Project Write-up

Facial key point detection and recognition.

March 2025

CS 419/619 Computer Vision

Course Instructor: Dr. Surya Prakash

Indian Institute of Technology Indore

Submitted By:

Gorantla Anuksha (210001018)

Jilagam Dinesh (210001026)

Objectives

The goal of this project is to predict facial key points on images using computer vision techniques. Each key point is represented by an (x, y) coordinate in pixel space. The model should accurately detect 15 predefined key points corresponding to facial features, namely left_eye_center, right_eye_center, left_eye_inner_corner, left_eye_outer_corner, right_eye_inner_corner, right_eye_outer_corner, left_eyebrow_inner_end, left_eyebrow_outer_end, right_eyebrow_inner_end, right_eyebrow_outer_end, nose_tip, mouth_left_corner, mouth_right_corner, mouth_center_top_lip, mouth_center_bottom_lip.

Scope

- Preprocessing image data and handling missing key points.
- Building a model to accurately predict key points.
- Evaluating model performance and optimizing accuracy.
- Applying the trained model to test images for key point detection.

Dataset Description

The dataset, obtained from Kaggle, consists of grayscale facial images with a resolution of 96×96 pixels. The image data is stored in the last column of the dataset as a row-wise sequence of pixel intensity values ranging from 0 to 255. The dataset is divided into two parts:

- **Training Set:** Contains 7049 images, each annotated with (x, y) coordinates for 15 facial key points along with the corresponding pixel data.
- **Test Set:** Consists of 1783 images without labeled key points. Each row includes an ImageId and the pixel data in a row-ordered format.

Expected Outcomes

- A trained model capable of detecting facial key points with high accuracy.
- Visualization of detected key points on test images.
- A submission file with predicted key points formatted as required.

Performance Metrics

- **Accuracy**: can be defined as a performance metric by setting a threshold for acceptable error in key point predictions. One approach is **Key Point Accuracy** (The percentage of key points predicted within a certain error threshold (e.g., 5 pixels) from the actual location.)

$$Accuracy = \frac{\text{Number of key points with error} \leq \delta}{\text{Total number of key points}} \times 100$$

  where:

  - δ is the predefined threshold (e.g., 5 pixels),
  - A key point is considered correct if $|y_i - \hat{y}_i| \leq \delta$ for both x and y coordinates.

- **Mean Absolute Error (MAE)**: Evaluates the average absolute differences between predictions and ground truth.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

- **Root Mean Squared Error (RMSE)**: Measures the square root of the average squared differences between predicted and actual key point locations. Compared to the Mean Absolute Error, RMSE punishes large errors.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

  where:

  - N is the total number of key points,
  - y_i is the actual key point location,
  - $\hat{y}_i$ is the predicted key point location.