Privacy via Select Obfuscation of Large Scale Visual Datasets

SUSHMITA KHAN, Clemson University

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Sushmita Khan. 2018. Privacy via Select Obfuscation of Large Scale Visual Datasets. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Large-scale visualization datasets (LSVD) like ImageNet [**?**] contain millions of images that were scraped from the internet, made publicly available, and published to further research in the fields of Computer Vision and Machine Learning. However, the pressing issue with such a dataset is that the images were obtained without the image owner's consent, thereby infringing upon their privacy. It is even more alarming because datasets like ImageNet have explicit, misogynistic, and even pornographic images in addition to images of children which were all collected without consent, and are very easily identifiable with applications like the reverse image search engine (RISE) [2] leaving these people vulnerable to unforeseen harm.

Considering the adverse effects one can suffer as a result of their image existing in LSVDs, it is imperative that we take action to protect the identity of these people. In an effort to get to know the dataset, I started with auditing a small subset of the training data (approximately 1GB) of the Imagenet dataset. My explorations focused on detecting people in the images (as compare to objects), number of people in each image, and age. Next, I worked on obfuscating faces in the validation dataset (approximately 6GB) in an effort to help up keep the privacy of the people who are in the publicly available and widely used ImageNet dataset.

The remainder of this paper talks about my two experiments, some interesting results that I found, and the methods I used.

2 EXPERIMENT 1: AUDITING THE TRAINING DATASET

The ImageNet data is the largest image dataset available and is approximately 1TB. Given the computational and time constraint, I decided to audit a very small subsection of the dataset, which was released in 2012. Although my original goal was to identify misogynistic, explicit, and pornographic content, the images in the dataset I downloaded are of dogs of different breed, people holding dogs, playing with dogs and such. While it was an option to download another sunset, I decided downloading and running analysis on a 155GB dataset wasn't a realistic option for me and thus I

pivoted. My auditing [?] included identifying human faces from objects and animals, finding the number of people in the images, their age, their gender, and the dogs breed's.

I ran my experiments on Clemson University's Palmetto server using their Jupyter Hub platform. Using the open sourced deep face analysis toolbox, *InsightFace*, I ran an initial model on the following image to find the age, gender of each of the individuals there:

The model detected 3 females who are of ages 24 and 25, and 3 males of ages 25, 33, and 35. After testing the model on this image, ran the experiment on a small portion of the subset. This experiment was looking for human faces in the images and the number of human faces. Images with label 0 indicate that there is no identifiable face in the image, label 1 implies 1 identifiable face, and label 2 implies 2 identifiable faces. The following illustrates the results of a few images:

As demonstrated in these images, most the classification of faces are done correctly. However, I want to focus particularly on the top left image in figure 1 and figure 3. It is interesting how the model classifies the top left image in figure 1 as '0' implying there is no person in the image. Furthermore, in figure 3, although clearly is a child holding a dog, the model classifies it as no person is present. This was an interesting find and I would like to explore this further in the future.

Next, I worked with classifying age and gender. While the model was able to classify correctly, there were also instances of misclassification of age. This might be because of the image quality, or because of incorrect labeling in the

Fig. 1. Identifying faces - 1

Fig. 2. Identifying faces - 2

labeled data. For example the person in figure 4 is labelled as a 27 year old woman. However, I would question the authenticity of the age label considering the image. I also detected and labeled faces, as they appeared in the images, as shown in figure 5. Finally, I find that this subsection of the imageNet dataset has an overwhelmingly high number of females, than males.

3 EXPERIMENT 2: FACE OBFUSCATION

I used Gaussian blurring [**?**] method to obfuscate the faces of the people in the validation dataset. This method of obfuscation removed sharp boundaries between the blurred face and the unblurred area in the image. In this method, first the bounding boxes are drawn around the faces and then enlarges, followed by truncating the out-of-rage coordinated. Then the unions of the enlarged boxes is presented. Each of these bounding boxes has a value of 1 inside it and value

Fig. 3. Identifying faces - 3

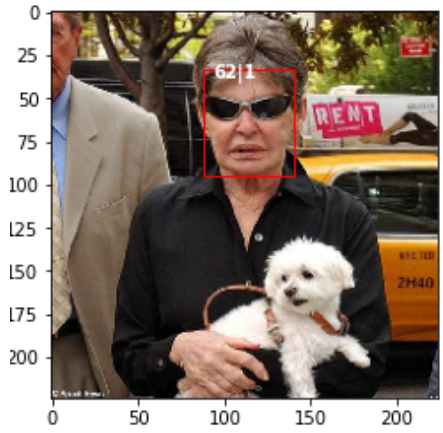

Fig. 4. Identifying faces - 3

of 0 outside it. Then the gaussian blurring technique is applied to the union of the boxed and the RGB values of the image. Although this method worked sufficiently to blur faces and can sucessfully determine between human faces and others, I am skeptical of how well the blurring will work in containing privacy of the users, because of recent reverse engineering methods. The following figures illustrates the obfuscation:

4 CONCLUSION AND DISCUSSION

The ImageNet dataset has considerable contribution to advancing research in Computer Vision and Machine Learning. However, it has infringed upon privacy of many a individuals, and also seems to have many biases namely age and gender. These biases adversely affects models, and classification and can have severe real life implications. As for obfuscating, although the Gaussian Blurring method is effective, in my opinion, it does not necessary blur images well

enough to make them unrecognizable. Furthermore in the age of deepFakes, and bot, having so many easiy identifiable images can pose threat to these people.