Racial Bias in Facial Recognition Technology

Jordi Sabatés De La Huerta[1], Ahana Deb[1], and Igor Kuzmin[1]

Universitat Pompeu Fabra
jordi.sabates01@estudiant.upf.edu
ahana.deb01@estudiant.upf.edu
igor.kuzmin01@estudiant.upf.edu

Abstract. With the recent advancement in technology, we can effortlessly collect a vast amount of visual data and access powerful ways of processing them. Such advancements in Facial Recognition Technology(FRT), have abetted its use in understanding, modeling and predicting human behaviour. Depending on the racial diversity of the dataset used to train these FRT algorithms, certain biases maybe be perpetuated by the algorithm and the researcher may not adequately account for these discrepancies. In this paper we aim to analyze recent publications on FRT benchmarks, and propose questions on the racial composition of datasets and accuracy reports on racial subgroups. Our analysis indicates that a significant portion of the papers does not consider any kind of bias, some racial groups are underrepresented in the datasets used, and there is a need for taking these factors into account while analyzing facial data, otherwise posing limitations in the performance of the FRTs.

Keywords: Facial Recognition Technology · Ethics in AI · Artificial Intelligence.

1 Introduction

Owing to the advent of artificial intelligence in every aspect of our lives, there is an increase in use of machine learning algorithms to predict and analyse human behaviour, from song recommendations on music apps, to FRT used in CCTV footages to identify a suspect, or non-intrusive detection of fever and contract tracing. [1], in a report proposed that substantial improvements in FRT accuracies have been achieved in the last 5 year(2013-2018) and subsequently, FRT has been widely used to improve security and surveillance, as well as healthcare, marketing and retail. [2] elaborates on how China and South Korea have utilised the effectivity of FRT along with other metrics to flatten the curve on COVID instances as well as COVID related mortalities. Despite the use of technology motivated by societal well-being, there might be serious consequences of misclassification by these algorithms, for example in case of decisions on identifying a potential criminal or a potential candidate for a job position, which is being increasingly automated through these algorithms [3], and through Israel based companies like Faception[4], which claims to be able to accurately predict intelligence and inclinations towards terrorism, solely through analysing facial data.

Many attempts have also been made to model and identify human emotions through only facial data [5].

There are a vast number of deep learning algorithms currently used in face classification and recognition, which have also been made easily available to users through most smart phones. The algorithms used here are often developed by training on pre-labelled datasets, and the problem is further magnified when we observe the racial and gender composition of the datasets used. [6] shows that algorithms trained on unbalanced and biased data, may perpetuate the history biases towards race or gender in its applications. Previous research like [7][8] have tried to explore the differences in algorithm accuracy across facial data from different races. In pre-deep learning era, [9] was one of the initial papers exploring racial bias in algorithms, and suggested that recognition accuracy was greater for the majority race composition of a dataset. [10] evaluated algorithms on multi-class demographic groups and concluded that VGG-Face, although outperformed other algorithms on classification, also exhibited a large difference in its evaluation metrics between images of Caucasian individuals and that of Black individuals.

Further many algorithms developed may not take race into consideration at all, for example, [11] explores convolutional networks to detect melanoma from image samples with high accuracy, but does not take into account the need for a balanced dataset having labels for racial characteristics, like skin colour, amount of hair, etc. This might lead to subpar performance of the algorithm for different races in a population which are not well represented in the dataset on which the algorithm is trained on. [12] ,through analysing data from 100 police departments North America, revealed that African American people are far more likely to be subjected to facial recognition searches than any other race or ethnicity. [13] shows that some FRT systems have high tendencies of misclassifying along both race and gender lines for the minority groups. [14] further characterizes the skin type distribution across IJB-A and Adience, two facial analysis benchmarks, and conclude that the data is majorly composed of light skin sample points, to a fascinating 79.6% and 86.2% for IJB-A and Adience, respectively.

