Real time Facial Emotion Detection using Convolutional Neural Networks

Presented by:
Yashpal Parmar (2019BCS-073)
Anoop Kammari (2020BCS-045)
Anish Thakur (2020BCS-076)

Overview:

1. Introduction
2. Problem Statement
3. Objectives
4. Methodology
5. Results
6. Limitations
7. Conclusion
8. References

INTRODUCTION

Emotion recognition, a fundamental facet of human interaction, is pivotal in numerous fields, ranging from psychology to artificial intelligence. This project presents a dynamic fusion of Convolutional Neural Networks (CNN) and OpenCV to create a real-time emotion recognition system. Leveraging the prowess of deep learning and computer vision, this system interprets facial expressions captured through live video streams, discerns underlying emotions, and instantaneously displays predictions.

This amalgamation not only showcases the potential of modern technologies in capturing human emotions but also offers an interactive and engaging experience in real-time. As we delve into the intricacies of this project, we explore how AI algorithms and real-time video processing converge to decode the complex language of emotions from facial cues.

Role of Facial Emotion Recognition in the Industry

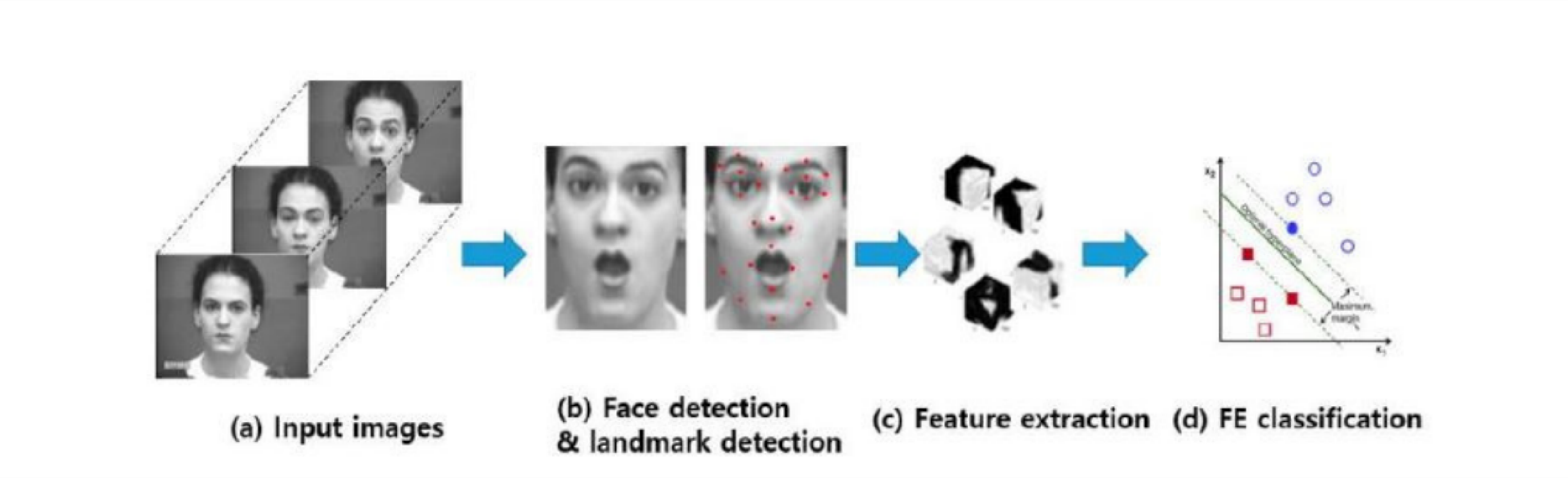
- **Human-Computer Interaction**: It enhances user experiences in applications like virtual assistants, video games, and interactive systems by allowing machines to respond sensitively to users' emotional states.
- **Healthcare:** In mental health diagnostics, emotion recognition assists therapists in understanding patients' emotional states. It's also used to develop tools for individuals with conditions like autism to interpret emotions.
- **Security and Surveillance:** It can augment security systems by detecting suspicious emotional states or stress in public places.
- **Human-Robot Interaction:** Robots can become more intuitive by recognizing and responding to human emotions.
- **Marketing and Advertising:** Emotion recognition aids in gauging customer reactions to advertisements and products. This helps tailor marketing strategies and improve customer engagement

