IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Facial Expression Recognition Using AI

Nikhil Floyd Dsouza,Aravind Naik,Jayasagar Y,Ashna Namira,Hrithik

[1]Student, [2]Assistant Professor, [3]Student, [4]Student,[5]Student

[1]Computer Science Engineering Department,

[1]Srinivas Institute of Technology,Mangalore,Karnataka,India.

Abstract: The emotions evolved in face have an excellent influence on decisions and arguments about various subjects. In psychological theory, emotional states of an individual are often classified into six main categories: surprise, fear, disgust, anger, happiness and sadness. Automatic extraction of those emotions from the face images can help in human computer interaction also as many other applications. Machine learning algorithms and particularly deep neural network can learn complex features and classify the extracted patterns. In this paper, a deep learning based framework is used for human emotion recognition. The proposed framework uses the feature extraction then a Convolutional Neural Network (CNN) for classification. The experimental results show that the proposed methodology increases both of the speed training process of CNN and therefore the recognition accuracy.

Index Terms **- Depthwise Separable Convolution, XCeption Model, XGBoost Algorithm.**

I. INTRODUCTION

For Artificial intelligence systems to recognize human expressions have attracted much research attention, and potential applications of such systems existing in large numbers, comprise domains such as customer-heeding, and emotionally astute robotic interfaces. In view of the important role that facial expression plays in communicating expressions in humans, there has been substantial research interest in computer vision systems to notice human expressions.Some facial expressions have universal meaning. Popularly identified six facial expressions that are universal across all cultures: anger, disgust, fear, happiness, sadness, and surprise. These are the same expressions/emotions that modern facial expression researchers aim to identify and classify using computer vision. In our project, the success of convolutional neural networks(CNNs) in tasks such as object classification is extended to the problem of facial expression recognition. The objective of this project is to classify images of human faces into discrete expression categories. Many established facial expression recognition systems use standard techniques of machine learning and their extracted features.

II. LITERATURE SURVEY

Yongmian Zhang et al. [1] explored the usage of the multisensory data fusion method with Dynamic Bayesian networks (DBNs) for modelling and know-how of the temporal behaviour of facial expressions withinside the picture sequences formed . This version lets in to actively pick the maximum informative visible statistics from the to be had data reassets to reduce the paradox and problems in popularity of the expressions.

Irene Kotsia and Ioannis Pitas [2] defined novel strategies for countenance popularity the usage of SVMs for countenance popularity. The consumer initializes style of the Candide grid nodes at the facial picture depicted on the number one body of the picture sequence.

Yongmian Zhang et al. [3] proposed probabilistic framework for fact duplicate of dynamic facial expressions on a face version with MPEG-four facial animation parameters (FAPs) wherein accomplishing extraordinarily low bitrate in statistics transmission. Casted the FAPs and facial motion writing (FACS) right into a dynamic Bayesian network (DBN) to account for uncertainties in FAP extraction and to version the dynamic evolution legitimate expressions. Thereby ensuing in manufacturing of practical expressions.

Yongqiang Li et al. [4] said facial characteristic factors round every facial component, i.e., eyebrow, mouth, etc., to seize the complete exact face form data. Facial motion gadgets which can be described withinside the facial motion writing represents the contraction of a particular set of facial muscles.Advanced gadget getting to know strategies had been delivered to are looking for out out the version supported each schooling statistics and subjective previous understanding which changed into received via way of means of special reassets.

Muhammad Hameed Siddiqi et al. [5] proposed Stepwise Linear Discriminant Analysis (SWLDA). SWLDA specializes in deciding on the localized functions from the expression frames the usage of the partial F check values.HCRF is able to approximating an advanced distribution the usage of a aggregate of Gaussian density functions.

Christopher Pramerdorfer and Martin Kampel [6] proposed country of the artwork in CNN primarily based totally FER, highlighted key variations among the character works, and in comparison and mentioned their overall performance with a focal point at the underling CNN architectures. They have tested that an ensemble of such CNNs outperforms country of the artwork strategies with out the usage of extra schooling statistics or requiring face registration.

