BME 548: Project Proposal
Ashish Vinodkumar
Pranav Manjunath

I See You: Facial Detection and Emotion Classifier

General Aim:

To build an application where the user can input an image of a person, and the Machine Learning model will be able to perform face detection and face emotion classification. Specifically, we will be analyzing how downsampling, image masking techniques, color jitter techniques, impact model performance with accurately identifying and classifying facial emotions. Furthermore, our thesis will be looking to answer the following questions:

1. Do models trained on images that are masked in the top-half or bottom-half or side-half, result in an improved classification accuracy, compared to a model trained on the full image?
2. Does a model trained on only "some" of the pixels result in a significant boost in performance? Can we achieve near-real time face detection and emotion classification while maintaining similar accuracy compared to the baseline model?
3. Does downsampling of the image result in better performance? Does downsampling of the image maintain accuracy metrics compared to the baseline model?
4. Do color jitter augmentations (brightness and contrast) improve face detection?

We plan to break down our project into the below 2 segments:

1. Face Detection:
    a. We aim to perform facial detection to identify only those pixels corresponding to the face, place a box around the face, and remove all "background noise" from the image. To facilitate this process we will use the Faster-RCNN model as our baseline model architecture. We will further enhance this model architecture for our domain and business use-case.
2. Face Emotion Classification:
    a. Having removed all noise and narrowing down to the face, we aim to employ image augmentation techniques as described below, to perform image classification into 7 categories (Happy, Sad, Angry, Neutral, Surprise, Fear, Disgust). Specifically, to facilitate image classification, we will use the VGG-16 model architecture as our baseline model architecture, which natively contains a non-linear activation for classification. Furthermore, to create an alternative model, we will extract the features from the VGG-16 and input it into an XG-Boost model. We will use these 2 models for comparison of performance metrics such as accuracy, precision, and recall.

We will be looking at two different types of data preprocessing/augmentation techniques for Face Detection and Face Emotion Classifier.

1. Image Augmentation Techniques: We plan to transform the image using color jitter transformation technique. Within color jitter, we will be exploring various brightness, contrast, and aspect-color ratio transformations, to assess model performance for facial emotion detection.
2. Masking Algorithms: After the face detection, we will implement various masks to the cropped image (output of the bounding box - facial detection) and send the masked image to a CNN model for facial emotion classification to understand which masking algorithm has the highest model performance. Specifically, we will analyze the tradeoffs between data augmentation techniques such as downsampling, top-half, button-half, side-half, and diagonal masking, and its impact on image classification.

Discussion

1. **Data Source**
    a. Face Detection
        - Kaggle Labeled Faces in the Wild (LFW) Dataset containing 13233 images
        - Link: https://www.kaggle.com/jessicali9530/lfw-dataset
        - We will create the annotations (bounding boxes) using the Cascade Classifier from the OpenCV module in Python.
    b. Emotion Classification
        - Kaggle Facial Expression Recognition Dataset consisting of 28821 images
            - Link:https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset
            - 7 Emotions: Happy (7164 images), Sad (4938 images), Anger (3993 images), Disgust (436 images), Neutral (4982 images), Fear (4103 images), Surprise (3205 images)
        - Edureka's Facial Emotion Classifier Dataset consisting of 350 images
            - Link:https://www.dropbox.com/s/w3zlhing4dkgeyb/train.zip?dl=0&file_subpath=%2Ftrain
            - 7 Emotions: Happy, Sad, Anger, Disgust, Neutral, Fear, Surprise
            - 50 images for each emotion

c. Expected Simulation: Please find below our expected simulation workflow:

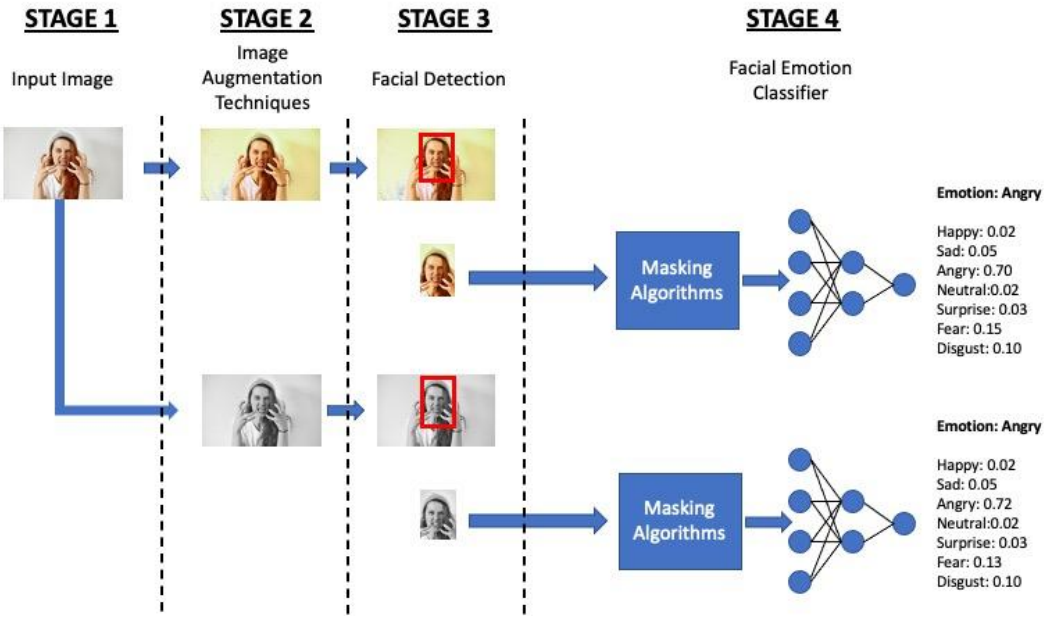

Figure 1. Expected Simulation Workflow

d. Expected CNN
   i. Face Detection
      - We plan to train a Faster R-CNN facial detection model that can output a bounding box around detected faces in the image.
   ii. Face Emotion Classifier
      - We plan to train a VGG-16 model in two ways. Firstly, using the VGG-16 model for image classification and secondly, for feature extraction and the features then are inputted into an XG-Boost model for classification.

e. Quantitative Analysis of Physical Layer/Physical Component
   The Physical Component to our project will consist of image augmentation and masking techniques that can be used to maximize facial detection and emotion classification performance, as described in the above sections. While the model will be trained on a variety of human stock photos, we plan on testing our model on laptop webcam images - as this project can be implemented for real-time webcam based face detection and face emotion classification.