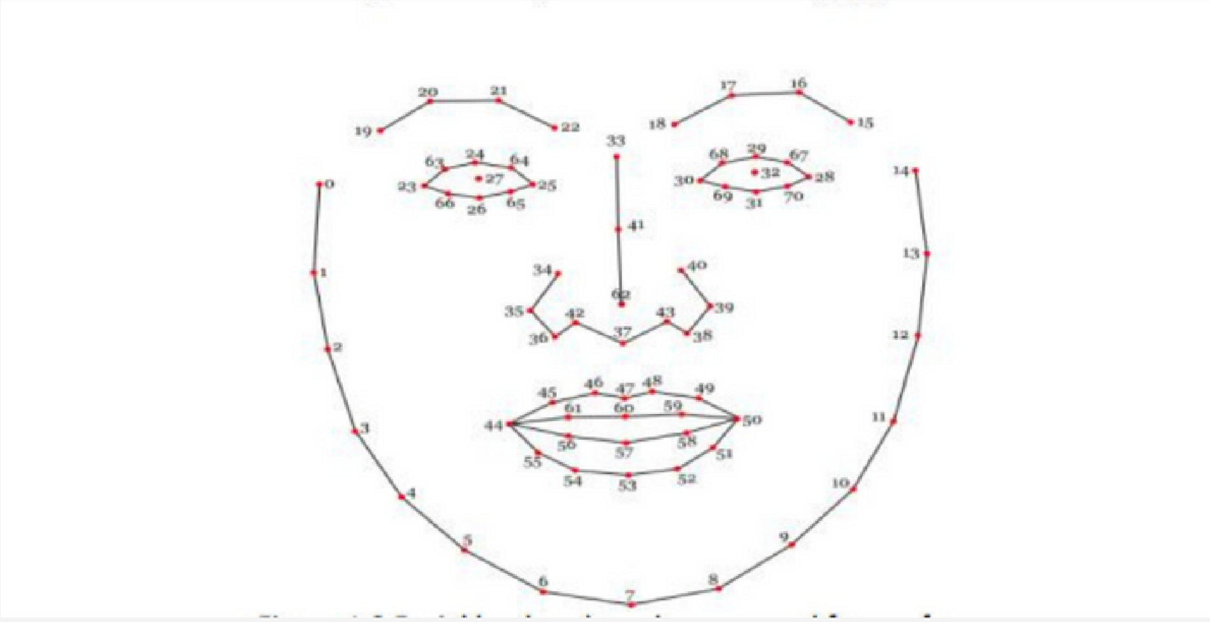
PROBLEM STATEMENT

The project addresses the challenge of real-time emotion recognition from facial expressions using a combination of Convolutional Neural Networks (CNN) and OpenCV. It aims to develop an accurate and responsive system that interprets live video frames, detects faces, and predicts emotions in real-time, while also addressing potential limitations in cultural variation and ethical considerations surrounding privacy and consent.

OBJECTIVES

1. Develop a real-time emotion recognition system by integrating Convolutional Neural Networks (CNN) and OpenCV.
2. Capture live video frames using a webcam and implement face detection to isolate facial regions for analysis.
3. Train a CNN model to learn complex facial features and predict a range of emotions from facial expressions.
4. Display predicted emotions overlaid on the live video feed, providing an interactive and dynamic user experience.
5. Address potential limitations related to cultural variations in expressing emotions and ethical considerations regarding privacy and data security.



