Facial Emotion Detection using CNN

Deekshita Rajesh

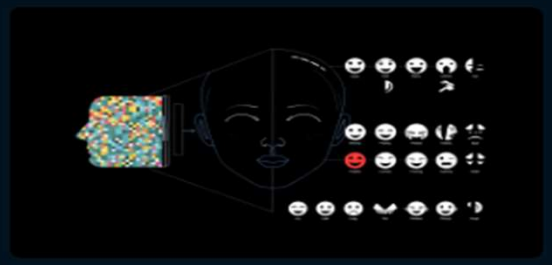
4al22cd012@aiet.org.in

Contents

Introduction to Facial Emotion Recognition

Facial Emotion Recognition (FER) is a technology that leverages computer vision and deep learning to automatically identify human emotions from facial images. This presentation delves into the use of Convolutional Neural Networks (CNNs) for real-time emotion detection, covering the model architecture, dataset, training process, evaluation, live demonstration, and future work. The presented approach highlights how deep learning has advanced the field of emotion recognition and enabled practical, real-time applications.

Understanding Convolutional Neural Networks (CNNs)



Definition and Purpose

A Convolutional Neural Network (CNN) is a specialized deep learning model designed to effectively process image data. It excels at learning and extracting features such as edges, textures, and objects from images, making it highly suitable for computer vision tasks like facial emotion recognition.

Key Layers in a CNN

Common layers in CNNs include the Input Layer (accepting 3D image inputs), Convolution Layer (applying filters to extract spatial features), Activation Layer (usually ReLU to add non-linearity), Pooling Layer (such as Max Pooling for downsampling), Fully Connected Layer (classic neural network layer), and Output Layer (using Softmax for classification).

Dataset: FER-2013

Composition and Structure

The FER-2013 dataset consists of over 35,000 grayscale facial images, each of 48x48 pixel resolution. Each image is labeled as one of seven emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, or Neutral.

Dataset Split

The dataset is organized into training, validation, and test sets, facilitating robust model training and fair evaluation.

Model Architecture

The CNN model was constructed using the Keras Sequential API. The architecture features multiple Conv2D layers for extracting image features, MaxPooling2D layers for reducing feature dimensions, followed by Flatten and Dense layers for classification. Dropout was implemented for regularization, and the output layer uses Softmax to predict one of seven emotion classes. This layered approach enables efficient and effective recognition of diverse facial expressions.

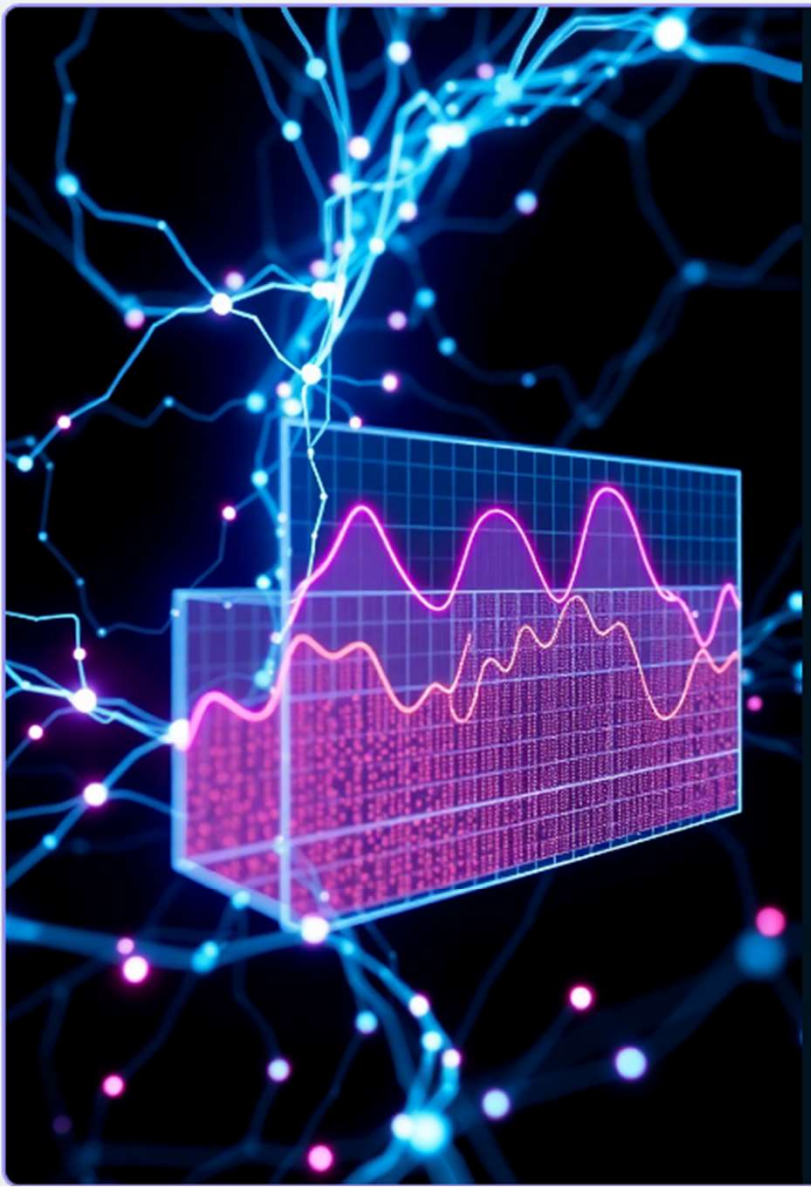
Training Details

Training Settings

The CNN model was trained using the Adam optimizer and Categorical Crossentropy as the loss function. ModelCheckpoint was employed as a callback to save the best-performing model checkpoint during training.

