Project Proposal Title: Automated Classification of Human Facial Expressions Using Deep Learning

Author:
Mathilde Cros (mathildecros@g.harvard.edu)

Data:
This project will use the FER-2013 dataset from Kaggle (https://www.kaggle.com/datasets/msambare/fer2013), which consists of 48x48 pixel grayscale images of human faces, which are approximatively centered and of about the same size. These are labeled into 7 categories: angry, disgust, fear, happy, sad, neutral, surprise. There are a total of 28,709 training images and 3,589.

Background and Motivation:
Emotion recognition from facial expressions is crucial in human-computer interaction or in recent advances for mental health monitoring (AI-driven therapy, virtual assistants…), so this project aims to improve the accuracy and robustness of automated facial emotion classification to help with ongoing research in computer vision and psychological studies on emotion detection.

Problem:
The goal of this project is to develop a deep learning model capable of accurately classifying human facial expressions into one of the 7 emotional categories. A key challenge is ensuring that the model generalises well across different lighting conditions, facial orientations, and demographic variations. Additionally, recognising subtle differences between emotions like fear and sadness or happy and surprise remains a difficult task for humans, so it might (or not!) be for computers…

Scope and Methods:
Using Convolutional Neural Networks as the backbone and initial performance benchmarking for this project's model, it is then possible to experiment with different architectures such as:

- Pre-trained CNNs (e.g., VGG16, ResNet, EfficientNet) for transfer learning
- Data Augmentation (e.g., rotation, flipping, brightness adjustments) for generalisation
- Facial landmark extraction for better feature representation
- Attention mechanisms or Vision Transformers (ViTs)

Concerns & Limitations:
Ethical Considerations: Could this model be biased across different demographic groups? What is the impact of emotion recognition models deployed at a larger scale in society?

Data Quality and Imbalance: The class for "disgust" has significantly less images, so there might be some data augmentation or cleaning to do. In general there might be some additional preprocessing necessary for the quality of images as well, even though it looks of good quality at a first glance.

Computational Resources: Training deep learning models on image data can be computationally expensive, so it might be necessary to use GPU acceleration through Google Colab for example.