MSA 8150 Final Project:
Behind the Mask

Quick Overview

Each team can have a maximum number of 3 members, and would need to pick 2 projects among the 4 possible projects. Notice that if you decide to team up with fewer people than 3, you are still responsible for the same number of projects.

This project deals with image data. Many of you have heard about the deep fake problem. You would be developing a model which can map pictures of people with masks to similar pictures with the mask removed. Notice that the training data does not include pairs of images of the same person, and you would need to use generative models to learn the mapping.

Problem Setup

The pandemic has been with us for quite a while and we all find our ways to deal with it. Everyday, we see people wearing masks, and wonder how they look without the mask. Let's see if machine learning can help us answer that question.

The goal of this project is to train a machine learning model, which receives the picture of a person wearing a mask, and and outputs a predicted image of the person without the mask. Notice that we do not deal with a regression problem, because the training data does not contain pairwise images of subjects taken with and without masks. Instead the data consists of a large number of pictures from people with masks, and a large number of pictures from people without masks, where the subjects in the two sets are most likely different. Notice that you are not allowed to use any external data, and your model is only allowed to use the data provided. For this problem, you would need to familiarize yourself

with more advanced machine learning tools such as generative models. At the end of the day, you would still be training some deep learning models, but the training procedure is different than how a standard regression or classification model is trained.

One of the seminal papers related to this research is available at [this link]. However, more advanced techniques have been published since then, which would produce more realistic results. As a research problem, you are encouraged to do a survey of the available technique and set up your own techniques to address this problem. Simply downloading someone's code from Github and running it on the data is not an acceptable project completion, and you would need to explore different directions, novelties and technologies, along with comparisons to ultimately produce the most realistic results for the competition. You would be asked to do a live demo of your code on the competition day.

Please make sure to communicate with the instructor and Piazza about any potential questions related to the data.