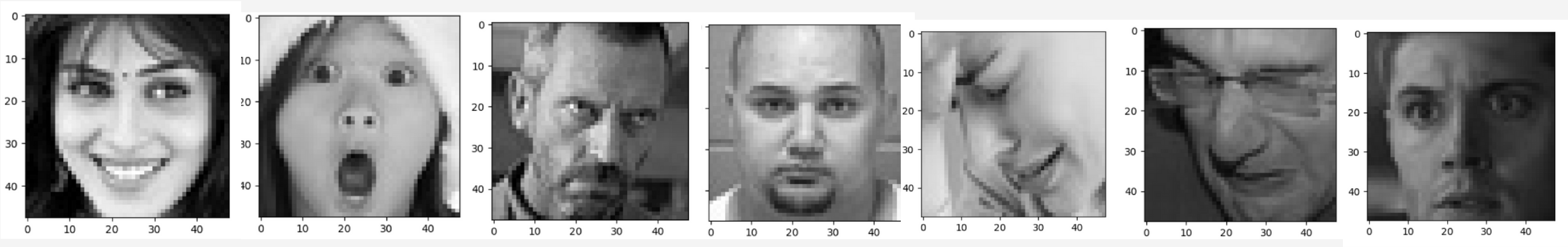


Facial landmarks to be extracted from a face.

METHODOLOGY

Data acquisition

1. The dataset-gathering process for the project involves obtaining the Face Expression Recognition Dataset from Kaggle.
2. The dataset contains 28,821 for training and 7066 images for validation all the images are grayscale.
3. The Original image dimensions are 64x64 pixels.
4. Images in the dataset are categorized into seven classes namely happy, sad, angry, surprise, fear, disgust, and neutral.

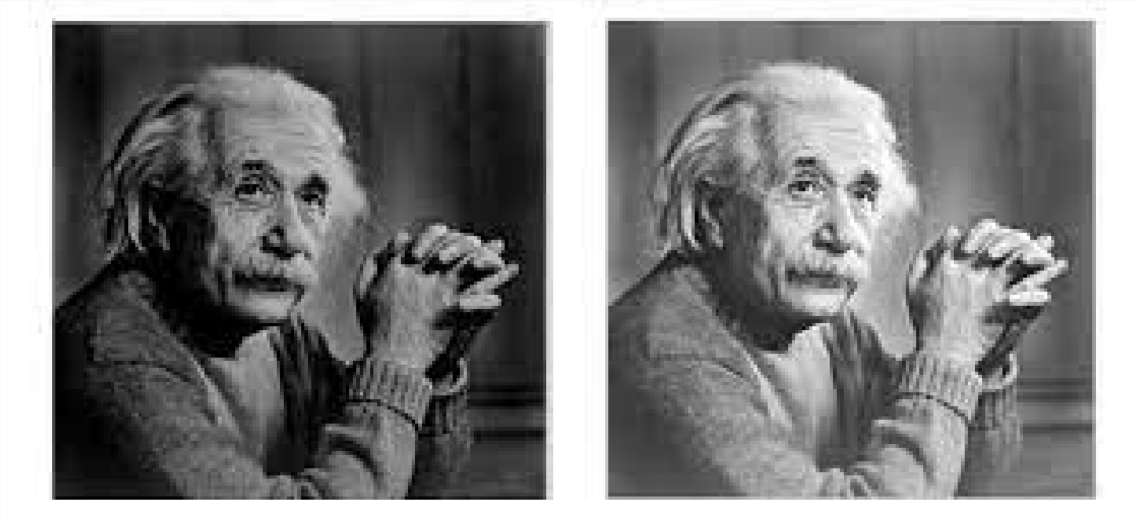


Above images are a preview of the image dataset taken from each class of emotion- happy, surprise, angry, neutral, sad, disgust and fear

