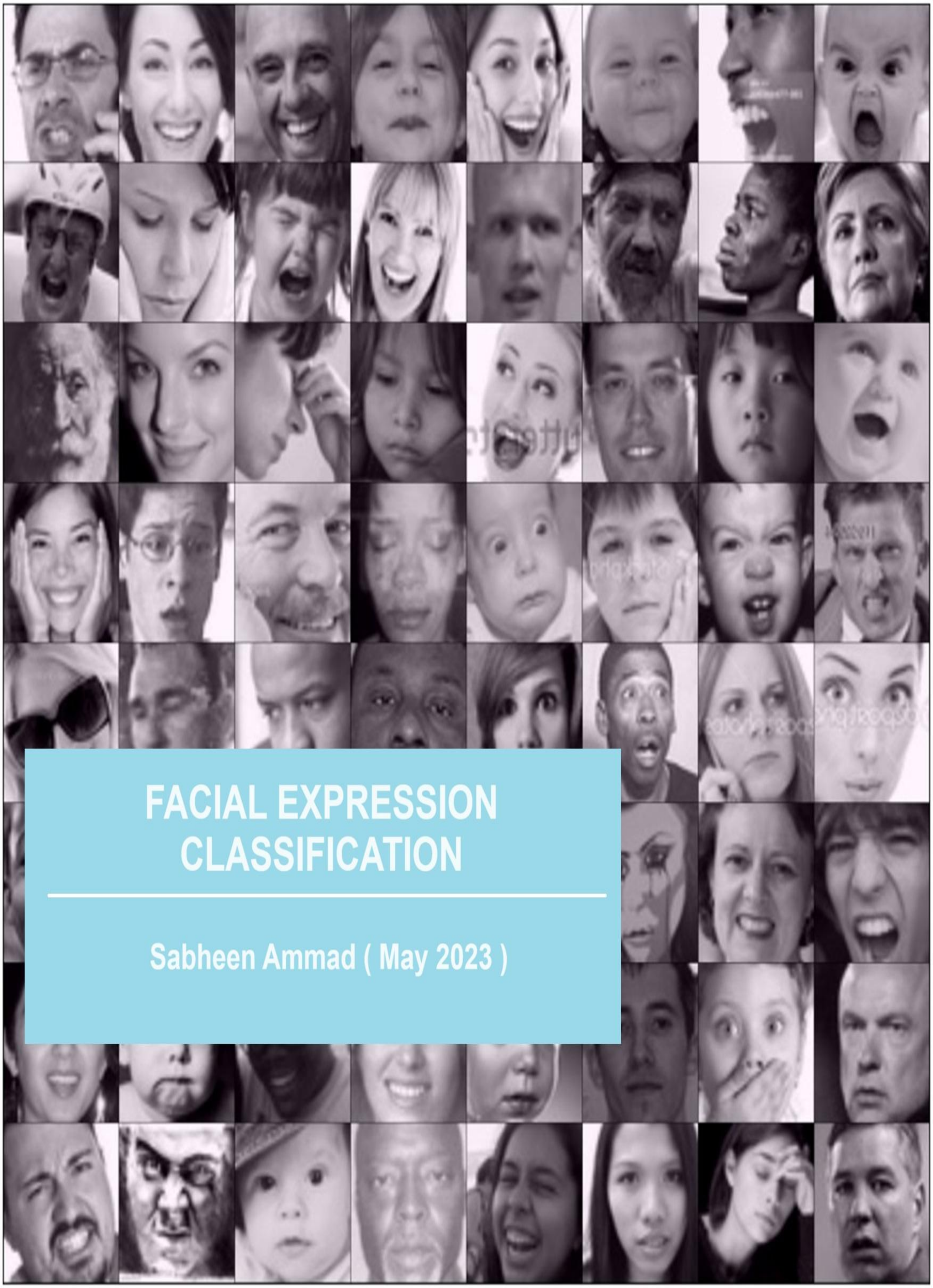


FACIAL EXPRESSION CLASSIFICATION

Sabheen Ammad (May 2023)

INTRODUCTION

Facial Expression Recognition (FER) is a technology that involves the detection and analysis of human facial expressions to identify the emotional state of an individual. The primary goal of FER is to recognize and interpret the various facial expressions that convey emotions such as happiness, sadness, anger, surprise, fear, and disgust. This technology has applications in diverse fields, including human-computer interaction, healthcare, marketing, and security.

