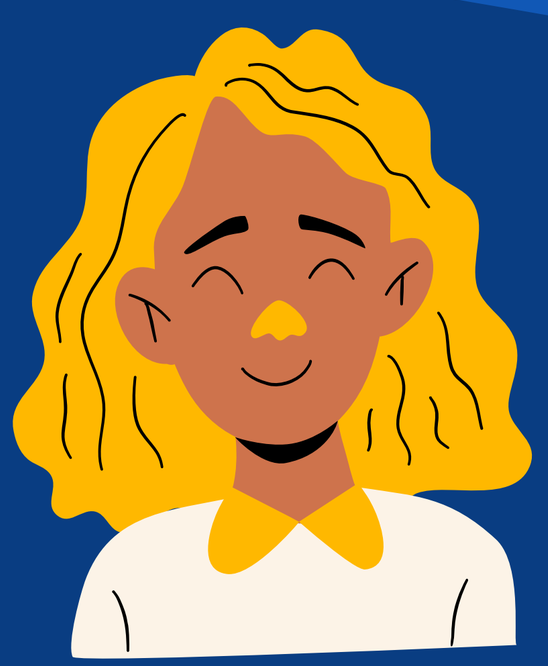
EmotiVision

Real-time Emotion Recognition through Dynamic Webcam Analysis using Deep Learning and OpenCV

by - Ananya Arora

Introduction



- **EmotiVision:** a significant project in the realm of CV and DL, focuses on the precise recognition of human facial expressions.
- **Significance:** Facial expressions are a fundamental mode of non-verbal communication, making EmotiVision pivotal for understanding human emotions.
- **Primary Goal:** real-time identification and classification of diverse facial expressions and providing insights.
- **Tech Stack used:** CNN , OpenCV

Let's Check it Out

(Demo)

Methodology

Step 1

Data Acquisition:

- FER2013 dataset from Kaggle.
- Each label corresponds to a specific emotion category (0-6)
- Anger, Fear, Neutral, Sad, Happy, Disgust, Surprised

Step 2

Dataset Preparation:

- Split the data in 80:10:10, for train, validation and test respectively.
- The data involved varies with every iteration.

Step 3

Data Pre-processing:

Using CRNO:

- convert strings to lists of integers
- reshape and normalise grayscale image with 255.0
- one-hot encoding label, e.g. class 3 to [0,0,0,1,0,0,0]

Model training

Step 1

Feature Extraction:

- CNN architecture consisting of multiple layers.
- analyze input facial images, using filters to identify essential patterns, edges etc.

Step 2

Regularization and Batch Normalization:

- Batch normalization layers are used after each convolutional and dense layer to prevent overfitting and enhance the model's generalization capabilities.
- normalize input to each layer, reducing internal covariate shift, stabilizing and accelerating the training process

Step 3

Flattending and Dense Layers:

- flattening operation to convert the 3D output to 1D
- Three densely connected layers that reduce the number of neurons
- These layers extract high-level features and relationships from input data, improving understanding of facial expressions.

Step 4

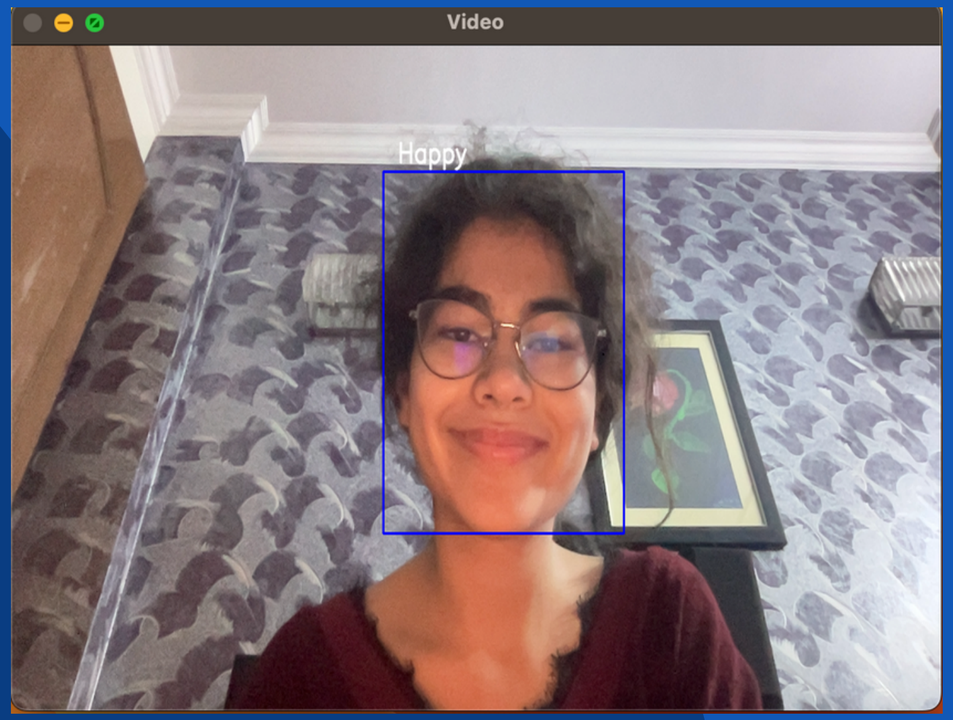
Softmax Activation and Output Layer:

