Facial Keypoints Detection Using Deep Learning

Viktoriia Kachanovska

Project Overview

Goal: Detect 15 key facial landmarks (30 coordinates) in grayscale face images

Task Type: Supervised Regression

- Input: 96x96 grayscale image
- Output: 30 (x, y) keypoint values
- Dataset: ~7,000 training samples, partial missing keypoints

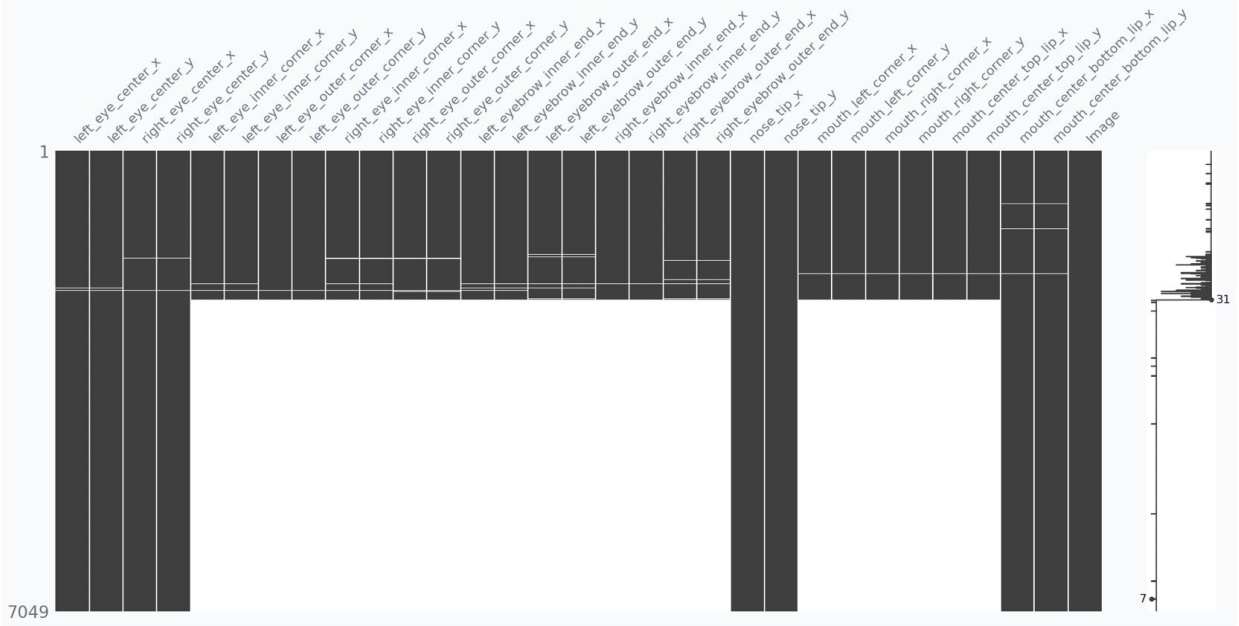
EDA

Missing Values

Missingness is not random.

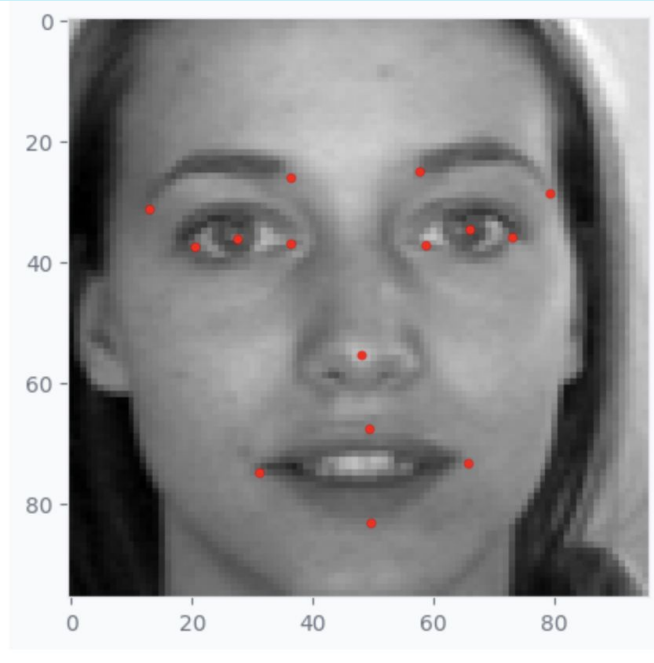
Two dominant missing regions.

