 

HAROHALLI, KANAKAPURA ROAD – 562112

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(DATA SCIENCE)

**STATISTICAL FOUNDATIONS OF DATA SCIENCE-** **23DS2404**

ON

**“Affect Detection Using Facial Landmarks and Action Units”**

2024-2025

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE)

Submitted by

Adithya N – ENG23DS0049

Shivamurthy B – ENG23DS0036

Sagar M – ENG23DS0029

Under The Supervision of:

MS Shindhu A.

Assistant Professor

Department of CSE (Data Science), DSU

DAYANANDA SAGAR UNIVERSITY

#### DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

**(DATA SCIENCE)**

#### SCHOOL OF ENGINEERING

**HAROHALLI, KANAKAPURA ROAD – 562112**



CERTIFICATE

#### It is certified that the mini project work entitled “Affect Detection using Facial landmarks and Action Units” has been carried out at *Dayananda Sagar University*, Bangalore, by *Adithya N (ENG23DS0049) Shivamurthy B (ENG23DS0036) Sagar M (ENG23DS0029),* Bonafide student of fourth Semester, B.Tech in partial fulfilment for the award of degree in *Bachelor of Technology in Computer Science & Engineering (Data Science)* during academic year *2024-25*. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in departmental library.

#### The project report has been approved as it satisfies the academic requirements in respect of project work for the said degree.

**Signature of the Guide Signature of the Chairperson**

**ACKNOWLEDGEMENT**

A project's successful completion offers a sense of satisfaction, but it is never finished without expressing gratitude to everyone who contributed to its accomplishment. We would like to convey our sincere gratitude to our esteemed university, Dayananda Sagar University, for offering the first-rate facilities.

We are especially thankful to our Chairperson, Dr. Shaila S G, for providing necessary departmental facilities, moral support and encouragement. The largest measure of our acknowledgment is reserved for MS Shindhu A. whose guidance and support made it possible to complete the project work in a timely manner.

We would want to thank everyone who has assisted us in successfully completing this project work, both directly and indirectly. The staff has provided us with a great deal of direction and cooperation.

Name - USN

Adithya N – ENG23DS0049

Shivamurthy B – ENG23DS0036

Sagar M – ENG23DS0029

**DECLARATION**

We hereby declare that the project entitled **"Affect Detection Using Facial Landmark and Action Unit"** submitted to Dayananda Sagar University, Bengaluru, is a bonafide record of the work carried out by our team under the guidance of MS Shindhu A Assistant Professor in the Dayananda Sagar University School of Engineering's Department of Computer Science and Engineering (Data Science). This work is submitted toward the partial fulfillment of the requirements for the award of a Bachelor of Technology in Computer Science and Engineering (Data Science).

Name-USN

Adithya N – ENG23DS0049

Shivamurthy B – ENG23DS0036

Sagar M – ENG23DS0029

**ABSTRACT**

It suggests a unique way to build a neural network to address three tasks in computer vision: finding faces in images, detecting Action Units and detecting emotions. To make predictions for different tasks, the authors design a single-model approach where the process is done one step at a time at different levels in the hierarchy. When using this method, neural networks use their learning capabilities to produce more relevant features, improving them over individual networks used for one task only. This work is important because it investigates optimizing neural networks for real-world usage, with findings on using separable convolutions, binarization and quantization. Each database is used to assess and assess the performance of this method according to its purpose: AffectNet, EmotioNet, RAF-DB and WiderFace are used.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| **CHAPTER** | **PG. NO.** |
| 1. INTRODUCTION | 7-7 |
| 1. OBJECTIVES AND SCOPE OF WORK | 8-8 |
| 1. DESCRIPTION OF WORK | 9-11 |
| 1. SOURCE CODE | 14-14 |
| 1. RESULT | 15-16 |
| 1. CONCLUSION | 17-18 |
| 1. REFERENCES | 19-20 |

**INTRODUCTION**

Showing emotions through your face is vital in non-verbal communication and reveals a great deal about your mood. Automatic detection of emotions from people’s faces can advance work in different fields, most significantly by making HCI tools and interfaces more effective. Two major approaches for predicting emotions are discussed in the paper. Relying on direct facial qualities Determining facial movements defined by the Facial Action Coding System (FACS) Every technique used in economics comes with certain benefits and drawbacks. Despite being able to identify expressions of certain emotions, appearance features have limited use. Alternatively, when using FACS to analyze movements called Action Units (AUs), it becomes possible to express more emotions, although it makes things more complicated and means AUs can be challenging due to the subtle differences between them. Traditionally, detecting emotions requires going through several pre-processing procedures. Face detection Alignment Normalization Feature extraction Classification Now, deep learning has made it possible to use neural networks for these pre-processing activities. However, doing each job singularly increases the workload and adds extra challenges for the architecture. They argue that a cohesive network is able to handle several tasks related to navigation due to learning in steps and using predictions in the process to increase the system’s stability. It is pointed out that for applications that may run on low-power devices, choosing the right architecture is essential for getting the best results. The study reports that applying separable convolutions, binarization and quantization helped reduce the complexity of the network by up to 60%, without affecting its ability to do practical work.