- predict the probability of each emotion class.
- Categorical cross-entropy measures the difference between predictions and true labels.
- The Adam optimizer is used for efficient weight updates, with hyperparameters tuned for optimal convergence.

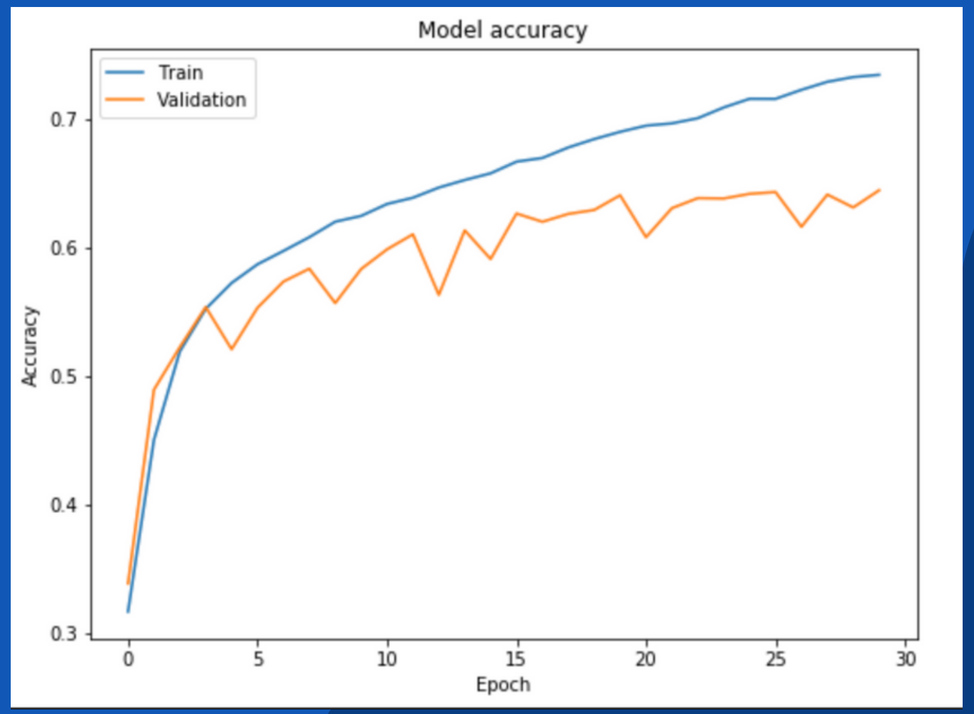
Image Processing

- Opted Real-Time Evaluation
- **Libraries used:** TensorFlow, Keras, NumPy, OpenCV and Haar Cascades
- **Workflow:**
    1. capturing expressions through a webcam
    2. using Haar cascades for initial detection
    3. processing images with a pre-trained model
    4. predicting emotions
    5. retrieving the emotion label from a predefined list based on predicted class index.

Diagrams



Model testing using WebCam



Training and Validation Accuracy

Future Scope

- A successful model having multi-layered Convolutional CNN architecture has been developed with accuracy of **65%.**

- **Future Scope:**

1. expanding the dataset to include : a range of facial expressions, variations in lighting , subtle facial gesture movements
2. exploring the integration of speech recognition

- EmotiVision promotes empathy in human-computer interactions, aligning with the broader goal of advancing emotionally intelligent technology.