FACIAL EXPRESSION DETECTION USING DEEP NEURAL NETWORKS

B Sai Teja, Ch Chinmai Durga Srinivas, M Sandeep Reddy, Dr. P M Ashok Kumar

Abstract: *Facial Expression conveys non-verbal communication, which plays an important role in acquaintance among people. The Facial Expression Detection system is an activity to identify the emotional state of a person. In this system, a captured frame is compared with trained dataset that is available in database and then state of the captured frame will be defined. This system is based on Image Processing and Machine Learning. For designing a robust facial feature descriptor, we apply the Xception Modelling Algorithm. Xception is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 71 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The detection performance of the proposed method will be evaluated by loading the data-set and pre-processing the images for feeding it to CNN model. Experimental results with prototypic expressions show the superiority of the Xception-Model descriptor against some well- known appearance-based feature representation methods. Experimental results demonstrate the competitive classification accuracy of our proposed method.*

Index Terms: Facial Expression, Xception Modelling, ImageNet, CNN

I Introduction

Facial expression is the representation of the affective state, intention, personality and psychopathology of a person and plays a communicative role in interpersonal relations. Though the implementation has been made, recognizing facial expression with a more accuracy and reducing the noise remains to be burdensome due to the complexity and having many types of various expressions [1].

Generally non-verbal ways of communication is used by humans when they want to express their feeling and these are things generally humans accomplish in day to day life. These include gestures, facial expressions and unconditioned languages. This system can be significantly useful for people to communicate with each other. The mode of communication is raising because it could be widely used in many fields like medical assessment, lie detection, human social/physiological interaction and human computer interface. The Facial Action Coding System (FACS), which was proposed in 1978 by Ekman and refined in 2002, is a very popular facial expression analysis tool [2].

On a frequent basis, humans repeatedly identify emotions by characteristic features, viewed as a piece of a facial expression. For instance, sad is irrefutably connected with a tragic or an upward development of the edges of the lips. Investigation into programmed discovery of outward appearances tends to the issues encompassing the portrayal and classification of static or dynamic attributes of these misshapenness of face pigmentation [3].

The framework groups outward appearance of a similar individual into the essential feelings emotions namely anger, disgust, fear, happiness, sadness and surprise.

The principle motivation behind this framework is proficient cooperation between people and machines utilizing eye stare, outward appearances, psychological displaying and so forth. Here, detection and classification of facial expressions can be a way for communication between man and machine And the framework power fluctuate from individual to individual and furthermore changes alongside age, sexual orientation, estimate and state of face, and further, even the appearances of a similar individual don't stay consistent with time.

Be that as it may, the natural inconstancy of facial pictures brought about by various variables like varieties in enlightenment, present, arrangement, impediments makes demeanour discovery a difficult errand. A few reviews on facial highlight portrayals for face recognition and demeanour investigation tended to these difficulties and potential arrangements in detail [4].

In today's networked world the need to maintain security of information or physical property is becoming both increasingly important and increasingly difficult. In countries like Nepal the rate of crimes are increasing day by day. No automatic systems are there that can track person's activity. If we will be able to track Facial expressions of persons automatically then we can find the criminal easily since facial expressions changes doing different As we know that Image processing is a method to perform some operations on an image, in order to get an more desirable image or to extract some beneficial information from it. It is a kind of sign processing in which input is an photo and output can also be picture or characteristics/features related with that image. Nowadays, image processing is amongst hastily growing technologies. It types core lookup area within engineering and laptop science disciplines too. Image activities. So we decided to make a Facial Expression Detection System.

We are interested in this project after we went through few papers in this area. The papers were published as per their system creation and way of creating the system for accurate and reliable facial expression detection system.

Therefore we are exceptionally energetic to build up a framework that perceives outward appearance and track the action of an individual.