Key Steps in Facial Expression Recognition:

- Face Detection
- Feature Extraction
- Classification of Expression

APPLICATIONS

Human-Computer Interaction (HCI):

FER can enhance human-computer interaction by allowing systems to respond to users' emotional states. This can be applied in gaming, virtual reality, and user interface design.

Marketing and Advertising:

Understanding consumer emotions through FER can help marketers tailor advertisements and campaigns to elicit specific emotional responses. This can improve the effectiveness of advertising strategies.

Healthcare:

FER can be used in healthcare for monitoring and diagnosing mental health conditions. It has potential applications in detecting conditions such as depression and anxiety based on facial expressions.

Education:

In educational settings, FER can be used to assess students' engagement and emotional states. It can provide insights into how well students are grasping the material and whether they are engaged in the learning process.

Human-Resource Management:

FER can be used in recruitment processes to analyze candidates' emotional intelligence and suitability for specific roles. It can also be applied in employee monitoring for well-being and job satisfaction.

Security and Surveillance:

FER can enhance security systems by detecting suspicious or unusual behavior based on facial expressions. It can be applied in airports, public spaces, and other security-sensitive areas.

Autonomous Vehicles:

FER can contribute to the development of more intelligent and adaptive autonomous vehicles. Understanding the emotional state of passengers can improve safety and comfort.

Customer Service:

FER can be used in customer service applications to analyze the emotions of users during interactions. This can help businesses provide more personalized and empathetic customer support.

Data Wrangling and EDA

https://www.kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge/data?select=icml_face_data.csv

- FER2013 is a well-established dataset used for facial expression recognition tasks . It was developed by Pierre-Luc Carrier and Aaron Courville and introduced in a Kaggle competition with the intention to promote researchers to improve FER systems. Google Image search API was used for the generation of this dataset.

- Dataset consists of 35887 images and having two columns name “Emotion” and “Pixels”

- The “emotion” field from the dataset is the target attribute which consists of six emotions namely
    - “Anger” : 0,
    - “Disgust” : 1,
    - “Fear” : 2,
    - “Happy” : 3,
    - “Sad” : 4,
    - “Surprise” : 5,
    - “Neutral” : 6.

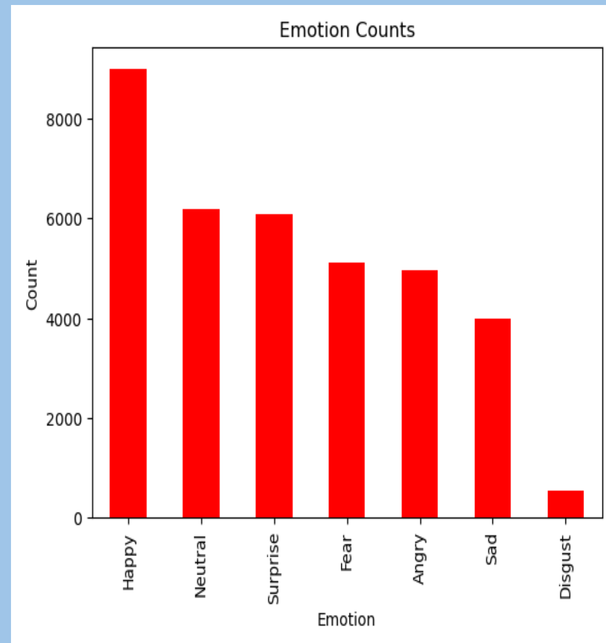
- The “pixel” field consists of a 48*48 facial images. It is stored as a flattened 1-dimensional string.

- The "pixel" field in the dataset represents 48x48 facial images stored as flattened 1-dimensional strings. To facilitate analysis, we have converted the "Pixel" column from string to array format.

- Fortunately, there are no null values present in either column.