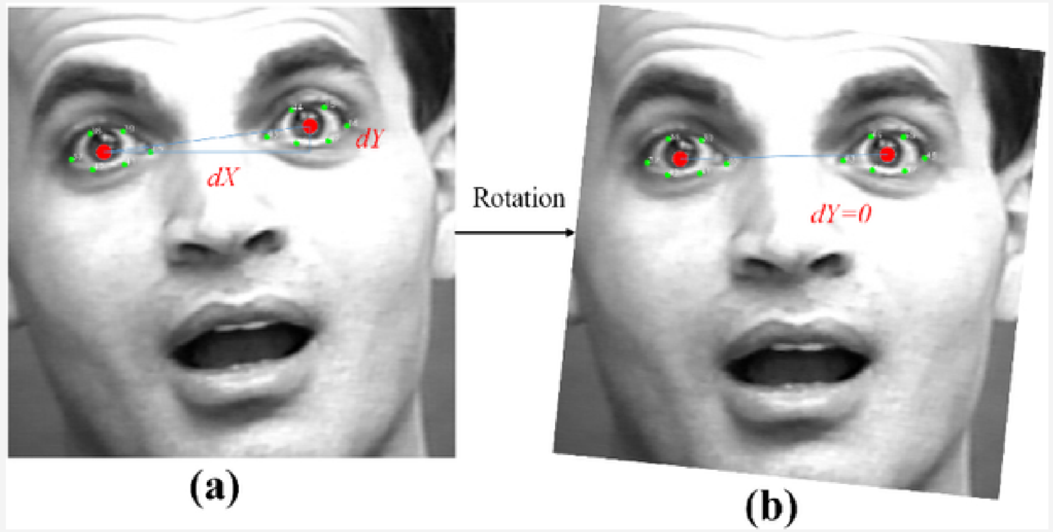
METHODOLOGY

Pre-Processing

1. Several pre-processing algorithms are applied to the model before sending it for learning.
2. The Original image dimensions are changed to 48x48 pixels.
3. The First technique used is rotation which rotates the image to an angle θ (theta).
4. Then Contrast change is added to the images which increases the distance between the maximum and minimum pixel intensities.
5. These techniques are then used to increase the effective dataset size.



Contrast change



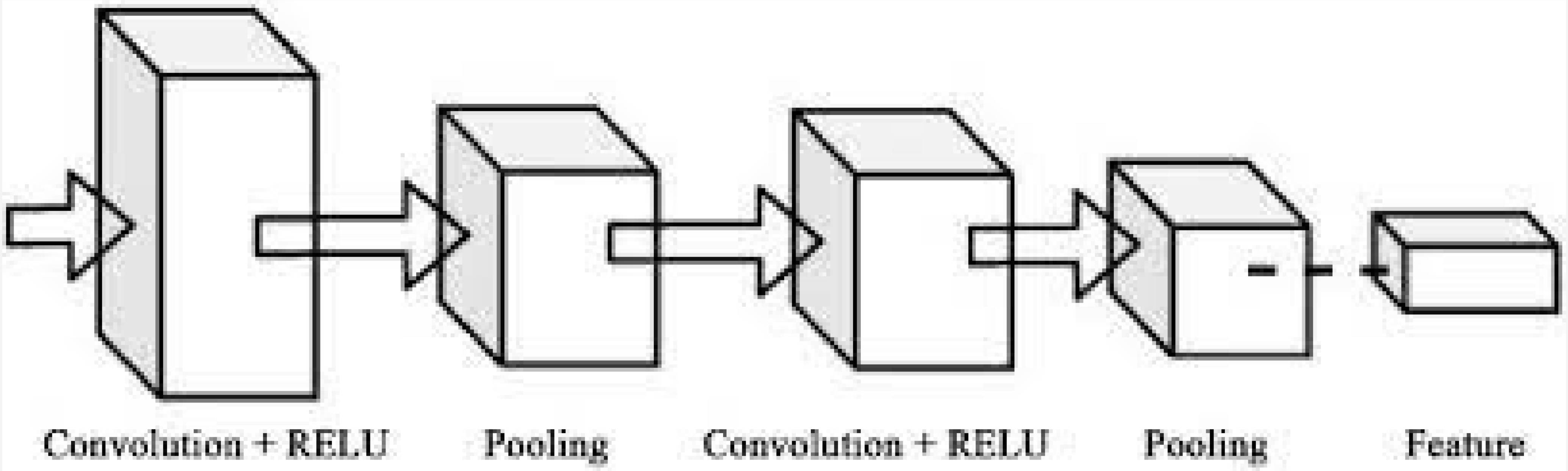
Rotation

Above images are a preview of the images before and after the pre-processing techniques are implemented.