Yimo Guo et al. [7] taken into consideration ranges expression popularity; atlas production level and popularity level. One challenge of the proposed approach is that it's far nevertheless now no longer strong sufficient to conquer demanding situations of sturdy illumination changes. The LDDMM registration set of rules used on this paper won't compensate sturdy illumination changes.

Arushi Raghuvanshi and Vivek Choksi [8] builds mission upon current studies to categorise photographs of human faces into discrete emotion classes the usage of convolutional neural networks (CNNs).They experimented with special architectures and strategies inclusive of fractional maxpooling and best tuning, in the long run accomplishing an accuracy of 48% in a seven-magnificence class task.

Mercy Rani and Durgadevi [9] proposed a singular face emotion popularity gadget from video frames is proposed. Face place is detected via way of means of Viola Jones set of rules and pores and skin shadeation segmentation with RGB shadeation area is used to extract the face pores and skin and non-pores and skin areas. After the segmentation, morphological operations are implemented to extract the boundary of the non-pores and skin areas particularly eyes and mouth. Then the feelings are identified via way of means of calculating the vicinity of the mouth place.

Saeed Turabzadeh et al. [10] proposed a gadget to assess the FPGA implementation and performance of regression strategies for automated emotional country detection and analysis. The database of 5 customers with 63,000 samples changed into recorded. The LBP approach changed into applied for assessment in MATLAB and the FPGA.

III. SYSTEM REQUIREMENTS

A. Web Camera:

A web camera could be a video camera that streams its image in real time to or through a pc network. Web camera is employed to capture the image and store in database. Testing image is captured and matched with the database image of an equivalent person.

B. Processor:

Intel Pentium ahead compatible hardware.

C. RAM:

3GB minimum.

D. Windows 10:

Windows 10 is a Microsoft product for private computers, and most of the other devices. Microsoft delineate Windows 10 as "operating system as a service" that may receive new updates to its options and functionalities, increased with the potential for enterprise environments to receive non essential updates at a slower speed, or use future support milestones which will solely receive critical updates, reminiscent of security patches, over their five year amount of thought support.

E. Flask:

Flask is one amongst the micro net framework written in Python. It will be classified as a micro framework as a result , it doesn't need specific tools or libraries. It does not information abstraction layer, type validation or the opposite parts wherever pre-existing third-party libraries bestow common functions.

F. Python:

Python is an interpreter, and high level artificial language with dynamic semantics. Its high-level intrinsical information structures, integrated with dynamic typewriting and dynamic binding, create it alluring for fast application development likewise as to be used as a scripting or glue language to connect existing element together. Python is simple, straightforward to find out syntax stresses readability and so cut back the charges of program maintenance. Python supports modules and packages, that motivates

code complexness and code reuse. The python interpreter and python standard library are on the market in resources or binary type without any cost for all major Platforms, and might be freely distributed.

IV. SYSTEM IMPLEMENTATION

Implementation is the last and most important phase of software development. It means a discussion of building a new system in operation. The new system may be completely new, replacing existing manuals or the default system or it may be a major overhaul of the existing system. Includes user training, system testing and advanced system performance. It is the stage at which text formatting is turned into an active program. Many arrangements are involved before and during the implementation of the proposed plan.

4.1 Trained model and Algorithm:

4.1.1 XCeption model:

XCeption is a complex design of the Convolutional neural network that incorporates depthwise separable convolutions, Developed by Google researchers. The data starts with an input flow, then passes through an intermediate flow eight times, and finally an exit flow. All separable Convolution and Convolution layers are followed by batch making. Figure 4.1 shows the architecture of XCeption model.