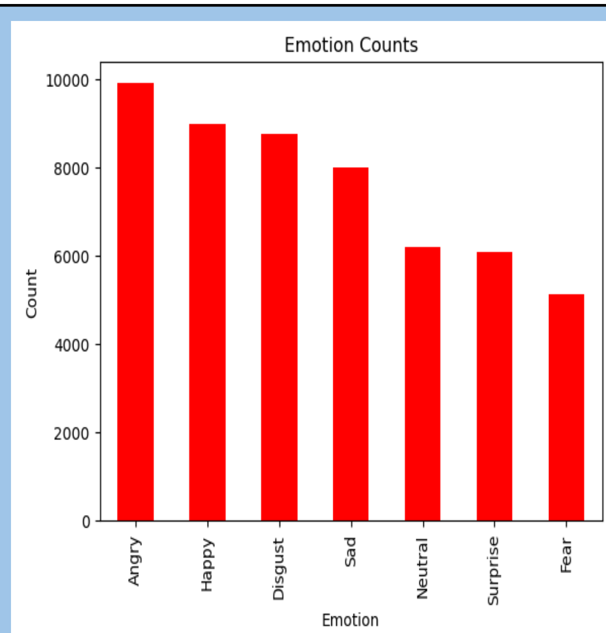
- Table 1 reveals a significant imbalance in the data distribution. To address this issue, we deem it essential to undertake image augmentation techniques to enhance the diversity of the dataset and mitigate the impact of class imbalances.

	Count	Distribution
Happy	8989	25.048068
Neutral	6198	17.270878
Surprise	6077	16.933709
Fear	5121	14.269791
Angry	4953	13.801655
Sad	4002	11.151671
Disgust	547	1.524229



Before Data Augmentation (Table:1)

	Count	Distribution
Angry	9906	18.674006
Happy	8989	16.945350
Disgust	8752	16.498577
Sad	8004	15.088506
Neutral	6198	11.683978
Surprise	6077	11.455879
Fear	5121	9.653703



After Data Augmentation (Table:2)

Model Selection: Convolutional Neural Networks (CNNs) and ResNet for Facial Expression Classification

Convolutional Neural Networks (CNNs):

For this project ,Facial Expression Classification , I decided to choose Convolutional Neural Networks.Convolutional Neural Networks (CNNs) have revolutionized computer vision tasks, particularly image classification, due to their ability to automatically learn hierarchical features from visual data. Inspired by the human visual system, CNNs are composed of layers that perform convolutional operations to capture spatial hierarchies and local patterns. This makes them well-suited for tasks where understanding visual context is crucial.

In the context of facial expression classification, CNNs excel at capturing intricate facial features, such as subtle changes in expression and spatial relationships between facial components. The architecture consists of convolutional layers that learn relevant filters to detect features like edges, textures, and facial landmarks. Pooling layers help downsample the spatial dimensions, and fully connected layers enable the network to make predictions based on the learned features.

Residual Networks (ResNet):

Residual Networks (ResNet) address challenges in training very deep neural networks by introducing residual connections. These connections enable the learning of residual functions, making it easier for the model to capture fine-grained details and avoid the vanishing gradient problem.

For facial expression classification, the depth of the network is crucial to capture the diverse and intricate features associated with different expressions. ResNet's residual connections facilitate the training of extremely deep networks (e.g., hundreds of layers) without degradation issues. This is particularly advantageous for tasks where nuanced features play a crucial role, such as recognizing subtle changes in facial expressions.

Why CNN and ResNet for Facial Expression Classification:

Hierarchical Feature Learning: CNNs automatically learn hierarchical features, capturing both low-level details and high-level abstractions relevant to facial expressions.

Spatial Hierarchy: Facial expressions exhibit spatial hierarchies, and CNNs are effective at capturing local patterns and spatial relationships critical for accurate classification.

Robustness to Variations: CNNs, including ResNet, have demonstrated robustness to variations in pose, lighting, and facial expressions, making them suitable for real-world scenarios.

Residual Connections for Deep Networks: ResNet's residual connections allow for the training of very deep networks, essential for capturing intricate facial features associated with different expressions.

How it Works:

Input Layer: The network receives flattened facial images as input, representing pixel values.