1.1 Image Processing

Image processing is a methodology to convert a picture to a digital type and perform some operations on that, so as to induce associate increased image or to extract some helpful data from it. It is a sort of signal dispensation during which input is a picture, like video frame or photograph and output could also be image or characteristics related to that image. Usually Image processing system includes treating pictures as 2 dimensional signals whereas applying already set signal process ways to them.

In alternative words, a picture are often outlined by a two-dimensional array specifically organized in rows and columns.

processing essentially consists of the following three steps:

- Importing the picture by means of photo acquisition tools
- Analysis and manipulation of an image
- Output in which result can be altered picture or record that is based on picture analysis.

1.2 Fundamental steps in Digital Image processing

1.2.2 Image Acquisition

This is the primary step or method of the elemental steps of digital image process. Image acquisition can be as straightforward as being given a picture that's already in digital type. Generally, the image acquisition stage involves pre-processing, like scaling etc. Digital photo processing offers with manipulation of digital pics via a digital computer. It is a subfield of alerts and systems however focus mainly on images. DIP focuses on developing a computer gadget that is in a position to perform processing on an image. The enter of that machine is a digital image and the machine method that photo the use of environment friendly algorithms, and gives an photo as an output. The most common example is Adobe Photoshop. It is one of the broadly used application for processing digital images.

1.2.2 Image Enhancement

Image enhancement is the process of adjustment of digital images so that, the results obtained are more suitable further image analysis.

Fig 1.1

Fig 1.1 is an example for Enhancing grayscale images with histogram equalization.

1.2.3 Color Image Process

Color is a powerful descriptor that simplifies object identification and extraction of objects from selected scenes.
Color image processing is divided into two major areas:

- Full-Color: The images in question typically are acquired with a full-color sensor, such as a color TV camera or color scanner.
- Pseudo-Color processing. The problem is on of assigning a color to a particular monochrome intensity or range of intensities.

1.2.4 Compression

Compression deals with techniques for reducing the storage needed to save lots of a picture or the information measure to transmit it significantly within the uses of web. It's substantially is necessary to compress knowledge. Image compression is the process of encoding or converting an image file in such a way that it consumes less space than the original file.

1.2.5 Morphological process

Morphological process deals with tools for extracting image elements that are helpful within the illustration and outline of form. Morphological image processing is a collection of non-linear operations associated to the shape or morphology of features in an image. Morphological strategies probe an image with a small shape or template called a structuring element.

1.2.6 Segmentation

Image segmentation is partitioning of a digital image into multiple segments. The goal of segmentation is to simplify and change the representation of an image so that the analysis on it can be simple and easier. Image segmentation is mainly used Research in the fields of face detection and tracking has been very active and there is exhaustive literature available on the same. The major challenge that the researchers face is the non-availability of spontaneous expression data. Capturing spontaneous expressions on images and video is one of the biggest challenges ahead. Many attempts have been made to recognize facial expressions. Zhang et al investigated two types of features, the geometry-based features and Gabor wavelets based features, for facial expression recognition. Appearance based methods, feature invariant methods, knowledge based methods, Template based methods are the face detection strategies whereas Local Binary Pattern phase correlation are the expression detection strategies in related field. Face reader is the premier for automatic analysis of facial expression recognition are some of the API's for expression recognition. Automatic facial expression recognition includes two vital aspects: facial feature representation and classifier problem. Facial feature representation is to extract a set of appropriate features from original face images for describing faces. Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP) are the algorithms used for facial feature representation. LBP is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighbourhood to locate objects and boundaries of images. Image segmentation is the process of assigning a label to each and every pixel in an image so that clustering of pixels based on characteristics can be done.

II Literature Review

of each pixel and considers the result as a binary number. HOG was first proposed by Dalal and Triggs in 2005. HOG numerates the appearance of gradient orientation in a local path of an image. For classifier problem we use algorithms like Machine learning, Neural Network, Support Vector Machine, Deep learning, Naive Bayes. The formation of histogram by using any of facial feature representation will use Support Vector Machine (SVM) for expression recognition. SVM builds a hyperplane to separate the high dimensional space. An ideal separation is achieved when the distance between the hyper plane and the training data of any class is the largest.