XCeption is an efficient architecture based on 2 key points

1. Depthwise Separable Convolution.
2. Shortcuts between Convolution blocks as in ResNet.

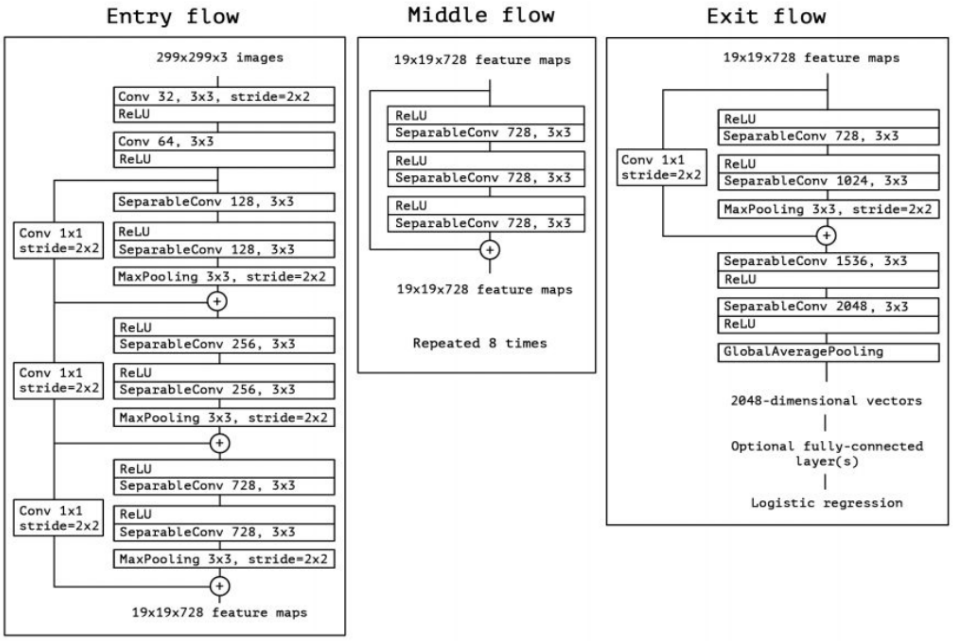

Figure 4.1: demonstrates the construction flow of XCeption model

4.1.2 Depthwise Separable Convolution

Depthwise Separable Convolutions are alternatives to classical convolutions that should be most effective in terms of calculation time. Depthwise Separable Convolutions are themselves divided into 2 main steps:

1. Depthwise Convolution
2. Pointwise Convolution

The Depthwise Convolution

Depthwise Convolution is a first step in which instead of applying convolution of size d×d×Cd×d×C, we apply a convolution of size d×d×1d×d×1. Figure 4.2 shows the illustration of the Depthwise convolution process.

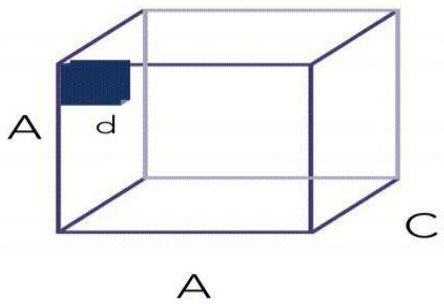

Figure 4.2: shows the diagram of the Depthwise convolution process

This creates a first volume that has size K×K×CK×K×C, and not K×K×NK×K×N as before. Indeed, so far, we have only performed the convolution functionality of 1 kernel / convolution filter, not their NN. This brings us to our second step.

Pointwise Convolution

Pointwise convolution operates a classical convolution, with size 1×1×N1×1×N over the K×K×CK×K×C volume. This allows creating a volume of shape K×K×NK×K×N, as previously.Figure 4.3 shows the illustration of the Pointwise Convolution.