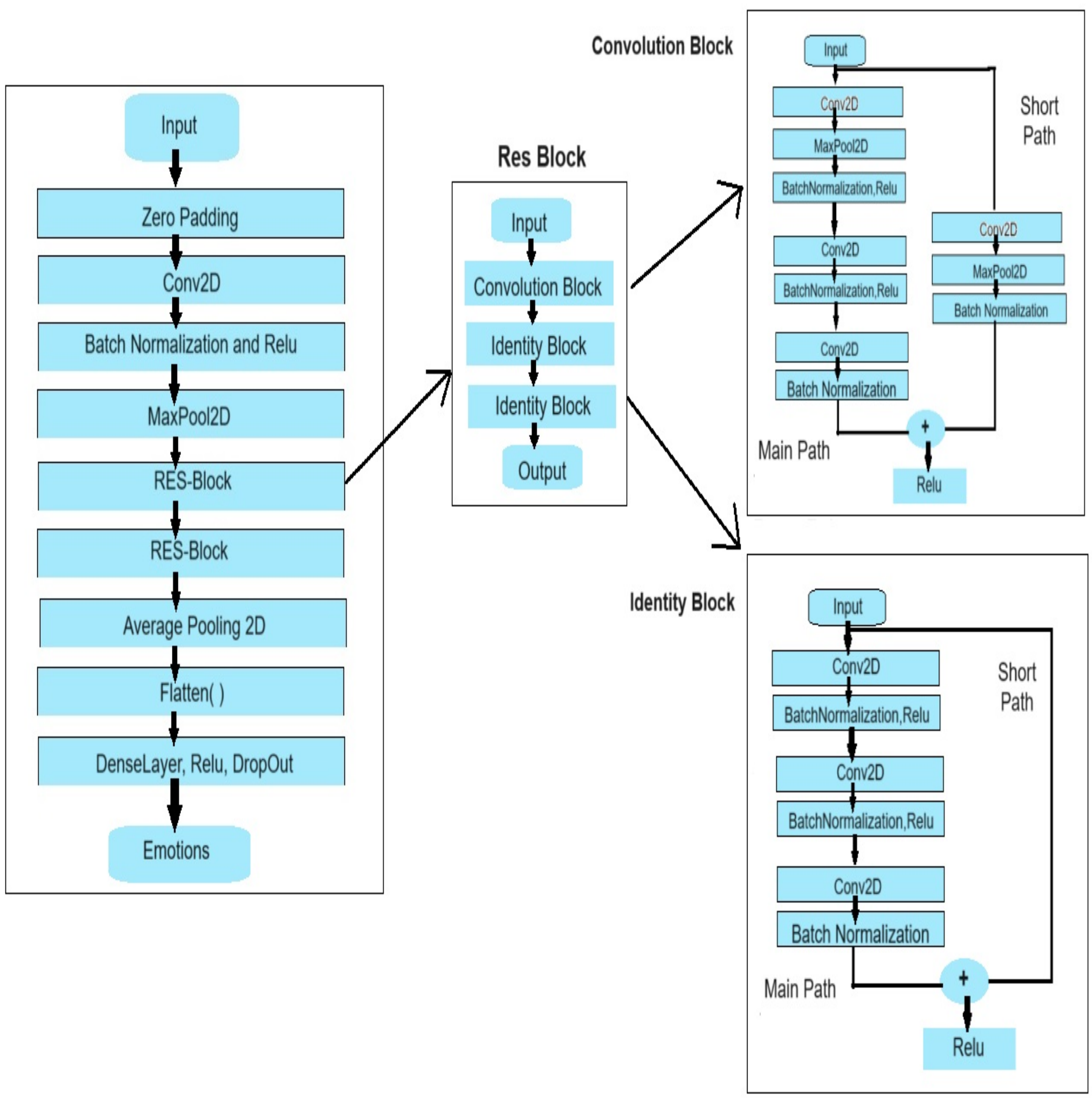
Convolutional Layers: Convolutional layers apply filters to detect features like edges, textures, and facial landmarks.