Although there has been marked improvement in facial analysis algorithm, both due to the ease of gathering visual data and analysing them with high-parameter deep neural networks, the improvement in its performance has not been universal to every section of the population. In this paper, we analyse 30 papers on FRT on how they approach facial recognition datasets, and how well the authors take into account the racial composition of datasets used, and compensate for those discrepancies in their methodologies and findings. The objective here is to find what factors researchers in FRT need to take into account while analyising facial data of a particular racial composition, and what limitations might be posed otherwise. In our approach, we analyse the articles based on the following questions - if the articles mention racial bias, or any other type of bias, proportion of the different races mentioned, the algorithms implemented in each paper, and the sources of the datasets the algorithms were trained on.

2 Research methodology

To understand how much modern machine learning algorithms are racially biased and if researchers take care about data acquisition, we will analyze more than 30 scientific articles from Mendeley and all the publications analyzed in our study are referenced in Annex I. To be more precise in finding research that demonstrates modern approaches in facial bias determination and its analysis in the machine learning algorithms we consider finding publications that apply the state-of-the-art approach and are mainly published in the last five years. The following words were chosen as keywords to search for articles: “bias”, “racial bias”, “gender bias”, “age bias”, “face recognition”, “race”, and “ algorithm”.

The current research takes into account the fact that the European Commission formulated restrictions for FRT usage[15], which influences methodology and limitations. Data collection must be neutral to prevent any bias and be ethical. The high level of accuracy in machine learning algorithms should be maintained. It is important to observe relationships between other types of biases to understand any correlation between them and their influence on FRT. This is why it was decided to analyze articles mainly related to the FRT and as one of the metrics to subdivide the type of bias referred to in the article.

To proceed with the data from given articles, to provide researchers in the field of face recognition the ability to be aware of current trends in FRT biases according to the mentioned limitations were formulated categories and their parameters (Table 1).

The first category is a reference to racial bias in collected articles to understand how much FRT on analysis of the influence of racial bias. The proportions of different races and other types of biases, and algorithms mentioned in articles are chosen as other categories because it is important to understand the weaknesses of FRT and each technology used in publications. The source of data could demonstrate the probability of biases related to data sources and specifics in the data processing.

Table 1: :Categories used to identify different biases described in each research.

Categories	Values
Does the article mention racial bias?	Yes / No
Proportion of different races mentioned	African, Asian, Black, Caucasian, Indian, White, Other
Does the research use particular algorithm?	Yes / No
Proportion of algorithms that were used?	ResNet, CNN, FG, NEC, DQN, etc.
Proportion of other types of bias mentioned in the research	Age, Gender, Skin, Gender
Source of data	EU, USA, UK, ASIA, Global, Turkey

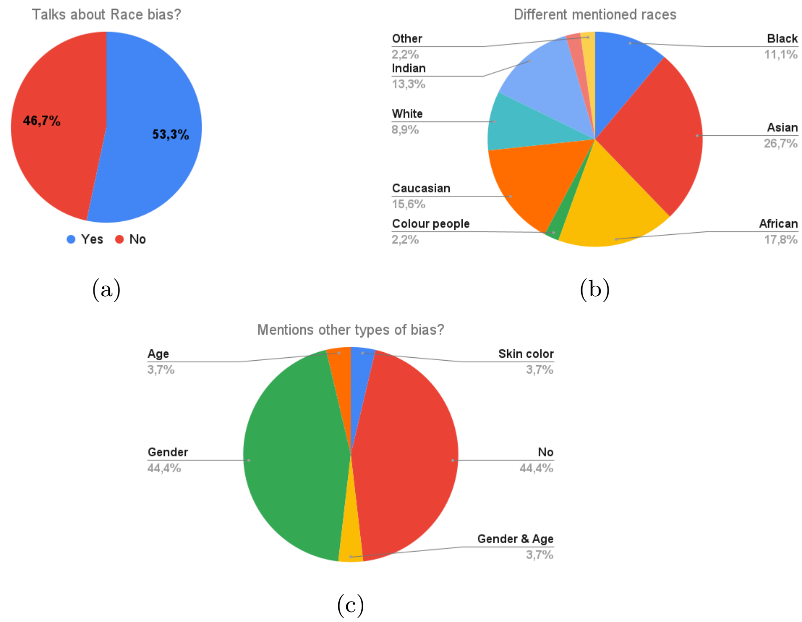

Fig. 1: Proportion of papers discussing a)race as a factor in FRT, b) different races mentioned, c) other biases.

