Caterev Robert
Coty Ma

Deliverable 1

Choice of Dataset

Our final project idea is to train a machine learning model to recognize human heads in images first, then to try and perform the same recognition in videos with the help of some libraries. Then, upon recognizing human heads, we will attempt to blur each head. We will attempt to perform this "live," or iteratively, and if it is too difficult of a task, we will switch to post processing blurring. This could be applied to public video feeds to preserve the privacy of individuals.

In order to accomplish this, we have gathered a few datasets. We are not yet sure of which of the following datasets we will choose, however, they have massive libraries of images containing humans that are segmented more or less precisely into a certain number of body parts. We will attempt to either choose the best dataset possible, or combine the data from multiple datasets if they have the same format of segmentation. We will only keep the segmentation information relating to the human head, as we are not attempting to create a general body segmentation model. **Table 1** contains the datasets that we will consider.

"**CCIHP (Characterized Crowd Instance-level Human Parsing)** dataset provides pixelwise image annotations for: human segmentation, semantic attribute segmentation and semantic attribute characterization."
The **PASCAL Part Dataset** is a modification of the PASCAL Visual Object Classes Dataset that "goes beyond the original PASCAL object detection task by providing segmentation masks for each body part of the object."
The **Multi Human Parsing Dataset** introduces "the problem of multi-human parsing in the wild," providing us with a dataset of human body segmentation images that contain multiple humans.
The **Freiburg Sitting People Dataset** provides once again a body segmentation dataset.
The **Human Parts Dataset**
In case we need it, the **Body Parts Dataset** provides a dataset of labelled specific smaller body parts such as noses, necks, ears, eyes, etc.

Table 1. *Datasets under consideration and their respective descriptions*

Methodology

a) There are many datasets online from projects related to face or body recognition, so there certainly isn't a lack of usable data. The most useful piece of data would be any kind of clear image with colours of the human head, notably ones with the parts of the face well defined. We would also have to get rid of pictures in black and white, or with any other colour altering filters. Otherwise, we most probably won't need to classify each image in a subcategory since we would like our model to recognize any face. We will have to carefully merge datasets and discard data about other body parts in the dataset as to not build a general body segmentation model, as that is not the goal of our project. If merging of datasets is possible, careful consideration will be given to the chosen method of "storing the segmentation," as to choose the easiest datasets that we can work with.

b) We would like our model to recognize the image of a face, or, to classify the correct subgroup of pixels as a visage. To do so, we would like to use supervised learning along with a neural network. Firstly, a large portion of our datasets contain labelled images of the head, describing the position of the different organs. Thus, the information is already properly formatted. Secondly, we believe that since we are dealing with faces with no set appearance, the neural network is the most intuitive model to use in order to recognize them. There definitely isn't a lack of pre-processed data online. However, we are worried about the large processing time that may ensue, as well as the difficulty of fixing appropriate hyperparameters. Further down the line, if we find a more appropriate model for our project, such as a logistic regression, then we would use that one.

c) Our evaluation metric will be the i) Confusion Matrix and Accuracy/Precision-Recall Logistic Loss because our problem is closely related to a classification one. We would like our model to classify a series of pixels as a face with an accuracy above 80%.

Application

The user will be able to input either an image or a video. As discussed previously, the videos will be processed either live, or in post-processing. In the case that live-processing capabilities are achieved, the input can be any camera-like object. If not, the input would have to be video files and image files only. If the video feed comes through a webcam, we would adapt our webapp as to process the incoming feed, possibly using the *OpenCV* library for python to facilitate manipulation.

If the final project is capable of blurring faces on live, then there will be three possible outputs. The best one would be a live blurring of the faces as the user is filming themselves on their computer or phone. The other two possible outputs will be image and video files where the faces are blurred as well. If the model is not able to blur faces fast enough, then we will only output image and video files in post-production. In addition, the output will either be a jpg, png, or any other convenient image file, and the video will probably be in mp4, and the user will be able to choose to download it. As for the live face blurring, it will simply be shown on screen. More details will follow as we work on the UI.