Residual Blocks: Residual blocks with skip connections allow the network to learn residual functions, facilitating the training of very deep networks.

Pooling Layers: Pooling layers downsample spatial dimensions, retaining relevant information.

Fully Connected Layers: Fully connected layers make predictions based on the learned features, classifying facial expressions.

By leveraging the power of CNNs, specifically ResNet, our facial expression classification model can effectively learn and represent complex patterns, enabling accurate and robust recognition of diverse facial expressions in real-world scenarios.



Interpreting Training and Validation Accuracy/Loss in a Facial Expression Classification

Understanding Accuracy:

Training Accuracy represents the accuracy of the model on the training dataset. It shows how well the model is learning from the training data over epochs whereas Validation Accuracy reflects the model's performance on a separate validation dataset not seen during training

A steadily increasing training accuracy is a positive sign, indicating that the model is effectively capturing patterns in the training set.

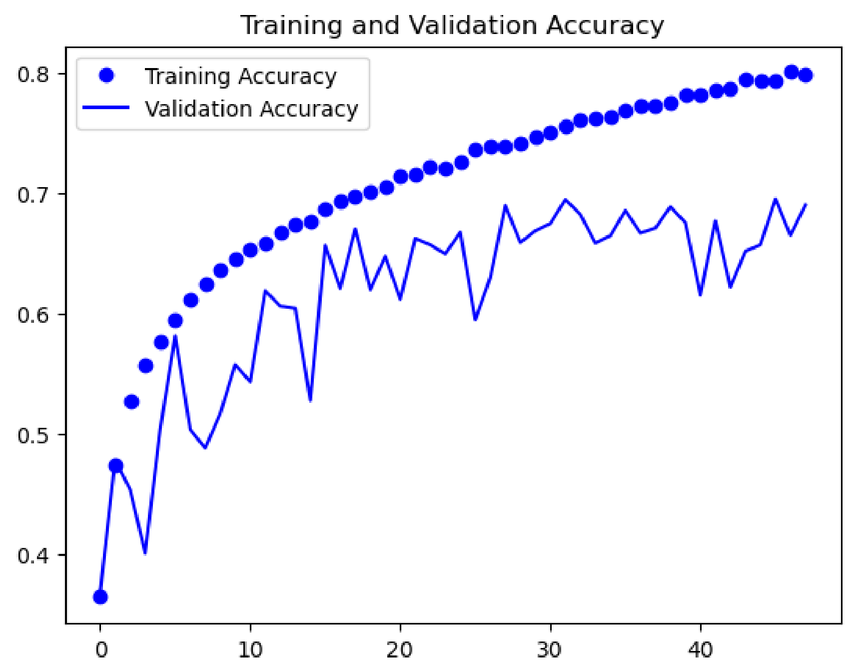
A rising validation accuracy indicates the model's ability to generalize to unseen data. A widening gap between training and validation accuracy may suggest overfitting.

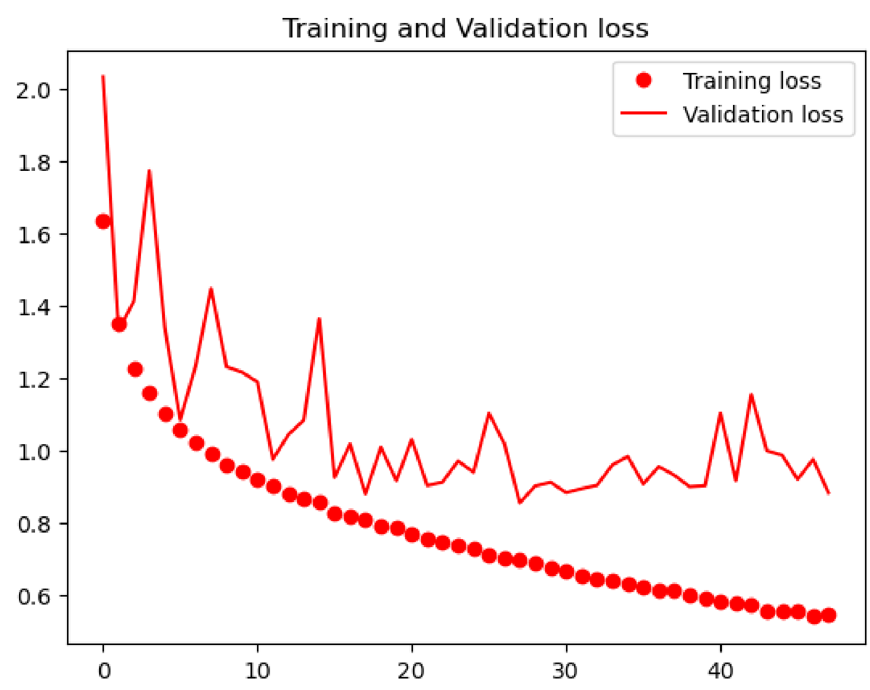
Analyzing Loss:

Training Loss represents the error or loss on the training data and Validation Loss reflects the error or loss on the validation dataset.

As training progresses, the goal is to minimize this value. A decreasing training loss indicates that the model is learning to make better predictions on the training set.

Similar to training loss, the aim is to minimize this value. Monitoring both training and validation loss helps assess whether the model is overfitting or underfitting.


Training and Validation Accuracy
Training Accuracy
Validation Accuracy
0.8
0.7
0.6
0.5
0.4
0
10
20
30
40

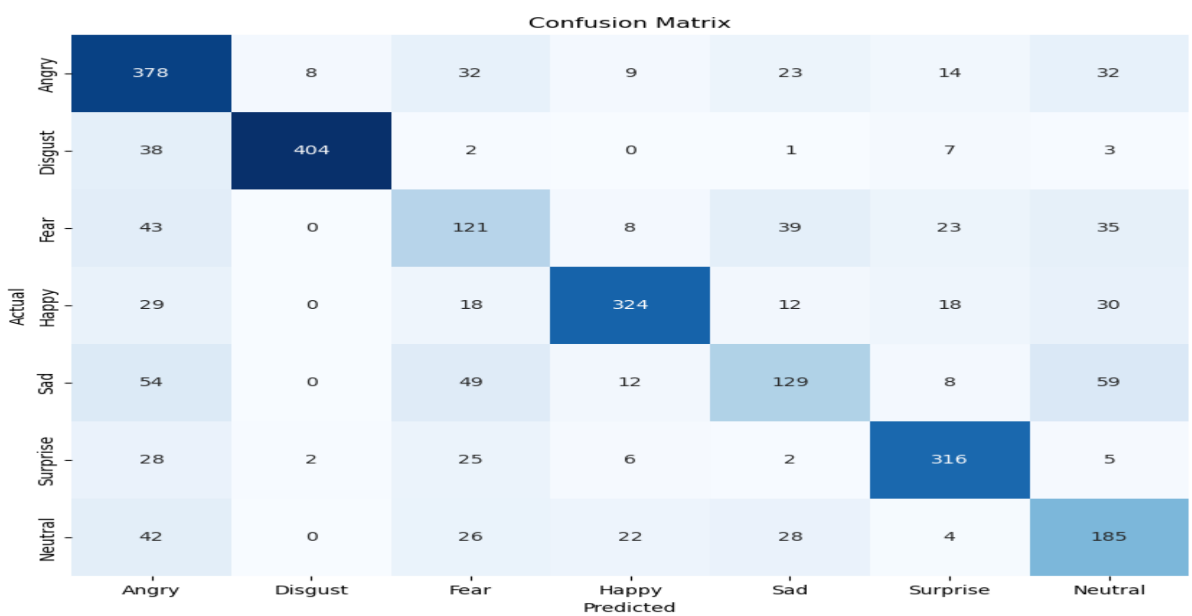

Training and Validation loss
Training loss
Validation loss
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0
10
20
30
40

CLASSIFICATION REPORT

The classification report is a valuable tool for assessing the performance of a classification model. It provides a detailed breakdown of various metrics that evaluate the model's ability to correctly classify instances for each class.