### OBJECTIVE AND SCOPE OF WORK

### Primary Objectives According to the research, there are three main goals. One neural network is proposed and supported to recognize faces, facial actions and the emotions shown by each face all in a linked and orderly way. By conducting evaluations on four different model databases (AffectNet, EmotioNet, RAF-DB and WiderFace), we showed that our approach achieved state-of-the-art results with a single system. Analyzing Network Optimization Techniques: The analysis ensures that several optimization techniques reduce the network’s parameter count, model size and still deliver acceptable performance.

### What is included in the project plan?

### The topic of the research includes: Generating a novel hierarchical network that processes each stage’s predictions using them as information for the following stages Set up a network with three parts and dense blocks that help with localizing a face, detecting its actions and categorizing emotions Building a reliable system to analyze the emotions shown on a person’s face from images Studying methods to ensure networks are efficient both computationally and in their performance Assessing each method by comparing it to the top existing ones for the three tasks It focuses on finding a way to develop facial analysis solutions with deep learning resources.

### DESCRIPTION OF WORK

### A three-tier network is introduced in the paper, with each dense block handling a specified aspect of how the satellite operates.

### 1. The DenseBlock-FL model was designed for face localization. The data processed in this step is a 100×100 picture which is resized to 0.125, 0.25 and 1 to assemble the image pyramid

### The network architecture is modeled after three subsequent networks, as in MTCNN. P-Net: Detects faces and marks the box that outlines each face. R-Net: Eliminates unwanted areas using bounding box regression. O-Net: Checks faces further by detecting and examining the important points in the face For each sub-network, several convolutional, max pooling and dense layers are used as presented in Table I of the paper

### 2. The task of detecting action units is approached with DenseBlock-AU in this paper. Study: Gets the output that comes from DenseBlock-FL after it goes through transition layers The test network has features depth 4 and a growth rate of 32. End-to-End Precision: Features from the previous layers of P-Net, R-Net and O-Net are all combined with the flattened features In the Output Layer, dense layers work with 256, 256 and 1024 neurons to recognize all 11 AUs. Loss: MSE with a threshold of 0.5 is used to classify each facial movement.

### 3. DenseBlock-EMO is a model that detects emotions. Once transition layers are done, Input Processing collects information from DenseBlock-AU. Network Configuration: There are three levels and its growth rate is 64. In Feature Fusion, the outcomes of AU detection and the features from face localization are merged. Classification Layers: Adds 1024, 1024 and 2048 neurons to predict the 7 basic emotions.

### Loss Function: Applies categorical cross-entropy to label emotions Training Procedure:

### The training process takes place step by step, covering different parts of data. In the first step, a program identifies all the faces. Built using a similar approach as MTCNN. Images having a size of 100×100 are fed into the image pyramid. Wider Face and EmotioNet using MTCNN as the source of ground truth results Set up for detecting faces, classifying non-face objects and predicting the locations of facial points. The approach used as the optimizer is Stochastic Gradient Descent (SGD). At this stage, gesture units are identified. Features from level one in P-Net, R-Net and O-Net are made 100×100 before being used as the initial input. Frozen DenseBlock-FL weights are used during this process. The training database used is EmotioNet. With multi-label classification, the MSE loss function and a 0.5 threshold are used.

### The optimizer being used is Stochastic Gradient Descent (SGD). During this stage, the system is able to detect emotions. Gets input from the output of the DenseBlock-AU model. In this part, both DenseBlock-FL and DenseBlock-AU weights are kept unchanged. You can practice using the RAF-DB and AffectNet databases. The loss function is called categorical cross-entropy. Optimizer name: Stochastic Gradient Descent (SGD).

### METHODOLOGY

### Optimization Techniques :

### The writers performed many tests using different optimization strategies.

### 1. Separable Convolutions

### Implementation: Filters previously separated by their depth (factorized 1D filters) replaced standard 2D filters

### Improvements: Can cut parameters by nearly 40% and halve the size of the model

### Problems with Accuracy: The model’s accuracy is just 1.5% lower Divides spatial and depthwise operations, so that the depthwise multiplier is only set to 1

### 2. Binarization

### Method: Leave the weights and activation maps in an array of -1 and +1.

### Using 9-processor, I opted for deterministic binarization where 0.5 was the threshold value. It makes the model around 22% smaller. The accuracy fell by about 2.5 percent. These layers are added to all the convolutional layers and all the dense layers.

### 3. Quantization

### The program shifts the weights from the high precision format to low precision.

### Advantage: The model is reduced in size by 15%.

### Most change provides little loss in the overall score (between 1-2%). Target Layers: Used on both alternate convolutional layers and dense layers.

