Investigating the Impact of Inclusion in Face Recognition Training Data

Chris Dulhanty University of Waterloo, Waterloo, Ontario, Canada
Alexander Wong University of Waterloo, Waterloo, Ontario, Canada
Email: {chris.dulhanty, a28wong}@uwaterloo.ca

Abstract

Modern face recognition systems leverage datasets containing images of hundreds of thousands of individuals' faces. Recently, there has been significant public scrutiny into the privacy implications of large-scale training datasets such as MS-Celeb-1M, as many people are uncomfortable with their face being used to train dual-use technologies that can enable mass surveillance. However, the impact of an individual's inclusion in training data on a derived system's ability to recognize them has not previously been studied. In this work, we audit ArcFace, a state-of-the-art, open-source face recognition system, in a large-scale face identification experiment. We find Rank-1 identification accuracy of 79.71% for individuals present in training data and 75.73% for those not present. These results demonstrate that modern face recognition systems work better for individuals they are trained on, which has serious privacy implications as all large-scale, open-source training datasets do not gather informed consent from individuals during their collection.

1 Introduction

Face Recognition (FR) systems using deep convolutional neural networks (DCNNs) depend on the collection of large image datasets containing thousands of sets of *specific* individuals' faces for training. Using this data, DCNNs learn a set of parameters that can map an *arbitrary* individual's face to a feature representation that has small intra-class and large inter-class variability. Computer vision researchers have benefited from the enabling power of the Internet to collect large-scale image datasets, leading to great advances in performance in the past five years. Consequently, FR systems are now being integrated into consumer and industrial electronic devices. However, along with improved performance has come increased public discourse on the ethics of face recognition systems and their development. In this study, we investigate privacy in the context of FR training data by assessing the impact of inclusion in the training data of a FR system on its ability to identify an individual.

2 Methodology

We frame this study as a closed-set face identification task. A *gallery* of known identities is constructed from images of individuals in advance of testing. Then, a new image of one of the gallery identities is presented to the system as the *probe*. The system attempts to match the probe with its identity in the gallery. All images in the gallery are ranked by distance in feature space to the probe, and the position of the correct identity in the ranked list is reported.

Face Recognition Model: We use the ArcFace model [1] in this work, trained on a cleaned version of Microsoft's MS-Celeb-1M dataset (MS1M) [2], containing 5.2M images of 93,431 identities. The model achieves 99.83% verification accuracy on Labeled Faces in the Wild [3] and 81.91% Rank-1 identification accuracy on MegaFace Challenge 1 [4], considered state-of-the-art results. We select this model for study as is the top academic, open-source entrant on the NIST Face Recognition Vendor Test[1], a government benchmark used by many commercial entities to validate the performance of their FR systems.

Probe Data: We construct two probe datasets from the VGGFace2 dataset [5] by matching identities by name with MS1M. We randomly select 1,000 identities present in both datasets as the *training probe set* and 1,000 present only in VGGFace2 as the *novel probe set*. For each identity, we randomly select 50 images and perform face detection and alignment with the Multi-Task Cascaded Convolutional Network (MTCNN) [6] to generate 112 x 112 pixel face crops. We then generate 512D feature representations for all images by running them through ArcFace.

Gallery Data: We leverage the MegaFace Challenge 1 "Distractor" dataset [4] of 1.0M images of 690,572 identities to form the basis of the gallery. We again apply MTCNN to generate normalized face crops for each image and generate feature representations with ArcFace.

Evaluation Protocol: The experiments conducted in this work follow the protocol of MegaFace Challenge 1, with our probe sets in place of the FaceScrub test set [7]. We employ the Linux development kit offered by MegaFace to perform experiments. We evaluate each probe set following Algorithm 1.

Algorithm 1: Closed-set face identification evaluation

```
Result: Rank-1, 10 and 100 face identification accuracies for a
        probe set.
$r_1, r_{10}, r_{100} = 0$;
gallery contains 1M distractor images;
for identity_i in identities_{1 to 1000} do
    for image_j in images_{1 to 50} do
        add image_j to the gallery;
        for image_k in images_{1 to 50} do
            if image_j == image_k then
                continue;
            else
                rank all images in gallery by L2 distance to
                  image_k;
                if image_j in first position in ranked list then
                    $r_1 = r_1 + 1$
                if image_j in first 10 positions in ranked list then
                    $r_{10} = r_{10} + 1$
                if image_j in first 100 positions in ranked list then
                    $r_{100} = r_{100} + 1$
        remove image_j from gallery;
Rank-1_Acc. $= r_1/(1000 \times 50 \times 49)$;
Rank-10_Acc. $= r_{10}/(1000 \times 50 \times 49)$;
Rank-100_Acc. $= r_{100}/(1000 \times 50 \times 49)$;
```

3 Results and Discussion

We present results in Table 1. We find a modest increase in face identification accuracy for identities that are present in the training data, compared to identities the model were not trained on. This relationship holds for Rank-1, 10 and 100 identification accuracies.

Table 1: Face identification accuracy of ArcFace model on different probe image sets with 1M distractor images.

Probe Set	Rank-1	Rank-10	Rank-100
Training	79.71%	90.82%	92.72%
Novel	75.73%	86.58%	89.22%

These results are concerning from a privacy perspective. There does not exist a major open-source FR dataset that gathers informed consent from the individuals it contains. As FR systems become more powerful and ubiquitous and regulatory measures are not enacted, the potential for misuse by governments or industry increases. While MS1M contains only "celebrity" identities, this classification of an individual should not negate informed consent requirements of data collection. By demonstrating the enhanced performance of FR systems in tracking certain individuals, without their knowledge, this work aims to inform best practices in the collection of large-scale FR datasets.

References

[1] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

[2] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*, pp. 87–102, Springer, 2016.

[3] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[4] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4873–4882, 2016.

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 67–74, IEEE, 2018.

[6] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[7] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 343–347, IEEE, 2014.

[1] https://www.nist.gov/programs-projects/frvt-11-verification