MOHAMED BIN ZAYED
UNIVERSITY OF
ARTIFICIAL INTELLIGENCE

Assignment 4
CV701 - Human and Computer Vision

Introduction:

In this assignment you must implement the following **two tasks** and submit the code (if required) along with a brief report summarizing your findings. This is a **group assessment**, where you are expected work in groups of no more than 4. You **must** implement all of the functions mentioned below from scratch without using any available libraries that directly carry out the full task/sub-task.

Submission. For this assignment, you must submit:

- The group report in PDF format, to be submitted to the link labelled with "Group Report Submission". Only one member of the group should submit this. You must add your code as editable text to this document. This report **must not exceed 6 pages,** excluding referencing and appendices.

- The group code and other materials (i.e. images) in a compressed (zip/tar) file to be submitted to the link labelled "Group Code Submission". Similar to the above, only one member of the group should submit this.

- Individual report in PDF format (no more than half a page), highlighting your individual contribution to the group.

- A group demo video showcasing the deployed model. The video should be **no longer than 3 minutes** and should highlight the performance of the model and its real-time performance.

Task 1: Model Development (6.5/12.5 Marks)

You are provided with a dataset of images of faces, each image with a ground truth set of facial keypoints. Starting from the dataset class defined in "dataset.py", carry out the following:

1. Create and train a deep learning model to detect the facial keypoints using the given dataset.

2. Test the model, report, and analyze the achieved performance using appropriate metrics.

3. Use the detected keypoints to determine whether the person in a given image is showing a positive or negative emotion. You can also consider the case where a person is showing a neutral expression.

Task 2: Model Optimization and Deployment (6/12.5 Marks)

Using the model developed in Task 1, carry out the following tasks:

1. Deploy the model on a device of your choice (Laptop, mobile device, etc..). The device must have a camera of which you can access the video stream and carry out inference.

2. Display the model's output keypoints on the video feed.

3. Optimize the model computational complexity to operate on the device in real-time.

4. Devise an approach to assess the performance of the model after deployment. Report the results you achieve and compare them to the test performance achieved in Task 1.

Constraints

All the above-mentioned tasks should be carried out with the following constraints:

1. You are only allowed to use the dataset provided to train your model.

2. You can start from ImageNet pre-trained models if you wish.

3. You can use a rule-based approach for Task 1.3.

4. You must use a device with a camera for Task 2. However, in case you do not have access to a device with a camera, you should simulate a camera feed using a video stream.

5. You do not have to support inputs containing more than one face.