Process Overview

Training involved running for several epochs, with both training and validation accuracy and loss monitored. The results were visualized to assess model performance and potential overfitting throughout the training period.



Evaluation Metrics and Results

1 **Performance Metrics**

Model evaluation involved plotting accuracy and loss graphs, generating a confusion matrix to compare actual labels with predictions, and producing a classification report with precision, recall, and F1-score.

2 **Model Accuracy**

The CNN achieved around 90% accuracy on the training set and 65–70% on the validation set, demonstrating strong learning with some generalization limitations.

Prediction Results

The trained model was tested on 10 randomly selected test images. Each image's predicted emotion label was directly compared with its actual label, providing clear insights into the model's strengths and weaknesses. The model performed particularly well in recognizing emotions like Happy, Angry, and Neutral, affirming its reliability in common scenarios encountered in FER tasks.

Webcam Live Demo

1

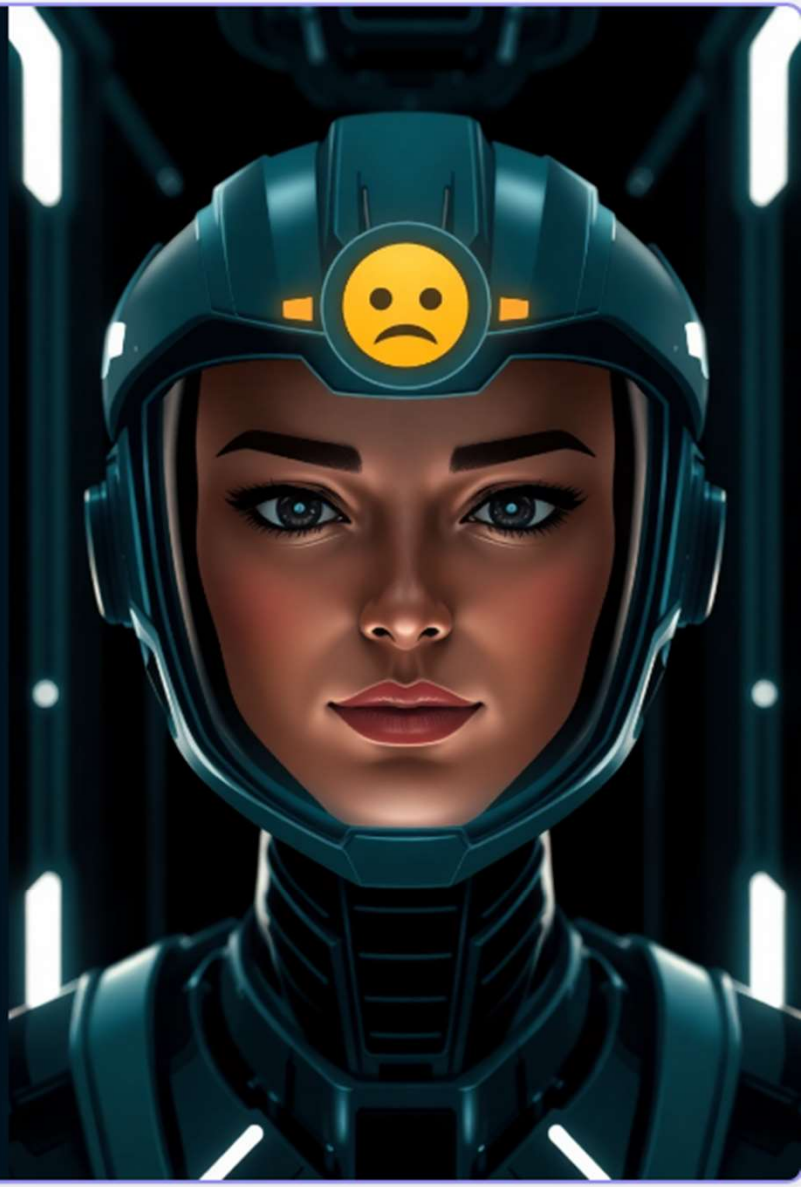
2

Real-time Prediction

Using OpenCV, the system captures real-time face images from a webcam, processes each frame into a 48x48 grayscale image, and predicts emotions using the trained CNN model.

Interactive Feedback

The predicted emotion is displayed directly on the live video feed, allowing users to see immediate results. The demo supports natural presenter interaction, enhancing engagement and realism.



Audio Feature

Although the implemented CNN model does not utilize audio data, the live demo was conducted using a webcam equipped with a microphone. This setup enabled the presenter to speak and respond naturally, improving the interactivity and human-like experience during the demonstration, even though only visual data were used for emotion prediction.

Conclusion and Future Work

Key Achievements

The project successfully built and deployed a CNN-based model capable of real-time facial emotion recognition, demonstrating strong performance on both static images and live video feeds.

Opportunities for Enhancement

Future improvements include boosting accuracy with advanced techniques such as Transfer Learning (using models like MobileNet or ResNet), integrating multi-modal data (audio or text), and deploying the model for web applications using frameworks like Streamlit or Flask.