METHODOLOGY

Model Architecture

1. We have created a Sequential model representing the CNN for emotion recognition. It contains four convolutional layers, one flatten layer, two fully connected layers followed by softmax activation layer.
2. Each convolutional layer contains BatchNormalization layer , ReLu Activation layer, MaxPooling2x2, and Dropout layers(0.25).
3. Construct a set of Fully Connected (Dense) layers with BatchNormalization, Activation, and Dropout.
4. Model is compiled using the Adam optimizer with a specified learning rate of 0.0001.



Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 48, 48, 64)	640
batch_normalization_6 (BatchNormalization)	(None, 48, 48, 64)	256
activation_6 (Activation)	(None, 48, 48, 64)	0
max_pooling2d_4 (MaxPooling 2D)	(None, 24, 24, 64)	0
dropout_6 (Dropout)	(None, 24, 24, 64)	0
conv2d_5 (Conv2D)	(None, 24, 24, 128)	204928
batch_normalization_7 (BatchNormalization)	(None, 24, 24, 128)	512
activation_7 (Activation)	(None, 24, 24, 128)	0
max_pooling2d_5 (MaxPooling 2D)	(None, 12, 12, 128)	0
dropout_7 (Dropout)	(None, 12, 12, 128)	0
conv2d_6 (Conv2D)	(None, 12, 12, 512)	590336
batch_normalization_8 (BatchNormalization)	(None, 12, 12, 512)	2048
activation_8 (Activation)	(None, 12, 12, 512)	0
max_pooling2d_6 (MaxPooling 2D)	(None, 6, 6, 512)	0
dropout_8 (Dropout)	(None, 6, 6, 512)	0
conv2d_7 (Conv2D)	(None, 6, 6, 512)	2359808
batch_normalization_9 (BatchNormalization)	(None, 6, 6, 512)	2048
activation_9 (Activation)	(None, 6, 6, 512)	0
max_pooling2d_7 (MaxPooling 2D)	(None, 3, 3, 512)	0
dropout_9 (Dropout)	(None, 3, 3, 512)	0
flatten_1 (Flatten)	(None, 4608)	0

flatten_1 (Flatten)	(None, 4608)	0
dense_3 (Dense)	(None, 256)	1179904
batch_normalization_10 (BatchNormalization)	(None, 256)	1024
activation_10 (Activation)	(None, 256)	0
dropout_10 (Dropout)	(None, 256)	0
dense_4 (Dense)	(None, 512)	131584
batch_normalization_11 (BatchNormalization)	(None, 512)	2048
activation_11 (Activation)	(None, 512)	0
dropout_11 (Dropout)	(None, 512)	0
dense_5 (Dense)	(None, 7)	3591

figure: Model architecture

METHODOLOGY

Early stopping and Reducing learning rate

- Early stopping is employed to prevent overfitting and enhance generalization. It involves monitoring the model’s performance on a validation set during training. If the performance does not improve for a certain number of epochs (controlled by a patience value), the training process is stopped. Early stopping prevents the model from memorizing the training data and ensures it learns useful patterns.

- Learning rate reduction is used to optimize the training process and model convergence. The learning rate determines the step size taken during parameter updates. By dynamically adjusting the learning rate, typically when validation performance plateaus or the loss function stops decreasing, the model can fine-tune its parameters more effectively. This adjustment aids in reaching better local or global minima during optimization.