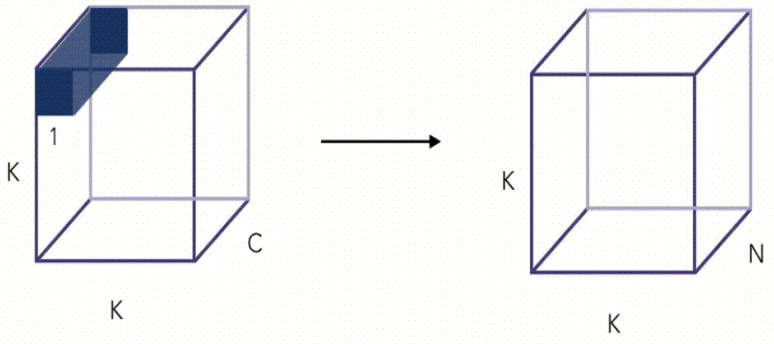

Figure 4.3: Illustration of the Pointwise Convolution

4.1.3 Implementation of the XCeption

XCeption offers built-in blocks of Depthwise Separable Convolution + Maxpooling, all connected to shortcuts as at the beginning of ResNet operations.

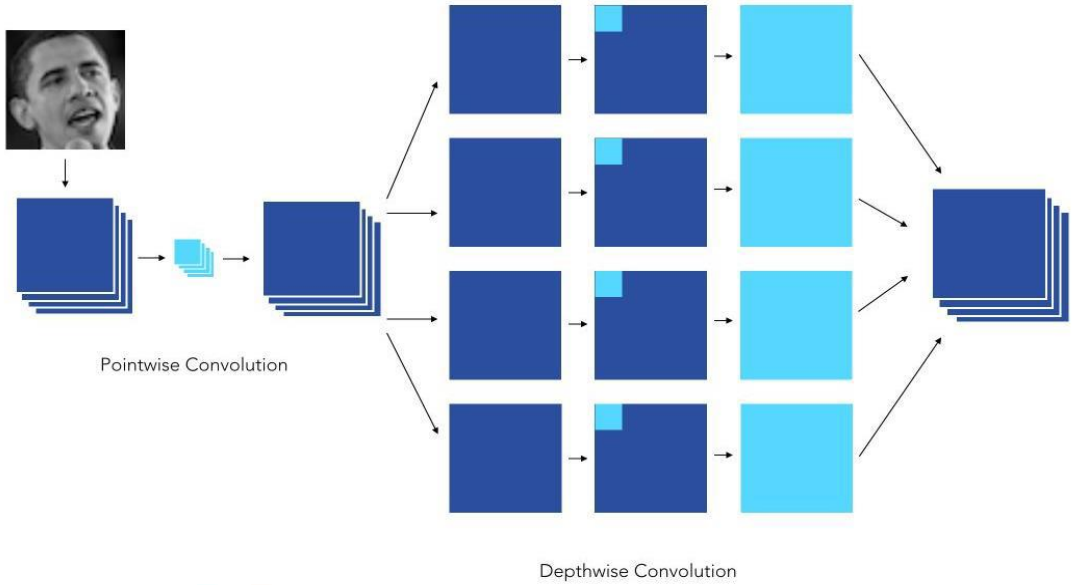

Figure 4.4: Implementation of XCeption incorporated module

4.2 XGBoost Algorithm

The XGBoost algorithm is an algorithm for combining tree-based learning tools using a frame-raising frame. Predictive problems involving unstructured data (images, text, etc.) Neural input networks tend to create all other algorithms or frameworks. XGBoost Library uses a gradient to maximize tree algorithm resolution. This algorithm comes with many different names such as gradient lifting, many additional retrofitting trees, stochastic gradient power upgrades or gradient expansion machines.

Boosting is an ensemble technique where new models are added to correct the errors made by existing models. Models are added one after other until no further improvements are found. A popular example is the AdaBoost algorithm which weighs data points that are difficult to predict. Raising the line is the process by which new models are created that predict the remnants or errors of previous models and then combine them together to make a final prediction. This is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when installing new models. This approach helps to solve regression and classification predictive modelling problems. The algorithm distinguishes itself in the following ways:

a) A wide range of applications: Can be used to solve regression, classification, ranking, and user-defined prediction problems.
b) Portability: Runs smoothly on Windows, Linux, and OS X.
c) Languages: Supports all major programming languages including C++, Python, R, Java, Scala, and Julia.
d) Cloud Integration: Supports AWS, Azure, and Yarn clusters and works well with Flink, Spark, and other ecosystems.