3 Results

As shown in Figure 1a, we see that more than half of the papers (53.3%) discuss the issue of race bias. As it is a very current topic, most researchers take it into account when doing their research.

We were also interested in investigating which races these publications were about, whether they focused more on some human races in particular or on several races at the same time. From what we see in Figure 1b, 28.9% of the publications mention "African" or "Black" races and 26.7% mention Asian races. These two being the most repeated human races in the papers analysed. On the one hand, the black races are the most legally disadvantaged by FRT[16]. On the other hand, we have seen after our study, that in the Asian continent there is a high growth of interest in this technology, especially oriented to face recognition using facial masks, due to the urgency caused by the Covid-19 pandemic[17].

We also wanted to analyse the origin of the datasets used in the analysed papers. With this we can observe in which regions of the globe we find more data

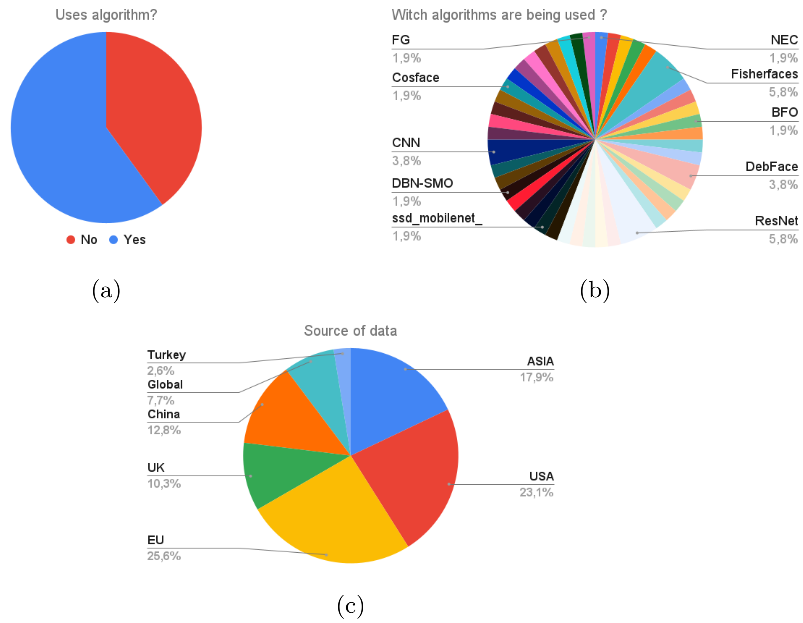

Fig. 2: Proportion of papers a)using algorithms, b) algorithms used, c) sources of dataset used.

regarding facial images. We see that most of the datasets come from Europe, Asia (China being the most relevant subregion) and USA in this order.

We also analysed what percentage of the publications used some algorithm in their research, or if on the contrary, they were limited to survey different people or to deepen the study of other papers. In Figure 2a we see that 3 out of 5 publications related to FRTs mention some algorithm related to this topic.

In Figure 2b we have plotted the frequency with which the algorithms appeared in the different papers. We have created the figure by obtaining the frequency of repetition of each algorithm as a function of the total used in the papers. We can see that there is an abundance of them and no one algorithm is the most used. The 3 most used ones are ResNet, FisherFaces and DebFace.