III Proposed Methodology

Xception is a convolutional neural network that is trained on more than million images from ImageNet database. The modelling consists of 71 layers in the format of deep structure and classification of images into more than 1000 different.

Xception is a convolutional neural network that is trained on more than million images from ImageNet database. The modelling consists of 71 layers in the format of deep structure and classification of images into more than 1000 different.

3.3 Xception Modelling

Xception is a convolutional neural network that is trained on more than million images from ImageNet database. The modelling consists of 71 layers in the format of deep structure and classification of images into more than 1000 different.

Architecture of Xception Modelling :

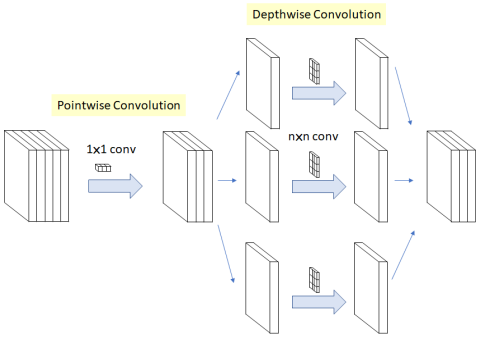

Fig 3.3 Xception Modelling Architecture

Xception is a deep convolutional neural network architecture that involves Depth-wise Separable Convolutions. Google researchers implemented Xception Modelling. Google presented an expound of Inception modules in convolutional neural networks as an initial step in-between regular convolution and the depth wise separable convolution working. A depth-wise separable convolution can be perceived as an Inception module with a great number of towers. This observation tends them to suggest a novel deep convolutional neural network architecture emboldened by Inception. Precedingly depth wise separable convolutions have been introduced instead of inception modules.

The reason behind using Xception modelling is:

- As compared to CNN they have less number of parameters.
- In mobile applications due to fewer computations they are cheaper in vision.

4.3.1 Point-wise Convolution

A 1×1 convolution is called as point wise convolution. 1x1 convolution work is uncomplicated task which is to connecting the input pixel to the output to its similar pixels. It is used to reduce depth channels size and the drawback is that it is an simpler and slower for the large data inputs to calculate multiplication process .
Here, in Xception modelling, to attain the formulation, the process for the dividing each frame into single task is explained above and frames are sent to depth-wise convolution step.

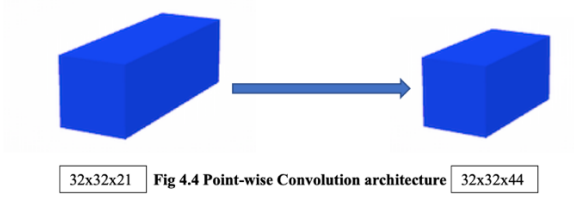

Fig 4.4 Point-wise Convolution architecture

Here, in fig 4.4, through point-wise convolution, the depth of the initial frame is filtered and reduced to point wise kernels and each kernel is sent to undergo depth-wise convolution for further feature extraction process.

4.3.2 Depth-wise Convolution

Depth-wise convolutions are a variation in regular 2D convolution equated over multiple number of input channels, the filter under goes in deep format as it can and lets us combine the channels and also generating each element in the output. Depth-wise convolution do not have combined channels. Here, each channel is kept separate and that is the reason for its name Depth-wise.

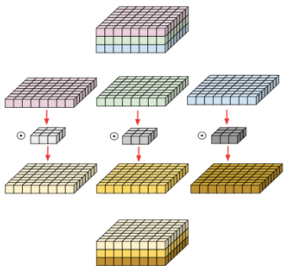

Fig 4.5 Depth-wise Convolution Architecture

There are three conceptual stages here:

1. Divide the input and filter into separate channels (The ratio of number of channels to the filtered channels must be equal).
2. In each channel combine input with the preceding filter to obtain the final layer.
3. Combination of the outputs of the previous is the final channel.