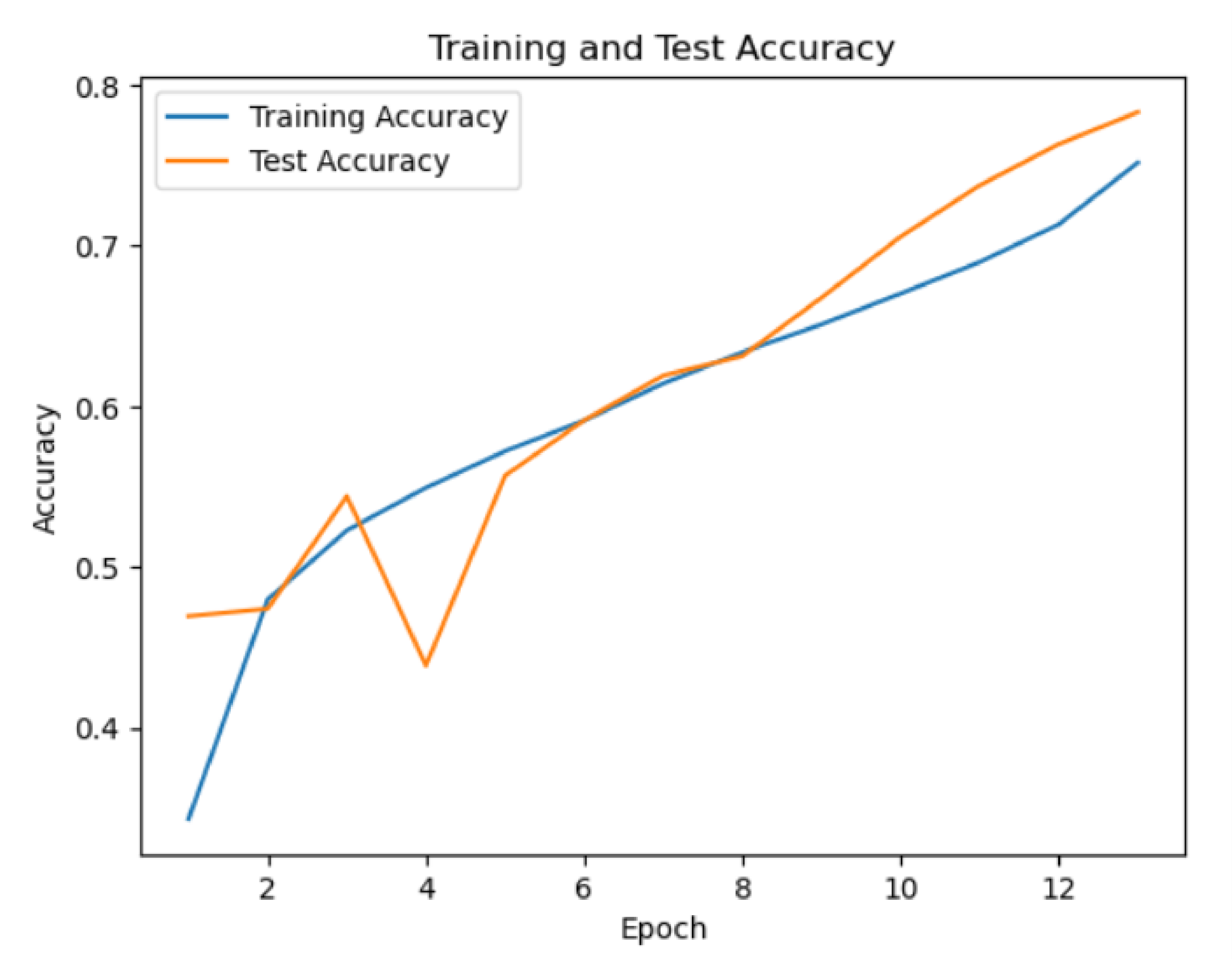
METHODOLOGY

Model Deployment using OpenCV

- OpenCV (Open Source Computer Vision Library) is essential for image preprocessing, face detection, and general visual analysis in the context of the emotion recognition project.
- The main loop of the code continuously captures video frames from the webcam. Faces are detected using the previously initialized Har-CascadeClassifier.
- For each detected face A rectangle is drawn around the face on the frame and the grayscale region of interest (ROI) is extracted from the frame.
- The ROI is resized to match the input size expected by the model (48x48 pixels).
- If the sum of pixel values in the ROI is not zero:
- The ROI is normalized and prepared for prediction and the pre-trained model predicts the emotion label for the ROI. The label with the highest predicted probability is selected. The emotion label is displayed on the frame next to the face.
- If the ROI sum is zero, a message indicating 'No Faces' is displayed on the frame.

