Compressed Models Decompress Race Biases on Face Recognition

Pedro C. Neto[1,2]
pedro.d.carneiro@inesctec.pt
Eduarda Caldeira[1,2]
up201906930@edu.fe.up.pt
Jaime S. Cardoso[1,2]
jaime.cardoso@inesctec.pt
Ana F. Sequeira[1,2]
ana.f.sequeira@inesctec.pt

[1] FEUP
Porto, Portugal
[2] INESC TEC
Porto, Portugal

Abstract

With the ever-growing complexity of deep learning models for face recognition, it becomes hard to deploy these systems in real life. Researchers have two options: 1) use smaller models; 2) compress their current models. The usage of smaller models might lead to concerning biases. However, compressing might be also responsible for an increase in the bias of the final model. We investigate the racial bias of a State-of-the-Art quantization approach when used with synthetic and real data. This analysis provides a few more details on the benefits of performing quantization with synthetic data, for instance, the reduction of biases on test scenarios. We tested four distinct architectures and evaluated them on a test dataset which was collected to infer and compare the performance of face recognition models on different ethnicity.

1 Introduction

Face recognition methods have made significant progress over the previous years. The urge to keep the current rate of improvement on these deep learning-based approaches led to an era of complex and obscure models. As such, despite their extraordinary performance, there are two pressing concerns. First, there are hardware limitations that affect the complexity of the models that can be deployed and used in real scenarios. These limitations affect both storage, memory and processing time. The second concern is that the behaviour of a deep neural network is not easily understood [7].

Addressing these two concerns is of utter importance. One must be careful to avoid a potential trade-off between their mitigation. For instance, considering the possibility of an existing bias on the models, further reducing the model size can lead to an increased bias. Moreover, unless that the model is reduced to an interpretable version of itself, these growing biases remain hidden within the black-box model.

Current work is investigating different model compression approaches. In this work, we aim to study the impact of quantization on the mitigation or amplification of existing biases. We framed our problem within the context of racial biases in face recognition systems. Our work, starting from Boutros *et al.* [1] quantization approach, further includes the usage of real and synthetic data. This study also aims to understand the current trade-offs between small models and hidden biases.

Within the context of this work, we aim to answer three research questions: 1) *Are smaller models more biased?* 2) *Are quantized models more biased?* 3) *What is the impact of using synthetic data to quantize these models*?. Furthermore, the usage of synthetic data is motivated by the possibility of further growing our current datasets without compromising ethical concerns or privacy. Given these research questions, we present the following contributions:

- A study on racial bias on four differently sized models trained on MS1MV2 [3];
- Using QuantFace [1] with real or synthetic data we study the racial bias of the quantized version of all the models;
- We discovered that quantization with synthetic data mitigates the racial bias of the final model.

The following sections are divided into two major sections and a conclusion. Section 2 describes the methodology and datasets in detail. Finally, the results are shown and discussed in Section 3.

2 Methods

The methodology was designed to understand if there is a bias problem on quantized models, and for this we have used the publicly available QuantFace models.

QuantFace has four different architectures available: MobileFaceNet [2], ResNet-18 [6], ResNet-50 and ResNet-100. Each of these architectures is available in five distinct shapes: the original full-precision model, the 8-bit model quantized with real data, the 8-bit model quantized with synthetic data, the 6-bit model quantized with real data and finally the 6-bit model quantized with synthetic data. This part is essential to understand if the behaviour of the quantized model changes with the selected precision and the network architecture. Hence, the dataset for this is fixed as the MS1MV2, so we can ignore the data as a factor of variability.

2.1 Datasets

2.1.1 MS1MV2

MS1MV2 is widely used in the literature to train and compare several deep face recognition models [3]. It is a refined version of the original MS-Celeb-1M dataset [5], which further improved the training of these systems. The dataset contains 85k different identities and almost six million images and it is not balanced with respect to the race.

2.1.2 Synthetic data

This dataset, introduced in [1] contains approximately 500k unlabelled synthetic images. These images have been generated by a generative adversarial network [4]. The usage of synthetic data is often seen to result in sub-optimal performances which might be caused by a domain gap between real and synthetic data [8]. In this work, the goal is not to use the synthetic data to learn the representations from scratch, and we further argue that there might exist advantages of this domain gap.

2.1.3 RFW

Racial Faces in-the-wild (RFW) [9] was designed as a benchmarking dataset for fair face verification. It includes labels for ethnicity, which allows for a fair assessment of potential biases. And it contains 3000 individuals with 6000 image pairs for face verification.

2.2 Evaluation Metrics

The performance of the evaluated models was measures in terms of accuracy. For the fairness evaluation of these models we have utilised two metrics: the standard deviation between the different accuracies (STD), and the skewed error ratio (SER) seen in Equation 1.

$$SER = \frac{100 - min(acc)}{100 - max(acc)} \tag{1}$$

The STD aims to evaluate the variance between the different accuracy values. On the other hand, SER measures or much larger is the worst error when compared with the better error. This is important to understand the relative differences between the different accuracy values. As a relative evaluation metric, SER is highly sensitive when the accuracy is above 99%. This happens because as the errors get below 1% their relative difference also change accordingly. For instance, a SER computed for a maximum accuracy of 90% and a minimum accuracy of 80% is the

Table 1: Table comprising the results, evaluated on RFW, from the different models trained on MS1MV2 and their respective quantized versions for different bits and quantization strategies (real or synthetic data).

