MAIS 202 - PROJECT DELIVERABLE 1

Léa Kassab

1. Choose your dataset:

This project aims to detect the emotion displayed by a human in real-time. The goal is to recognize how a set of facial traits are positioned (eyebrows raised, eyes narrowed, lips curved upwards...) and based on that, decide what the person is displaying.

Thus my input is going to be live video feed, my model will use deep learning, and the type of problem I am solving is a classification problem.

My aim is to recreate the project done in the following paper:
Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network, Shervin Minaee, Amirali Abdolrashidi,2019
and the paper can be found on this url: https://arxiv.org/pdf/1902.01019v1.pdf

The dataset I will train my model is the FER2013 dataset from Kaggle. Fer2013 contains approximately 30,000 facial RGB images of different expressions with size restricted to 48×48, and the main labels of it can be divided into 7 types: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral.
The data base's citation is *The Facial Expression Recognition 2013 (FER-2013) Dataset, Pierre-Luc Carrier and Aaron Courville*
and it can be found on this url: https://www.kaggle.com/msambare/fer2013

2. Methodology:

a. Data Preprocessing:

The training set consists of 28,709 examples and the public test set consists of 3,589 examples.
First, for each of the 7 categories, we don't have the same amount of pictures. This means that the dataset has a class imbalance problem, and I need to fix it with data augmentation.
Second, another problem is intra-class variation, that I need to fix by making sure there is no overfitting for my model.
Finally, the contrast of some images can be too high, and there could be objects (sunglasses or hands) blocking some part of the person's face. A possibility to tackle this would be to assess the contrast and then brighten the picture if it's too high. For the blocks on people's faces, if it's a hand blocking the left side of the face for example: I could program my ML model to copy paste the "clear part" of the face onto the blocked part, but mirror it (so that forms the two sides of the same face). For the glasses, I could paste neutral eyes on it, thus my model will focus on the rest of the parts (nose, muscle contraction, mouth, cheek height, eyebrows) to determine the emotion.

b. Machine learning model:

Since this is a classification problem, I need to use a deep learning model. The input of my dataset are images, which will be understood as numbers, will be read as a matrix of numbers representing RGB values.
Thus my image will be a matrix and to extract features from it (such as expressions on the face i.e. eyebrows raised up) I need to run a convolutional neural network. Then I need to pick the best feature, and to do this I need to input a max-pooling layer, which will also avoid overfitting. Finally, after implementing a DropOut layer (which drops nodes in the neural network to avoid overfitting), I need to add a "fully connected layer" whose last layer is comprised of 7 nodes because we have 7 classes representing our different emotions/expressions.

c. Evaluation Metric:

I will evaluate the model with a confusion matrix and since we have 7 categories, I will be doing a confusion matrix for each of the 7 categories.
I am taking this approach of "only measuring the accuracy of each submodel alone" because if I measure the overall accuracy of my model (equal to the product of the accuracy of each submodel) there will be the issue of "error propagation".

d. Final conceptualization:

The final product would be demo-ed as follows:
A person opens their webcam, and the algorithm first detects the person's face (there is a square that appears around that person's head). Then we can stimulate a specific emotion by either talking to the person or sending them a message or image. Once they see/hear something that triggers an emotion, that emotion will be displayed on their face and the model will annotate the square with that emotion. The person testing this could also just make a facial expression of her/his choice and the model would detect it.

3. Application:

I will integrate the model in alanding-page webapp, and will attend the *Intro to Flask Workshop*, because I am not familiar with web applications. This will allow the model to connect to the webcam of any laptop user, and if the page is opened on phone or tablet, it could do the same as well.