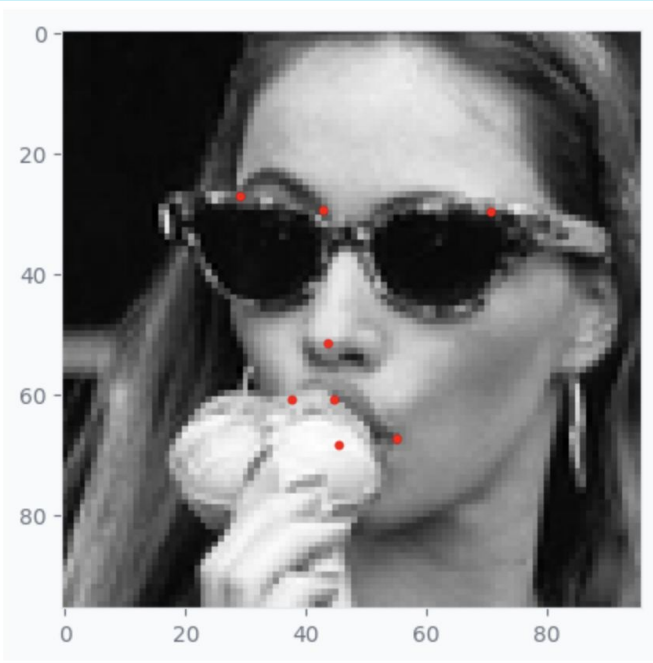
A few samples are fully complete



EDA

Samples

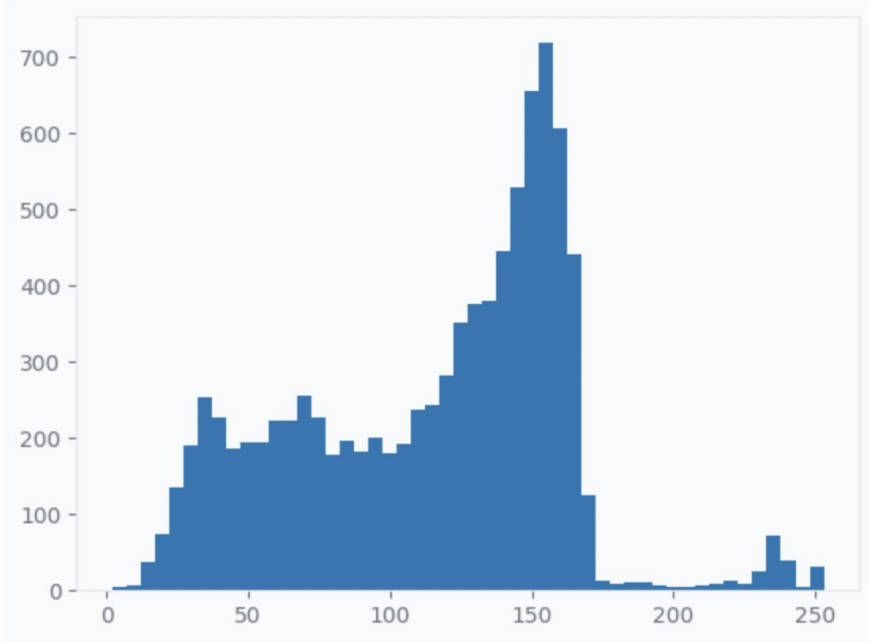




EDA

Pixels Distribution

- Most pixels are mid-range, so the faces are likely well-contrasted against the background.
- No uniform spread → not normalized yet



Data Preprocessing

- Images normalized to [0, 1]
- Converted string to 96x96 float arrays
- Handled missing values

Baseline CNN Model

- 2 Convolutional layers + Pooling
- Flatten + Dense + Dropout (x2)
- Final output: Dense(30) for all keypoints
- Loss: MSE, Metric: RMSE

Results: Quick convergence, some overfitting on validation

Region-Specific Modeling

- Split keypoints into regions:

1. Eyes & Nose (14 outputs)
2. Mouth & Eyebrows (16 outputs)

- Trained separate CNNs per region

Benefits:

- Better use of partially labeled data
- Lower validation RMSE
- Reduced overfitting

