A Minimal Real-Time Facial Expression Recognition System: Lightweight CNN on FER-2013 with OpenCV-based Inference

1 Abstract

We propose a lightweight, real-time facial expression recognition (FER) system trained on FER-2013. With OpenCV, we implement face detection, alignment, and per-frame emotion inference, and assess robustness and fairness in real-world settings. The outcome will be a reproducible training pipeline and deployable prototype balancing accuracy, latency, and portability.

2 Introduction

Real-time affect recognition is valuable for HCI, education feedback, and edge applications. State-of-the-art FER often relies on large backbones and high compute, limiting privacy-preserving, low-latency on-device deployment. FER-2013 (48×48 grayscale, 7 classes) is a common benchmark but suffers from label noise, low resolution, and distribution shift. We target an end-to-end approach centered on a small CNN with robustness-oriented training and deployment optimization, validated in live camera scenarios.

3 Research Questions

RQ1 Can a lightweight CNN achieve robust accuracy on FER-2013 without large backbones?
RQ2 Which augmentations and loss designs mitigate label noise and class confusion ?
RQ3 How stable is performance across illumination, pose, and demographic proxies, and what biases emerge?

4 Related Work

FER datasets and baselines. FER-2013 (48×48 grayscale, 7 classes) is a common benchmark; typical baselines include shallow CNNs and ResNet-18 variants.

Model compression. Pruning, quantization, and knowledge distillation help preserve accuracy on mobile/edge devices.

Detection and alignment. OpenCV Haar/LBP cascades, DNN-based detectors (e.g., Res10 SSD), and face alignment improve downstream classification stability.

Real-world generalization. Mixed data augmentation, contrastive learning, and label cleaning mitigate label noise and domain shift.