Here after training the model with 40 EPOCHs we received almost 70% accuracy which is not bad overall.

	precision	recall	f1-score	support
0	0.62	0.76	0.68	496
1	0.98	0.89	0.93	455
2	0.44	0.45	0.45	269
3	0.85	0.75	0.80	431
4	0.55	0.41	0.47	311
5	0.81	0.82	0.82	384
6	0.53	0.60	0.56	307
accuracy			0.70	2653
macro avg	0.68	0.67	0.67	2653
weighted avg	0.71	0.70	0.70	2653



FORECASTING

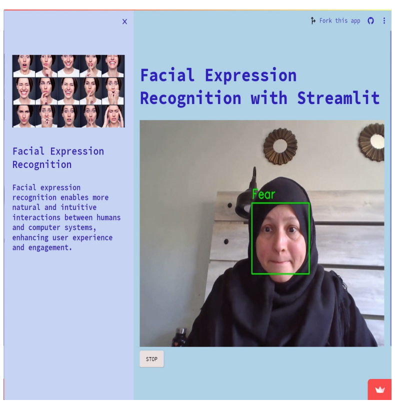
We visualize some images from our dataset with actual and predicted values

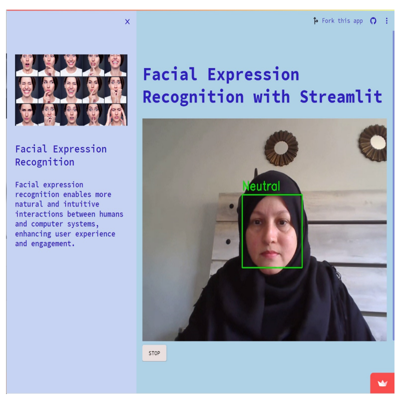
ML MODEL DEPLOYMENT

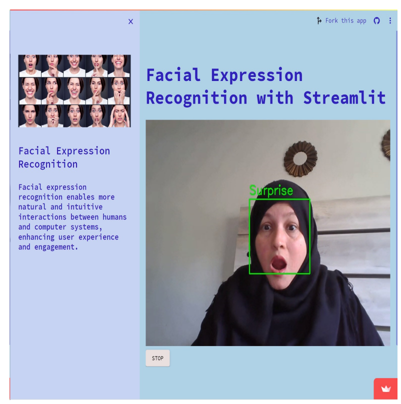
I successfully deployed the Facial Expression Classification project on the Streamlit platform, leveraging the streamlit-webrtc library for enhanced functionality. This integration allows real-time video processing, enabling users to interact with the facial expression model seamlessly. The deployment on Streamlit not only provides an intuitive user interface but also incorporates the dynamic capabilities of the streamlit-webrtc library for efficient webcam access during model inference.

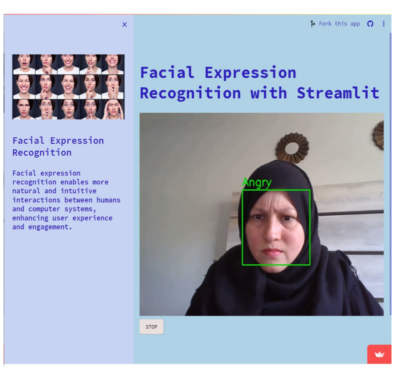
Project Link on Streamlit

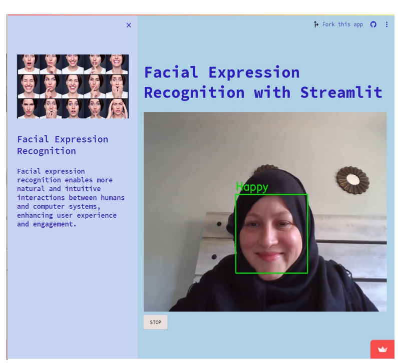
https://firstappapp-h7puengxkucbekipr46ia3.streamlit.app/











FUTURE RESEARCH OR IMPROVEMENTS

- An avenue for future exploration involves assessing the performance of different ResNet architectures for Facial Expression Recognition. Comparing and contrasting the outcomes of distinct ResNet models, such as ResNet-50 and ResNet-101, can provide valuable insights into the impact of network depth on recognition accuracy

- Additionally, expanding the dataset to encompass a broader range of demographics and expressions will contribute to a more robust and inclusive model.