### Use Several Types of Optimization Separable Convolutions combined with Binarization. Most significant reduction in the size of the model (down by 45%) The biggest decrease is in accuracy (5% drop). You can

### SOURCE CODE

### % main.m

### clc; clear; close all;

### % Read input image

### img = imread('face.jpg');

### imshow(img);

### title('Input Image');

### % Detect landmarks

### landmarks = detectLandmarks(img);

### % Estimate Action Units (simplified)

### activeAUs = detectAUs(landmarks);

### % Map AUs to emotion

### emotion = mapAUsToEmotion(activeAUs);

### % Display result

### fprintf('Detected Emotion: %s\n', emotion);

### function landmarks = detectLandmarks(img)

### % Detect face

### faceDetector = vision.CascadeObjectDetector();

### bbox = step(faceDetector, img);

### if isempty(bbox)

### error('No face detected.');

### end

### % Use shape predictor (simulate 5 landmarks for demo)

### % Replace with actual landmark detection or third-party library

### landmarks = zeros(5, 2);

### for i = 1:size(bbox, 1)

### x = bbox(i,1); y = bbox(i,2);

### w = bbox(i,3); h = bbox(i,4);

### landmarks = [x+w/3 y+h/3; x+2\*w/3 y+h/3; % eyes

### x+w/2 y+h/2; % nose

### x+w/3 y+2\*h/3; x+2\*w/3 y+2\*h/3]; % mouth corners

### break; % Only process one face

### end

### end

### function activeAUs = detectAUs(landmarks)

### % Dummy rule-based AU detection from landmarks

### % For real systems, use trained models like OpenFace or CNNs

### % Example: if mouth corners higher than nose => smile (AU12)

### nose = landmarks(3, :);

### mouth\_left = landmarks(4, :);

### mouth\_right = landmarks(5, :);

### if mouth\_left(2) < nose(2) && mouth\_right(2) < nose(2)

### activeAUs = {'AU12'}; % Smile

### else

### activeAUs = {'AU01'}; % Neutral/raised brow

### end

### end

### function emotion = mapAUsToEmotion(activeAUs)

### % Map AUs to emotions (simplified)

### if ismember('AU12', activeAUs)

### emotion = 'Happy';

### elseif ismember('AU01', activeAUs)

### emotion = 'Neutral';

### else

### emotion = 'Unknown';

### end

### end

### 

### RESULT

### All experimental results for the three tasks are reported in the paper.

### Results of Face Detection (Stage 1)

### Got an average precision of 0.51 when trained using Wider Face. I was able to achieve 0.69 average precision when trained using EmotioNet (MTCNN). Similar results were achieved to other baseline approaches ACF (0.52), Faceness (0.57) and VJ (0.33).

### The findings from the Action Unit Detection process are presented at this step (Stage 2).

### High achievement scores ranging from 0.60 to 0.61 represented different combinations of training databases Performance is comparable to I2R-CCNU-NTU-2 (0.72), JHU (0.71) and I2R-CCNU-NTU-3 (0.69) but a bit lower. The network performed better when it was trained using EmotioNet.

### Detecting Emotions (Stage 3)

### The RAF-DB accuracy was 63.22%, higher than the baselines’ ranges ranging from 56.93% to 74.22%. The best performance with AffectNet is at 59.47%, while baselines were between 57% and 63%. It performs well, even given the structure of the architecture which makes later stages rely heavily on what came before.

### Optimization Results

### Looking closely at the differences between the main model and the many optimized versions:

### Primary model: It achieves the best outcome, but it is also the most complex. Separable convolutions manage to decrease performance and size with equal effectiveness. Quantization lowers the performance only slightly but reduces the size the most.

### CONCLUSION

### Key Findings

### The study provides evidence that using a hierarchical organization enables a network to carry out tasks for face detection, AU detection and emotion recognition simultaneously. According to the research, this is correct: Training a network on various related facial analysis tasks can be done by using a hierarchy. The architecture does as well in performance as specialized models but also offers important integration features. By using networking techniques, the computer workload can be lower without much impact on performance. Such a network can be used on machines or devices with limited computer components. Practical Implications This type of network can be applied to many real-life situations. Improved performance at a low cost to the accuracy of the results Use one model for all data, rather than using different models for each type of data The steps from an image to detecting emotions are done smoothly and efficiently. Appropriate to use in areas where resources are limited Future topics that need to be examined The authors mention important paths that future researchers should consider. Analyzing the connection between hierarchical networks and multi-task joint networks Once the existing techniques are implemented, new optimization approaches can be explored for more improvement. Using a hierarchical approach in other related areas of computer vision Recognizing emotions in videos by acknowledging the sequence of events Evaluating how different ways of training can affect the performance of hierarchical networks.

### REFERENCES

### E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1113–1133, 2015.

### B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," IEEE Transactions on Affective Computing, 2017.

### Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39–58, 2009.

### Y. Wu and Q. Ji, "Constrained joint cascade regression framework for simultaneous facial action unit recognition and facial landmark detection," in IEEE conference on computer vision and pattern recognition, 2016, pp. 3400–3408.

### K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

### F. Chollet, "Xception: Deep learning with depthwise separable convolutions," arXiv preprint, pp. 1610–02 357, 2017.

### M. Courbariaux, I. Hubara, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks: Training deep neural networks with weights and activations constrained to plus 1 or minus 1," arXiv preprint arXiv:1602.02830, 2016.

### I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," Journal of Machine Learning Research, vol. 18,