Finally, we wanted to see what other types of biases are taken into account in the publications we analysed. As we can see in Figure 1c, 44.4% of the papers analysed do not take into account any type of bias related to the subject analysed. On the other hand, the same proportion also takes into account gender bias. This indicates that the majority of publications that analyse race bias also do so with gender.

It should be noted that with a larger volume of papers analyzed we would have a better accuracy in drawing any conclusions, but due to the limited time we had to do the research, we were only able to analyze 30 papers.

4 Conclusion

In our paper, we attempt to explore to what extent researchers in the field of FRT take into account the racial composition of their training datasets, and how they report their findings as a function of the different races. We analysed 30 papers in the field of FRT, and identified the datasets and algorithms used in those publications, along with whether racial or any other bias was addressed in the paper, which races were discussed while defining the dataset, and the source of the visual data used to train their algorithms. We can conclude from our study that racial bias is still an ongoing issue in the field of visual data, especially FRT, and there is a need for inclusive and balanced datasets and accuracy reports on different racial subgroups for evaluation of a particular algorithm, and an active effort to make up for the discrepancies in the availability of data for particular subgroups as well as to mitigate unequal performances of algorithms on minority data.

The analysis in this paper is only limited to 30 publications, which may not be a sufficient representation of the current state of bias. We also haven't adequately analysed the intersectional aspect of bias in FRT across race, gender and sexuality, and how the factors interact with each other. Future work is required on larger sample set of publications, and further explore the intersectionality of bias exhibited by FRTs, based on race, gender, age, sexuality and other criterias.

References

1. P. Grother, M. Ngan, and K. Hanaoka, "Face recognition vendor test (frvt) part 2: Identification," 2019-09-13 00:09:00 2019.
2. S. Whitelaw, M. A. Mamas, E. Topol, and H. G. C. Van Spall, "Applications of digital technology in covid-19 pandemic planning and response," *The Lancet Digital Health*, vol. 2, no. 8, pp. e435–e440, 2020.
3. C. O'Neil, *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books, 2017.
4. Faception, "Facial personality analytics." `https://www.faception.com//`, Accessed November, 2022.
5. A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "DAGER: deep age, gender and emotion recognition using convolutional neural network," *CoRR*, vol. abs/1702.04280, 2017.
6. A. Caliskan, J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, vol. 356, pp. 183–186, 04 2017.
7. G. Givens, J. Beveridge, B. Draper, P. Grother, and P. Phillips, "How features of the human face affect recognition: a statistical comparison of three face recognition algorithms," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, pp. II–II, 2004.
8. B. F. Klare, M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge, and A. K. Jain, "Face recognition performance: Role of demographic information," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1789–1801, 2012.

9. A. O'Toole, K. Deffenbacher, H. ABDI, and J. Bartlett, "Simulating the 'other-race effect' as a problem in perceptual learning," *Connection Science*, vol. 3, pp. 163–178, 10 2007.
10. H. E. Khiyari and H. Wechsler, "Face verification subject to varying (age, ethnicity, and gender)demographics using deep learning," *Journal of biometrics & biostatistics*, vol. 7, pp. 1–5, 2016.
11. A. Esteva, B. Kuprel, R. A. Novoa, J. M. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
12. A. B. Clare Garvie and Jonathan, "Frankle. the perpetual line-up: Unregulated police face recognition in america," *Georgetown Law, Center on Privacy Technology,*, 2016.
13. B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1931–1939, 2015.
14. J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (S. A. Friedler and C. Wilson, eds.), vol. 81 of *Proceedings of Machine Learning Research*, pp. 77–91, PMLR, 23–24 Feb 2018.
15. T. Madiega and H. Mildebrath, "Regulating facial recognition in the eu." `https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698021/EPRS_IDA(2021)698021_EN.pdf`, Accessed September, 2021.
16. J. Vitriol, J. Appleby, and E. Borgida, "Racial bias increases false identification of black suspects in simultaneous lineups," *Social Psychological and Personality Science*, vol. 10, 07 2018.
17. J. Yu, X. Hao, Z. Cui, P. He, and T. Liu, "Boosting fairness for masked face recognition," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1531–1540, 2021.