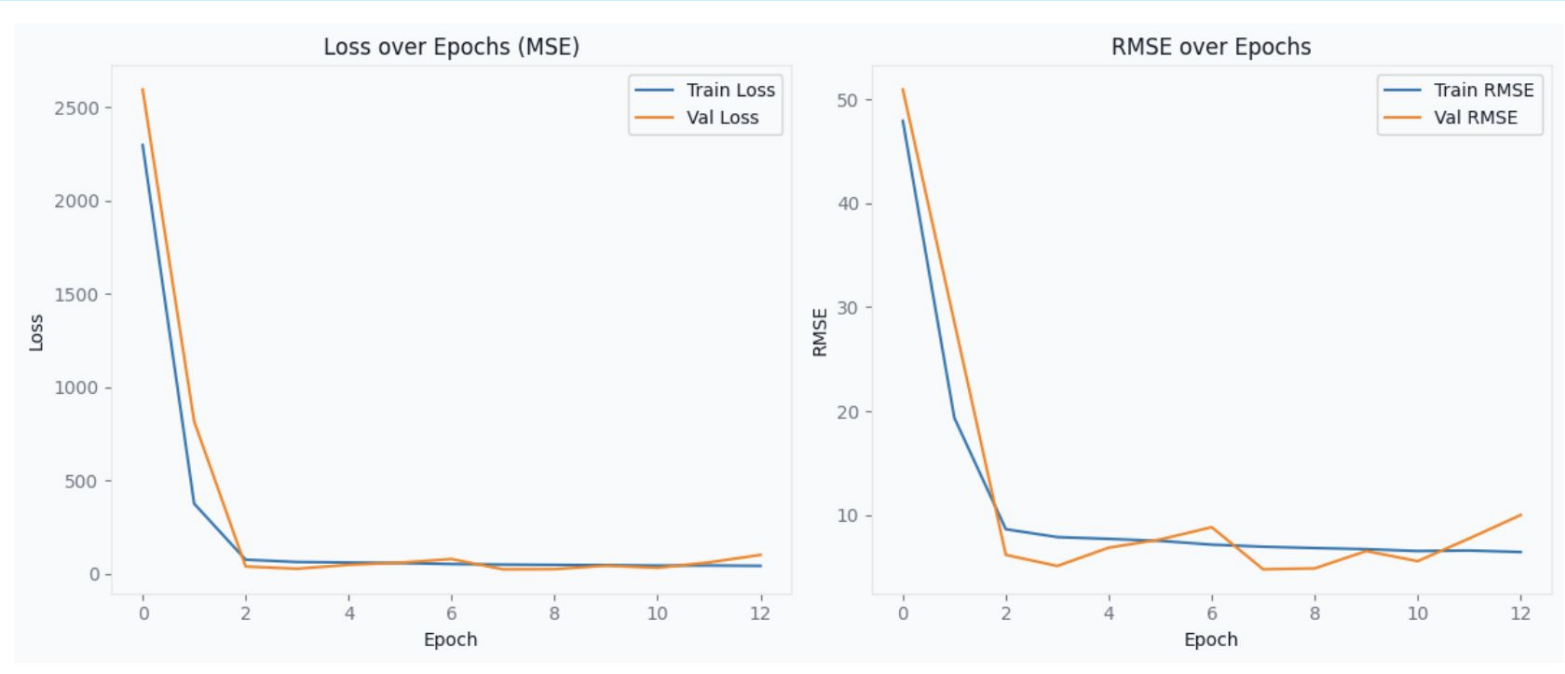
Multi-Branch CNN

- Shared input image
- Two convolutional branches (per region)
- Merged outputs into final Dense(30)
- Modular, interpretable, scalable