IV Results and Performance Measure

Confusion Matrix: A confusion matrix is the measurement of the effectiveness and performance of our algorithm. It is a table with a combination of actual observations and predicted observations and it is useful for measuring Accuracy, Precision, Recall, and F1score. We used four performance measures like True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). True Positive means you anticipated positive and it is true True Negative means you anticipated negative and it is true. False Positive means you anticipated positive and it is false. False Negative means you anticipated negative and it is false. Accuracy, in general, tells about how regularly is our algorithm or model is correct and it is the proportion of correctly estimated observations to the total observations.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

Confusion Matrix:

Total Images taken: 70

CM	Neutral	Happy	Sad	Disgust	Angry	Surprised	Scared
Neutral	9	1	0	0	0	0	0
Happy	1	9	0	0	0	0	0
Sad	1	0	8	1	0	0	0
Disgust	1	0	1	6	1	1	0
Angry	0	0	0	1	8	1	0
Surprised	1	0	0	0	1	7	1
Scary	0	0	0	1	1	1	7

Table 6.9

V Conclusion

This project proposes an approach for recognizing the category of facial expressions. Face Recognition and Extraction of expressions from these images is useful in many applications, such as robotics, mobile cameras, digital cameras, security and human-computer interaction. This project's objective was to develop a facial expression detection system implementing the computer visions and enhancing the advanced feature extraction and classification in face expression detection. In this system, seven basic facial expressions of different person images from different datasets have been analysed. This project involves facial expression pre-processing of captured facial images followed by modelling through Xception methodology. This project recognizes more facial expressions based on FER face database. To measure the performance of proposed algorithm and methods and check the results accuracy, the system has been evaluated using Precision, Recall and F-score. Experiment results on

FER dataset, show that our proposed method can achieve a good performance. Facial expression detection is a very challenging problem. Our future work will focus on improving the performance of the system and deriving more appropriate classifications which may be useful in many real world applications.

VI Future Scope

Face expression detection systems have improved a lot over the past decade. The focus has definitely shifted from posed expression detection to spontaneous expression detection. Promising results can be obtained under face registration errors, fast processing time, and high correct detection rate (CDR) and significant performance improvements can be obtained in our system. Our system is fully automatic and has the capability to work with images feed. It is able to recognize spontaneous expressions. Our system can be used in Digital Cameras wherein the image can be captured only when the person smiles. Security systems can identify a person in any form of expression he presents himself. Rooms in homes can set the lights, television to a person's taste when they enter the room. Hospitals can use this system to understand the pain or illness of a deaf or dumb patients. Our system can be used to detect and track a user's state of mind, and in mini-marts, shopping center to view the feedback of the customers to enhance the business.

VII References

1. Shan, C., Gong, S., & McOwan, P. W. (2018, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2018. ICIP 2018. IEEE International Conference on (Vol. 4, pp. II-370). IEEE.
2. Bhatt, M., Drashti, H., Rathod, M., Kirit, R., Agravat, M., & Shardul, J. (2019). A Studyof Local Binary Pattern Method for Facial Expression Detection. arXiv preprint arXiv:1405.6130.
3. Chibelushi, C. C., & Bourel, F. (2017). Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 11.
4. Ahmed, F., Bari, H., & Hossain, E. (2019). Person-independent facial expression recognition based on compound local binary pattern (CLBP). Int. Arab J. Inf. Technol., 11(2), 195-203.
5. Zhang, S., Zhao, X., & Lei, B. (2017). Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans. Signal Process, 8(1), 21-31.
6. Happy, S. L., George, A., & Routray, A. (2018, December). A real time facial expression classification system using Local Binary Patterns. In Intelligent Human Computer Interaction (IHCI), 2019 7th International Conference on (pp. 1-5). IEEE.