V. TESTING AND RESULTS DISCUSSION

5.1 Testing Methodologies

5.1.1 **Unit Testing**: This is the first phase of testing; the different modules or components are tested individually, often performed by coder himself.

5.1.2 **Integration Testing:** In this type of testing many unit tested modules are combined into subsystems, which are then tested. The goal here is to see if the modules can be integrated properly.

5.1.3 **System Testing**: Here the entire software system is tested. The reference document for this process is the requirement specification and the goal is to see if the software meets the requirements. This form of testing is popularly known as black box testing.

5.1.4 **Acceptance Testing**: It is performed with realistic data of the client to demonstrate that the software is working satisfactorily. It is the test conducted to determine if the requirements of a specification are met.

Table 5.1: **Test cases for Facial Expression Recognition**

Sl. No	Description	Input	Expected Result	Obtained Result
1	Image selection from video source	Video source	Input must be captured from video source	As expected
2	Capturing a number of images and storing in train database	Number of images	Images must be stored in train database	As expected
3	Capturing a number of images and storing in test database	Number of images	Image must be stored in database	As expected
4	Test image must be compared with test database	Load database	Expression recognition will be done and values will be displayed in percentage.	As expected

5.1 Experimentation Results

The system proposed in this project, accepts image continuously as the input. After getting the image the facial region will be detected and all the important features will be extracted. The features which are extracted from the image will form a knowledge-base. In expression recognition phase, captured face image is classified based on the match score from the knowledge-base. The matched label of the Facial Expression will be displayed.

In this project we used popular FER2013 Kaggle Challenge dataset. The data consists of gray scale face images of 48x48 pixel size. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image. The data set remains quite challenging to use, since there are empty pictures, or wrongly classified images.

In the experimentation, we have used 28709 face images for training and 3589 faceimages for testing. The recognition accuracy of our system is 64%.

That is ;

Accuracy % = ((TP + TN) ÷ (TP + TN + FP + FN)) *100 %
= ((1531+765) ÷ (1531+765+862+431)) *100 %
= ((2296) ÷ (3589)) *100 %
=63.9732 (~64 %).

Where TP=True Positive ,TN=True Negative, FP=False Positive and FN=False Negative.