Model	Bits	Quant.	Caucasian	Indian	Asian	African	Avg.	STD	SER
MobileFaceNets	32	-	95.18%	92.00%	89.93%	90.22%	**91.83%**	2.41	2.09
	8	Real	95.32%	91.60%	89.27%	90.08%	91.57%	2.68	2.29
	8	Synth.	94.18%	91.83%	88.85%	89.72%	91.15%	**2.38**	1.92
	6	Real	90.05%	86.52%	82.88%	83.18%	85.66%	3.36	1.72
	6	Synth.	89.97%	86.95%	83.13%	84.40%	86.11%	3.02	**1.68**
ResNet-18	32	-	97.48%	95.38%	93.72%	94.27%	**95.21%**	1.66	2.49
	8	Real	97.42%	95.33%	93.55%	94.20%	95.13%	1.70	2.50
	8	Synth.	96.95%	95.07%	93.30%	93.87%	94.80%	**1.61**	**2.20**
	6	Real	96.93%	94.65%	92.52%	93.22%	94.33%	1.95	2.44
	6	Synth.	96.80%	94.78%	92.35%	93.28%	94.30%	1.94	2.39
ResNet-50	32	-	99.00%	98.15%	97.62%	98.32%	98.27%	**0.57**	**2.38**
	8	Real	99.07%	98.07%	97.65%	98.40%	**98.30%**	0.60	2.53
	8	Synth.	99.02%	97.72%	97.33%	97.88%	97.99%	0.73	2.72
	6	Real	98.32%	96.27%	94.55%	95.87%	96.25%	1.56	3.24
	6	Synth.	97.95%	96.63%	94.97%	96.20%	96.44%	1.23	2.45
ResNet-100	32	-	99.65%	98.88%	98.50%	99.00%	**99.01%**	**0.48**	4.29
	8	Real	99.57%	98.87%	98.15%	98.77%	98.84%	0.58	4.30
	8	Synth.	99.37%	98.72%	98.13%	98.78%	98.75%	0.51	2.97
	6	Real	95.27%	93.15%	90.32%	91.70%	92.61%	2.12	2.05
	6	Synth.	95.93%	93.40%	91.92%	92.60%	93.46%	1.75	**1.99**

same if these accuracy values were 99.9% and 99.8%. STD is highly sensitive to absolute differences, and grows large on sets with lower accuracy values.

3 Results

A careful analysis of the performance of the different sized models at full precision (Table 1) shows that smaller models tend to have higher biases and lower performance in terms of average accuracy. ResNet-100 is an exception and this difference might be connected to the fact that SER becomes highly sensitive when the errors are below 1%.

The quantized version of these models seems to retain the performance and bias advantaged when compared to simpler models. As theorised, the quantization has a negative impact on the bias, and in most cases on the performance too. The lower the number of bit, the higher the bias. However, the usage of synthetic data has shown, for all the different precisions, a capability to reduce the bias while retaining the performance.

4 Conclusion

In this document, we answered three research questions. 1) and 2) It was possible to infer that models quantized with real data and smaller models are indeed more biased; 3) it was also verifiable that using synthetic data for quantization positively impacts the fairness metrics. We have extended previous literature on the assessment of the information that is lost by quantized models and further introduced a novel topic regarding the usage of synthetic data for bias mitigation.

Despite the interesting results shown by our experiments, there are several gaps in the literature that should be tackled in future work. For instance, it is not known if this behaviour is the same for gender biases. The usage of the combined real data that has and has not been seen, and synthetic data should be also analysed to understand how can we, just by changing the training data, mitigate these biases while retaining the original performance.

5 Acknowledgements

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project HfPT, with reference 11 and by National Funds through the Portuguese funding agency, FCT-Foundation for Science and Technology Portugal, a PhD Grant Number "2021.06872.BD".

References

[1] Fadi Boutros, Naser Damer, and Arjan Kuijper. Quantface: Towards lightweight face recognition by synthetic data low-bit quantization. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 855–862. IEEE, 2022.

[2] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings*, volume 10996 of *Lecture Notes in Computer Science*, pages 428–438. Springer, 2018. doi: 10.1007/978-3-319-97909-0_46. URL https://doi.org/10.1007/978-3-319-97909-0_46.

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00482.

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014.

[5] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 87–102. Springer, 2016.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90. URL https://doi.org/10.1109/CVPR.2016.90.

[7] Pedro C Neto, Tiago Gonçalves, João Ribeiro Pinto, Wilson Silva, Ana F Sequeira, Arun Ross, and Jaime S Cardoso. Explainable biometrics in the age of deep learning. *arXiv preprint arXiv:2208.09500*, 2022.

[8] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, Ser-Nam Lim, and Rama Chellappa. Learning from synthetic data: Addressing domain shift for semantic segmentation. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 3752–3761. Computer Vision Foundation / IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00395.

[9] Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.