Result: Balanced performance and flexibility

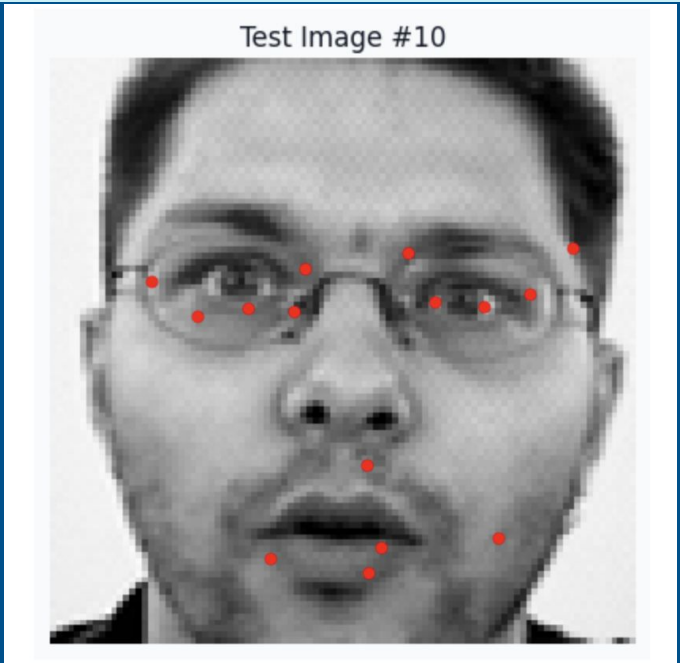
Training History Comparison

Baseline CNN



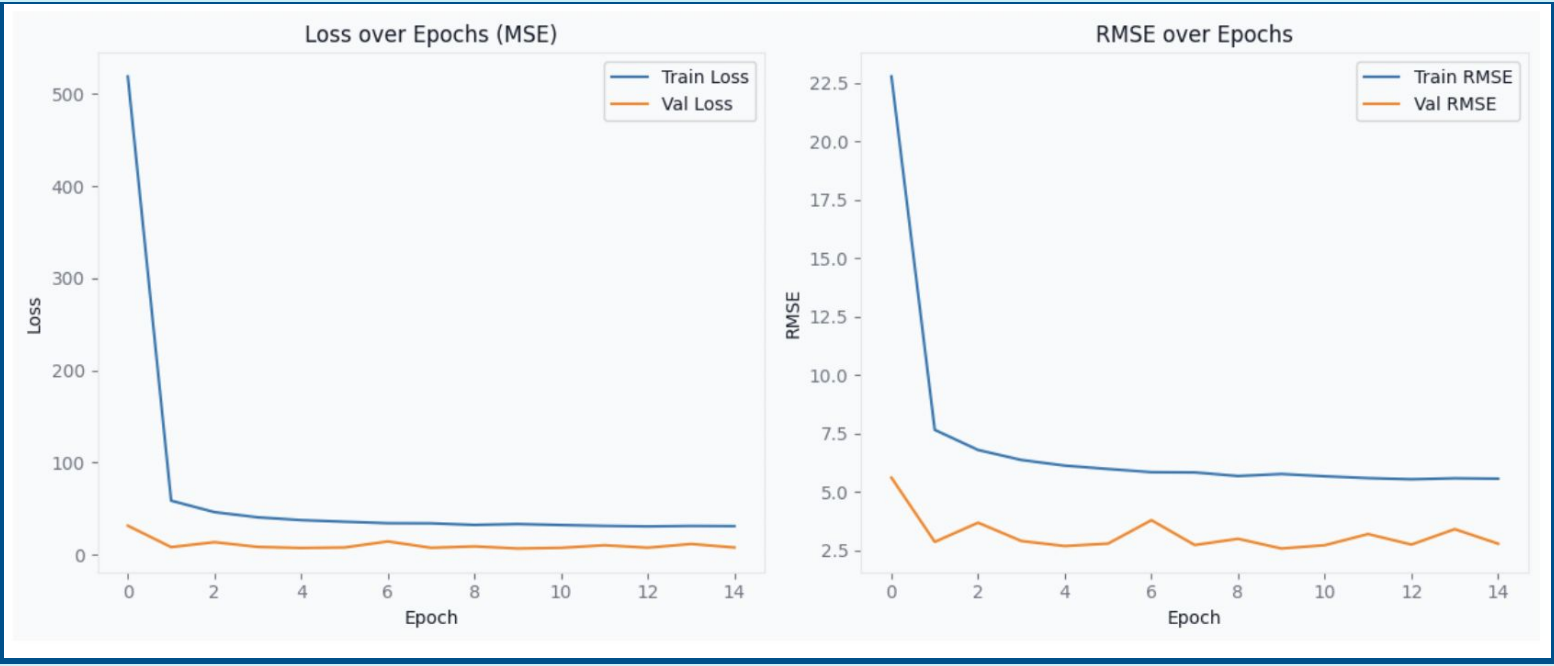
Training History Comparison

Baseline CNN



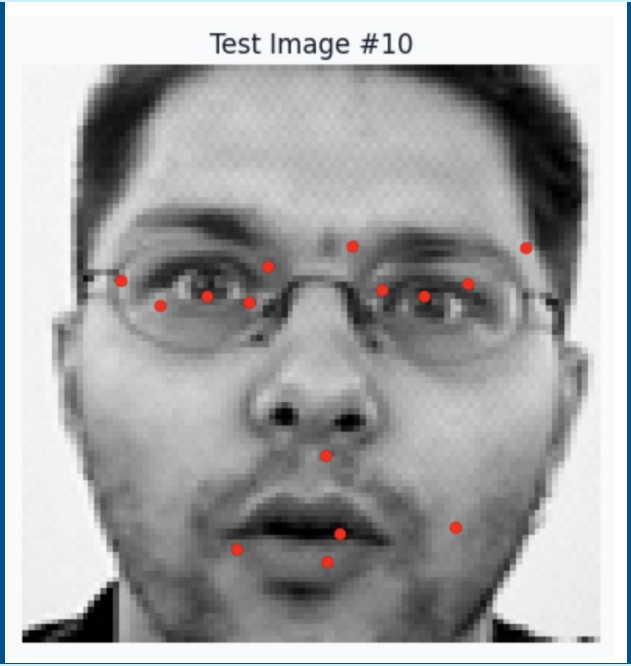
Training History Comparison

Region-Specific CNNs



Training History Comparison

Region-Specific CNNs



Conclusions

Modeling Insights:

- Baseline CNN models performed well initially. However, slight overfitting was observed without regularization.
- Introducing region-specific models significantly improved generalization. These models had lower and more stable validation RMSE.

With further work, including:

- landmark refinement techniques (e.g., heatmap-based regression)
- ensembling multiple specialized models ...the accuracy and robustness of facial keypoint detection can be pushed even further.