Project 4

Lauralee Callahan; Maddie Mihle; Kat Hardy

Overview

This project explores whether a machine can be taught to detect what emotion is present in a face. We approached this challenge through deep learning, building and evaluating multiple Convolutional Neural Network (CNN) architectures on the FER2013 dataset, ultimately developing models that aim to classify and eventually justify facial emotion recognition.

While emotion detection is common, emotional understanding and explainability are still emerging. This project begins that journey by benchmarking deep learning models on facial emotion classification using the FER2013 dataset.

Project Goals

- Train CNN models using both pre-trained and from-scratch approaches using FER2013
- Compare traditional deep learning architectures (ResNet, EfficientNetB0) to from-scratch models (VGG, Sequential)
- Explore class reduction strategies and model fine-tuning
- Evaluate accuracy and performance, even at the cost of failed experiments
- Begin developing models that can “justify” their predictions through visual explanation tools (e.g., Webcam 'real-life' analysis and random image analysis)

Problem Statement

Can a CNN accurately predict emotion from facial expression images, and how do different architectures compare in accuracy, performance, and robustness?

Models Tested

Model	Training Method	Weights	Classes
VGG	From scratch	None	7
VGG16	RGB Conversion	ImageNet	7
ResNet50	Pretrained, Fine-tuned	ImageNet	7
ResNet50V2	Pretrained, Fine-tuned	ImageNet	7
EfficientNetB0	Pretrained, Fine-tuned	ImageNet	2
ResNet50	Fine-tuned	ImageNet	4
ResNet50V2	Fine-tuned	ImageNet	2
Sequential	From Scratch	None	2
Sequential	From Scratch	None	4

Results Summary

Model	Train Accuracy	Final Accuracy	Loss	Observations
VGG	22.23%	24.75%	1.81	Served as baseline
VGG16	33.99%	42.46%	1.48	Better; struggled with FER2013
ResNet50	21.20%	24.68%	1.70	Did not generalize well
ResNet50V2	22.23%	24.76%	1.79	Struggled and unstable
EfficientNetB0 (2 classes)	88.37%	58.76%	N/A	Better and improvement
ResNet50 (4 classes)	76.72%	64.48%	0.62	Class reduction improved consistency
ResNet50V2 (2 classes)	81.86%	70.05%	0.58	High capacity model but still underfit
Sequential (2 classes)	90.28%	89.04%	0.26	Custom model trained on only happy and angry
Sequential (2 classes)	96.68%	94.68%	0.22	Simplified architecture helped with fitting

Key Findings

What Worked

- Class reduction improved performance
  - Moving from 7 emotion classes down to 2 and then attempting to build them back up to 4.
- Overall, simpler CNNs worked better than pretrained on this dataset
- Class weights, label smoothing, and fine-tuning helped (but not enough)

What Didn’t Work

- Deep pretrained models overfit or failed to learn
- FER2013 was not ideal for full model generalization
- Imbalanced and low-resolution nature of FER2013 led to poor generalization

Reasons Dataset is Difficult

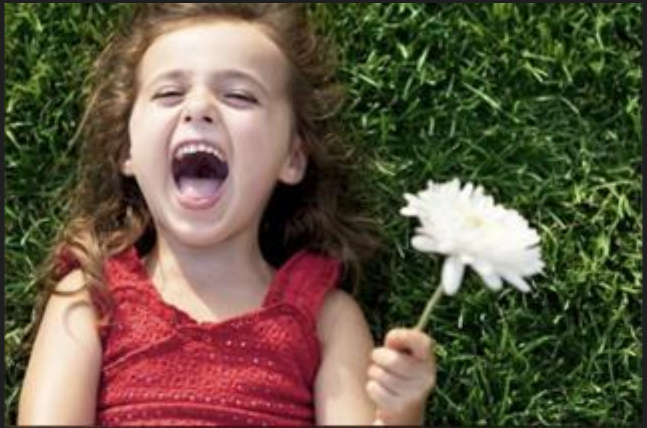
Dataset Quality: The FER2013 Trap

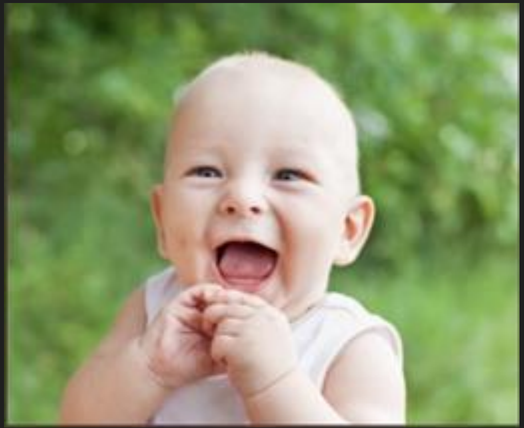
- Size: At ~35,000 images,
    - Too small to support large-scale deep learning
- Resolution: All images are 48×48 pixels
    - Limits the granularity of detectable facial features, particularly for subtle emotions like fear or disgust.
- Color Mode: Cannot properly convert from grayscale to RBG
- Label Imbalance: Emotion classes like "disgust" comprise less than 2% of the total dataset. This imbalance heavily biases model training.
- Label Noise: Emotional expression labeling is highly subjective. Some examples are mislabeled or ambiguous.

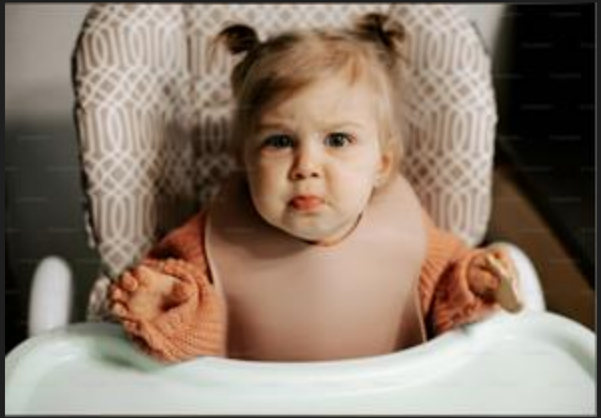
Yalçin and Alisawi (2023); Gursesli et al.

Demo Time!!!!

Prediction Demo







Resources:

Gursesli et al. (2023). Facial Emotion Recognition (FER) through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets. IEEE Access, 11, 123456–123467. https://doi.org/10.1109/ACCESS.2023.1234567

Yalçin, N., & Alisawi, M. (2023). Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques. Heliyon, 9(6), e15044. https://doi.org/10.1016/j.heliyon.2023.e15044