Results

Performace evaluation:



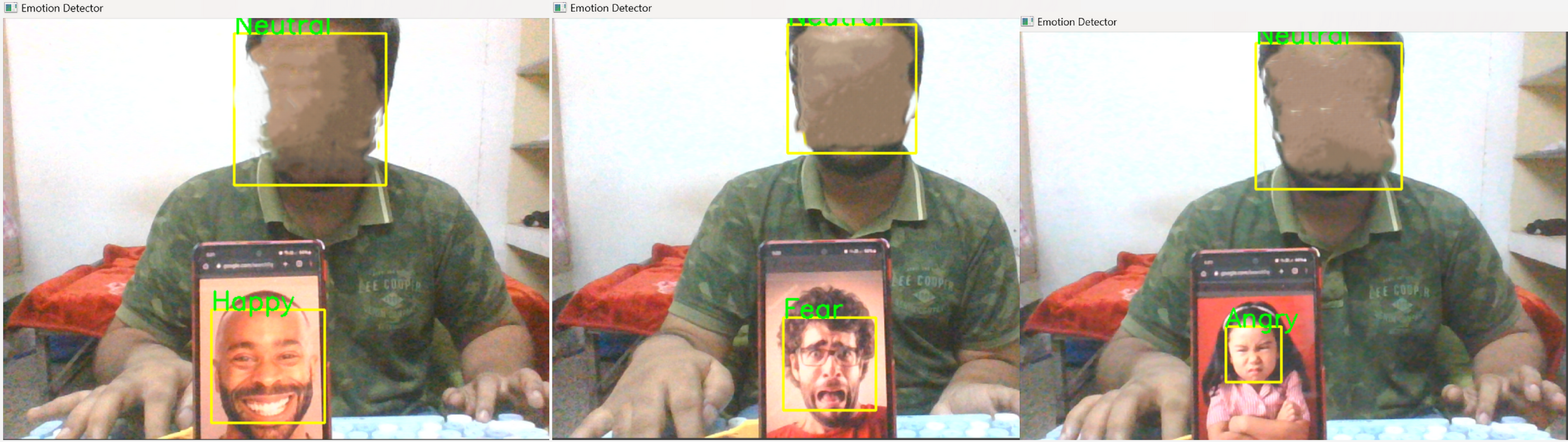
The above graph shows results of the CNNmodel. At the end of 14 epochs the test accuracy is 78.4 % and training accuracy is 75.7% respectively

Results

Performace evaluation:

Results

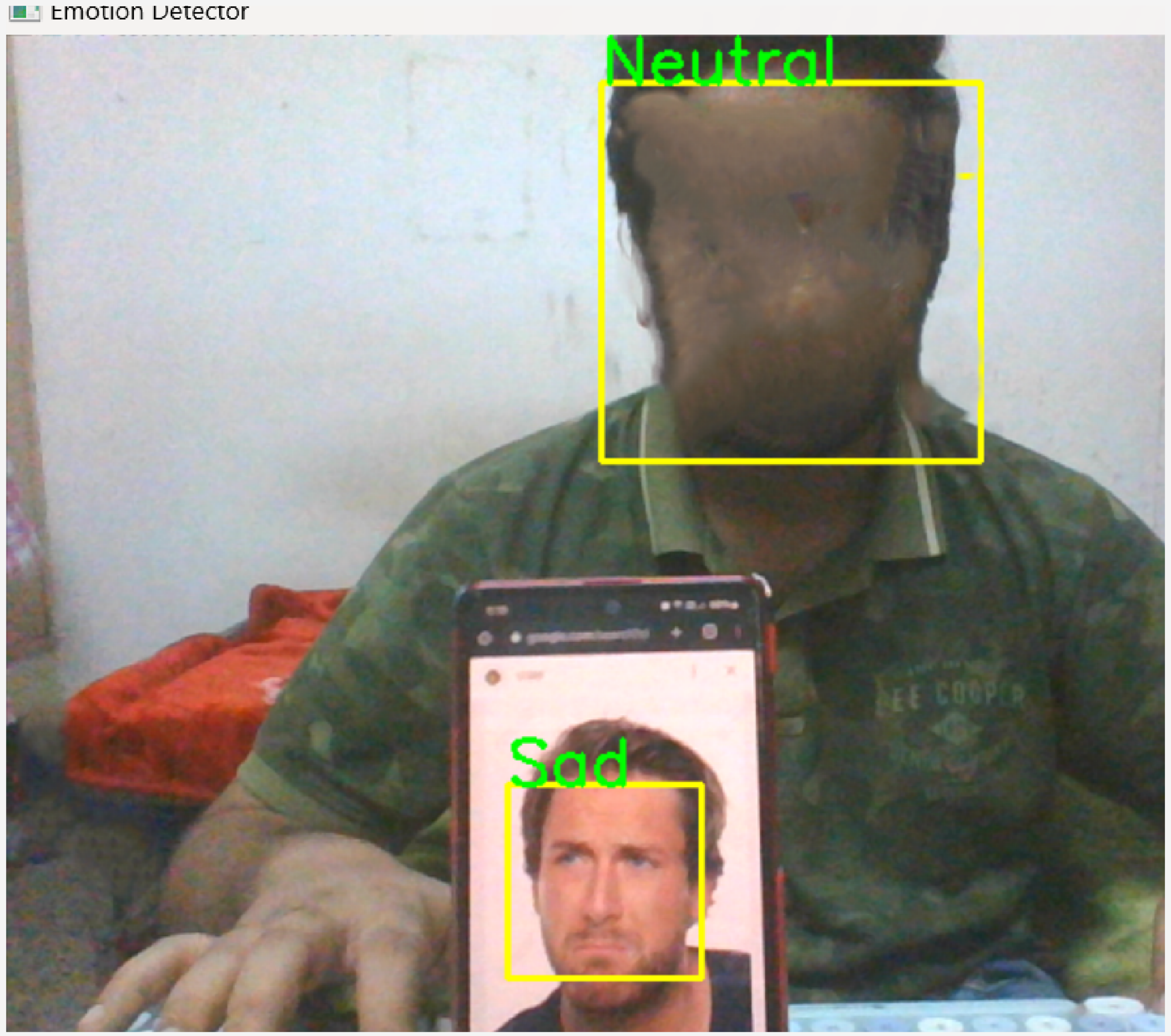
Working of the deployed model :



Above images are screenshots of the working model in OpenCV that is successfully able to recognise the facial emotions from images of random people.

Results

Working of the deployed model :



Few more output images

CONCLUSION

Limitations:

- Viewpoint & Environment Constraints: Variations in lighting, camera angles, and obstructions may hinder accurate emotion recognition.
- Hardware & Computational Demands: Real-time processing demands powerful hardware; less capable systems may experience delays.
- Facial Expression Limitation: The model only detects emotions from facial expressions, disregarding other cues like body language and tone.
- Ethical Concerns: Implementation should address privacy concerns and consent issues when deployed in public spaces.

CONCLUSION

Future Scope:

- Multi-Modal Emotion Recognition: Incorporating voice and text inputs for a comprehensive emotion recognition system.
- Diverse Dataset: Training the model on a diverse dataset to improve performance across demographics.
- Contextual Awareness: Integrating contextual information for better emotion interpretation in different situations.
- Privacy Measures: Implementing techniques to address webcam usage and data privacy concerns.
- Emotion Intensity: Extending the model to predict emotion intensity, providing nuanced insights.
- Ethical Guidelines: Establishing ethical guidelines for responsible deployment and usage.

References:

[1] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In Proceedings of the IEEE Conference on Automatic Facial and Gesture Recognition, 2006.
[2] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on, 2005.
[3] Sanket Shah Pratik Kanani Khushi Chavan, Devanshu Shah. Real-time facial emotion recognition. In 2nd Global Conference for Advancement in Technology (GCAT) Bangalore, India., 2021.
[4] Chieh-En James Li and Lanqing Zhao. Emotion recognition using convolutional neural networks. In Purdue Undergraduate Research Conference, volume 63, 2019.
[5] Ashish Lonare and Shweta V. Jain. A survey on facial expression analysis for emotion recognition. International Journal of Advanced Research in Computer and Communication Engineering, 2(12), 2013.
[6] M. Pantic and J.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. IEEE Transactions on Systems, Man and Cybernetics, 34(3), 2004.

THANK YOU