

New advances in the detection of bias in face recognition algorithms

January 20 2021



Credit: CC0 Public Domain

A team from the Computer Vision Center (CVC) and the University of Barcelona has published the results of a study that evaluates the accuracy and bias in gender and skin color of automatic face recognition

algorithms tested with real world data. Although the top solutions exceed the 99.9% of accuracy, researchers have detected some groups that show higher false positive or false negative rates.

Face recognition has been routinely used by both private and governmental organizations worldwide. Automatic [face recognition](#) can be used for legitimate and beneficial purposes (e.g. to improve security) but at the same time its power and ubiquity increases a potential negative impact that unfair methods can have on society (e.g. discrimination against ethnic minorities). A necessary, albeit not sufficient, condition for a legitimate deployment of face recognition algorithms is the equal [accuracy](#) for all demographic groups.

With this purpose in mind, researchers from the Human Pose Recovery and Behavior Analysis Group at the Computer Vision Center (CVC) – University of Barcelona (UB), led by Sergio Escalera, organized a challenge within the European Conference of Computer Vision (ECCV) 2020. The results, recently published in the journal *Computer Vision—ECCV 2020 Workshops*, evaluated the accuracy of the submitted algorithms by the participants on the face verification task in the presence of other confounding attributes.

"[The study] attracted 151 participants, who made more than 1,800 submissions in total, exceeding our expectations regarding the number of participants and submissions," explained Sergio Escalera, also member of the Institute of Mathematics of the UB.

The participants used a not balanced image dataset, which simulates a real world scenario where AI based models are supposed to be trained and evaluated on imbalanced data (considerably more white males than dark females). In total, they worked with 152,917 images from 6,139 identities.

The images were annotated for two protected attributes: gender and skin color; and five legitimate attributes: age group (0-34, 35-64, 65+), head pose (frontal, other), image source (still image, video frame), wearing glasses and a bounding box size.

The obtained results were very promising. Top winning solutions exceeded 99.9% of accuracy while achieving very low scores in the proposed bias metrics, "which can be considered a step toward the development of fairer face recognition methods," said Julio C. S. Jacques Jr., researcher at the CVC and at the Open University of Catalonia.

The analysis of top 10 teams showed higher false positive rates for females with dark skin tone and for samples where both individuals wear glasses. In contrast, there were higher false negative rates for males with light skin tone and for samples where both individuals were aged 35 and below. Also, it was found that in the dataset individuals younger than 35 wear glasses less often than older individuals, resulting in a combination of effects of these attributes. "This was not a surprise, since the adopted dataset was not balanced regarding the different demographic attributes. However, it shows that overall accuracy is not enough when the goal is to build fair face [recognition](#) methods, and that future works on the topic must take into account accuracy and bias mitigation together," concluded Julio C. S. Jacques Jr.

More information: Tomáš Sixta et al. FairFace Challenge at ECCV 2020: Analyzing Bias in Face Recognition, *Computer Vision – ECCV 2020 Workshops* (2021). [DOI: 10.1007/978-3-030-65414-6_32](https://doi.org/10.1007/978-3-030-65414-6_32)

Provided by University of Barcelona

Citation: New advances in the detection of bias in face recognition algorithms (2021, January 20)
retrieved 19 April 2024 from

<https://techxplore.com/news/2021-01-advances-bias-recognition-algorithms.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.