Annex I: Analysed papers

1. Andrejevic, Mark Selwyn, Neil. (2019). Facial recognition technology in schools: critical questions and concerns. Learning, Media and Technology. 45. 1-14. 10.1080/17439884.2020.1686014.
2. Talahua, Jonathan Buele, Jorge Calvopiña, P. Varela Aldás, José. (2021). Facial Recognition System for People with and without Face Mask in Times of the COVID-19 Pandemic. Sustainability. 13. 6900. 10.3390/su13126900.
3. Kostka, Genia Steinacker, Léa Meckel, Miriam. (2021). Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. Public Understanding of Science. 30. 096366252110015. 10.1177/09636625211001555.
4. Xu, Feng Zhang, Yun Zhang, Tingting Wang, Jing. (2020). Facial recognition check-in services at hotels. Journal of Hospitality Marketing Management. 30. 1-21. 10.1080/19368623.2020.1813670.
5. Avital Meshi. 2022. Deconstructing Whiteness: Visualizing Racial Bias in a Face Recognition Algorithm. In 10th International Conference on Digital and Interactive Arts (ARTECH 2021). Association for Computing Machinery, New York, NY, USA, Article 91, 1–4. https://doi.org/10.1145/3483529.3483738

6. Zhang, Wen Kang, Min. (2019). Factors Affecting the Use of Facial-Recognition Payment: An example of Chinese Consumers (July 2019). IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2927705.
7. Fussey, Pete Davies, Bethan Innes, Martin. (2020). 'Assisted' Facial Recognition and the Reinvention of Suspicion and Discretion in Digital Policing. The British Journal of Criminology. 61. 10.1093/bjc/azaa068.
8. Trawiński, T., Aslanian, A., Cheung, O. S. (2021). The effect of implicit racial bias on recognition of other-race faces. Cognitive research: principles and implications, 6(1), 67. https://doi.org/10.1186/s41235-021-00337-7
9. Zhou, Y., Gao, T., Zhang, T., Li, W., Wu, T., Han, X., Han, S. (2020). Neural dynamics of racial categorization predicts racial bias in face recognition and altruism. Nature human behaviour, 4(1), 69–87. https://doi.org/10.1038/s41562-019-0743-y
10. Ullah, Naeem Javed, Ali Ghazanfar, Mustansar ali Alsufyani, Abdulmajeed Bourouis, Sami. (2022). A novel DeepMaskNet model for face mask detection and masked facial recognition. Journal of King Saud University - Computer and Information Sciences. 10.1016/j.jksuci.2021.12.017.
11. Xu, T., White, J., Kalkan, S., Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition.. ECCV Workshops (6), 12540 506-523. https://doi.org/10.17863/CAM.56933
12. Yong Peng, Suhang Wang, Xianzhong Long, Bao-Liang Lu, Discriminative graph regularized extreme learning machine and its application to face recognition, Neurocomputing, Volume 149, Part A, 2015, Pages 340-353, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2013.12.065.
13. Dirin, Amir Kauttonen, Janne. (2020). Comparisons of Facial Recognition Algorithms Through a Case Study Application. International Journal of Interactive Mobile Technologies (iJIM). 14. 121. 10.3991/ijim.v14i14.14997.
14. Rutuparna Panda, Manoj Kumar Naik, B.K. Panigrahi, Face recognition using bacterial foraging strategy, Swarm and Evolutionary Computation, Volume 1, Issue 3, 2011, Pages 138-146, ISSN 2210-6502, https://doi.org/10.1016/j.swevo.2011.06.001.
15. N. Srinivas, M. Hivner, K. Gay, H. Atwal, M. King and K. Ricanek, "Exploring Automatic Face Recognition on Match Performance and Gender Bias for Children," 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2019, pp. 107-115, doi: 10.1109/WACVW.2019.00023.
16. Sixta, T., Jacques Junior, J.C.S., Buch-Cardona, P., Vazquez, E., Escalera, S. (2020). FairFace Challenge at ECCV 2020: Analyzing Bias in Face Recognition. In: Bartoli, A., Fusiello, A. (eds) Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science(), vol 12540. Springer, Cham. https://doi.org/10.1007/978-3-030-65414-6_32
17. Ignacio Serna, Aythami Morales, Julian Fierrez, Nick Obradovich, Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning, Artificial Intelligence, Volume 305, 2022, 103682, ISSN 0004-3702, https://doi.org/10.1016/j.artint.2022.103682.
18. Song, Ziwei Nguyen, Kristie Cho, Catherine Gao, Jerry. (2022). Spartan Face Mask Detection and Facial Recognition System. Healthcare. 10. 87. 10.3390/healthcare10010087.
19. Xiaojun Lai, Pei-Luen Patrick Rau, Has facial recognition technology been misused? A public perception model of facial recognition scenarios, Computers in Human Behavior, Volume 124, 2021, 106894, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2021.106894.

