connected Layer

Regular neural network. Can view as the final learning phase, which maps extracted visual features to desired outputs. Usually adaptive to classification/encoding tasksCommon output is a vector, which is then passed through softmax to represent confidence of classification. The outputs can also be used as "bottleneck"

6. RESULT

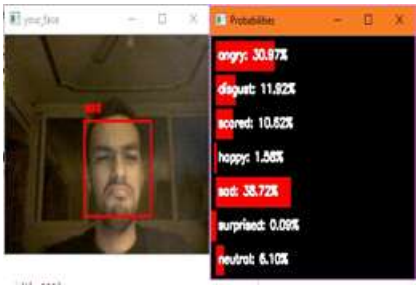


Fig:6.1Sad

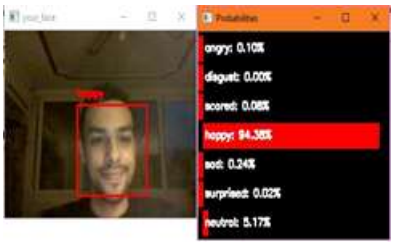


Fig: 6.2Happy

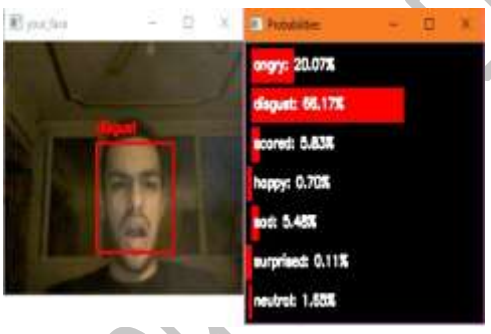


Fig:6.3 Disgust

7. CONCLUSION

The technological improvements in information and communication technologies, a highly anticipated key contributor to improve the customer experience and satisfaction in service episodes is through the application of video analytics, such as to evaluate the customer's emotions over the complete service cycle. The visual emotion recognition system to detect the universal three emotions (happy, angry, sad,) from video data. The detected customer emotions are then mapped and translated to give customer satisfaction scores. Proposed system is to training our database using convolutional neural network after that classification part implementation in it.

8. FUTURE SCOPE

The future work can be extended in following ways. One way is to include more different expressions into the dataset like anger, disgust, fear etc., and in other way the machine will be trained with occluded images also.

REFERENCES

[1] Jatmiko. W, Nulad. W.P, Matul. I.E, Setiawan. I.M.A andMursanto.P (2011) ,'Heart beat classification using wavelet feature based on neural network'. WSEAS Trans. Syst, pp. 17–26, , Vol 10.

[2] Kumar.B.V and Ramakrishnan.A.G. (2002), 'Machine Recognition of Printed Kannada Text', Lect. NotesComput. Sci, pp. 37–48.

[3] Ying. Z and Fang. X (2008),'Combining LBP and Adaboost for facial expression recognition. In Proceedings of the International Conference on Signal Processing', Beijing, China, 26–29 October.