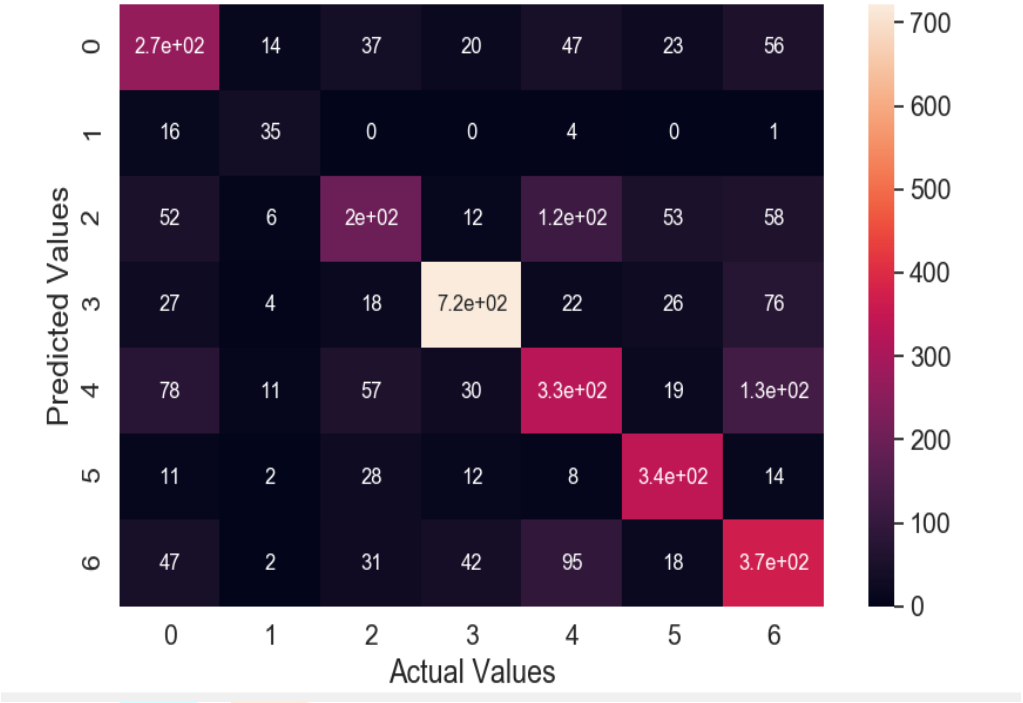

Figure 5.1: Implementation of XCeption incorporated module

VI. CONCLUSION AND FUTURE SCOPE

Here we proposed an correct and high-pace facial features popularity machine. The foremost contribution of this machine is that it is able to discover the expression and the effects tested in order that deep CNN's are able to gaining knowledge of facial traits and enhance facial features popularity. Here the Convolutional networks can intrinsically analyze the important facial capabilities via way of means of the use of most effective raw pixel records. We observed how the pixels are being activated in another way relying at the emotion being labelled. The happiness appears to depend upon the pixels related to the eyes and mouth, while the disappointment or the anger appears as an instance to be extra associated with the eyebrows. In this System we've used the XGBoost library. This library implements the gradient boosting selection tree set of rules. Boosting is an ensemble approach in which new fashions are introduced to accurate the mistakes made via way of means of current fashions. By the use of this set of rules we were given one of the quality version named Xception version. Xception is a deep Convolutional neural community structure that includes Depth smart Separable Convolutions. In this version the records first is going thru the access glide, then thru the center glide that is repeated 8 times, and in the end thru the go out glide. Xception version lets in a shorter schooling time on GPUs, extra pictures processing in line with 2d in actual-time prediction, and stops over fitting. Because of this one we've were given the coolest accuracy and the expressions are diagnosed accurately.

In future the recognition accuracy may be multiplied via way of means of increasing the information records base via way of means of along with extra face pictures having distinct expressions. Android primarily based totally implementation may be carried out to get the actual time effects. The distinct set of rules also can be used to enhance the accuracy.

REFERENCES

[1] Yongmian Zhang and Qiang Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," in IEEE Transactions on Pattern Analysis and MachineIntelligence,vol.27, no.5,pp.699-714,May 2005.

[2] I.KotsiaandI.Pitas,"FacialExpressionRecognitioninImageSequencesUsingGeomet ricDeformation Features and Support Vector Machines," in IEEE Transactions on ImageProcessing,vol.16,no.1,pp.172- 187,Jan.2007.

[3] Y. Zhang, Q. Ji, Z. Zhu and B. Yi, "Dynamic Facial Expression Analysis and Synthesis withMPEG4 Facial Animation Parameters," in IEEE Transactions on Circuits and Systems forVideoTechnology,vol.18,no.10,pp.1383-1396,Oct.2008.

[4] Y. Li, S. Wang, Y. Zhao and Q. Ji, "Simultaneous Facial Feature Tracking and Facial Expression Recognition," in IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2559-2573, July 2013.

[5] Pramerdorfer, Christopher, and Martin Kampel. "Facial expression recognition using convolutional neural networks: State of the Art.", 2016.

[6] Y.Guo, G.Zhao and M.Pietikäinen, "Dynamic Facial Expression Recognition With Atlas Construction and Sparse Representation," in IEEE Transactions on Image Processing, vol. 25,no.5,pp.1977-1992,May 2016.

[7] Arushi Raghuvanshi, Vivek Choksi, "Facial Expression Recognition with Convolutional Neural Networks,"vol.61,no.9,pp.610-628, Jan 2017

[8] Rani, A. Mercy, and R. Durgadevi. "Image Processing Techniques to Recognize Facial Emotions." In International Journal of Engineering and Advanced Technology(IJEAT), vol.6,issue 6,Aug 2017