20. Vedantham, R., Reddy, E.S. A robust feature extraction with optimized DBN-SMO for facial expression recognition. Multimed Tools Appl 79, 21487–21512 (2020). https://doi.org/10.1007/s11042-020-08901-x
21. Wen, Ge Chen, Huaguan Cai, Deng He, Xiaofei. (2018). Improving Face Recognition with Domain Adaptation. Neurocomputing. 287. 10.1016/j.neucom.2018.01.079.
22. Shi, Sheng Wei, Shanshan Shi, Zhongchao Du, Yangzhou Fan, Wei Fan, Jianping Conyers, Yolanda Xu, Feiyu. (2020). Algorithm Bias Detection and Mitigation in Lenovo Face Recognition Engine. 10.1007/978-3-030-60457-8_36.
23. M. Wang and W. Deng, "Mitigating Bias in Face Recognition Using Skewness-Aware Reinforcement Learning," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 9319-9328, doi: 10.1109/CVPR42600.2020.00934.
24. Georgopoulos, M., Oldfield, J., Nicolaou, M.A. et al. Mitigating Demographic Bias in Facial Datasets with Style-Based Multi-attribute Transfer. Int J Comput Vis 129, 2288–2307 (2021). https://doi.org/10.1007/s11263-021-01448-w
25. P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner and A. Kuijper, "Face Quality Estimation and Its Correlation to Demographic and Non-Demographic Bias in Face Recognition," 2020 IEEE International Joint Conference on Biometrics (IJCB), 2020, pp. 1-11, doi: 10.1109/IJCB48548.2020.9304865.
26. Gong, S., Liu, X., Jain, A.K. (2020). Jointly De-Biasing Face Recognition and Demographic Attribute Estimation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12374. Springer, Cham. https://doi.org/10.1007/978-3-030-58526-6_20
27. Georgopoulos, Markos Panagakis, Yannis Pantic, Maja. (2020). Investigating bias in deep face analysis: The KANFace dataset and empirical study. Image and Vision Computing. 102. 103954. 10.1016/j.imavis.2020.103954.
28. N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme and M. King, "Face Recognition Algorithm Bias: Performance Differences on Images of Children and Adults," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 2269-2277, doi: 10.1109/CVPRW.2019.00280.
29. Xu, T., White, J., Kalkan, S., Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. In: Bartoli, A., Fusiello, A. (eds) Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science(), vol 12540. Springer, Cham. https://doi.org/10.1007/978-3-030-65414-6_35
30. Yu, J., Hao, X., Xie, H., Yu, Y. (2020). Fair Face Recognition Using Data Balancing, Enhancement and Fusion. In: Bartoli, A., Fusiello, A. (eds) Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science(), vol 12540. Springer, Cham. https://doi.org/10.1007/978-3-030-65414-6_34