AIH Technology FaceAlgo racial biasness benchmark testing update

AIH Technology

Addressing racial biasness in facial recognition through algorithm engineering

Overview

Racial biasness is a critical problem facing the artificial intelligence (AI) industry. Especially in the context of facial recognition, a facial recognition algorithm that exhibits palpable bias across various ethinic groups can lead to severe consequences for the marginalized communities.

AIH Technology (AIH) is devoted to the mission of bringing inclusivity into AI and machine learning. This report details AIH's latest achievements in minimizing racial biasness in facial recognition.

In brief, AIH facial recognition algorithm (FaceAlgo) incorporated a number of breakthroughs in machine learning methods and deep learning models. These methods have enabled AIH's facial recognition algorithm to effectively minimize biasness and improve recognition accuracy across the spectrum of ethnic groups. The results of AIH's racial biasness minimization strategy is presented in this report in Page 4.

Current Technology Gap

Racial biasness and facial recognition algorithms

Over the past years, facial recognition has been the focal point of criticisms against AI and machine learning technologies on the subject of racial biasness. Most notably, Amazon's AWS Rekognition was reported having an unusually high false-positive rate in identifying African American faces: it incorrectly matched the photos of 28 U.S. congressmen with the faces of criminals, especially the error rate was up to 39% for non-Caucasian people.[1]

On December 2019, NIST released an analysis of 189 software algorithms from 99 developers. The NIST report shows a majority of commercially available facial recognition algorithms had unusually high false positive rates for Asian and African American faces relative to images of Caucasians. The differentials often ranged from as high as 10 to 100 times, depending on the individual algorithm.[2]

[1] Jacob Snow, Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots. ACLU. July, 2018. Accessed from: "https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28"

[2] National Institute of Standards and Technology, NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software, December 2019, accessed from: <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>

Lack of standardized benchmarking tests evaluating algorithm racial biasness

The issue with racial biasness in facial recognition is not just simply reflected in the poor performance of the algorithms in certain ethnic groups. What is more alarming is that major benchmarking tests, including NIST Face Recognition Vendor Test (FRVT), lack the proper evaluation mechanism for measuring racial biasness in public safety applications. Specifically, the NIST FRVT lacks the following proper controls to provide a clear evaluation of racial biasness:

1. **the database does not have comparable distribution of sample sizes** of various ethnic demographics, which makes the database inadequate in producing statistically significant comparative analysis of racial bias;
2. **the database does not use in-the-wild photos** (i.e. photos captured from video surveillance or internet). Only standards-compliant photos were used (e.g. passport photos, mugshots, traveller entry images taken from dedicated cameras at the US Customs). This database does not measure hidden ethnic differentials (i.e. racial biasness) that may occur in wild photographs that may arise due to camera limitations; and
3. **performance saturation**: the databases containing standards-compliant photos (i.e. passport photos) do not pose significant challenges for most commercial facial recognition algorithms to achieve a high accuracy. In other words, most participating algorithms would score fairly high on the accuracy scale. Despite having such high accuracy scores, those benchmark tests do not accurately reflect the racial biasness of algorithms, as reflected in the NIST Report.[3]

Other public benchmarking standards, like Labeled Faces in the Wild (LFW) created by the University of Massachusetts, provide wild photographs (photos captured from video surveillance or internet) for testing purposes.[4] However, the LFW benchmark does not offer assessment on racial biasness either.

[3] *Ibid.*

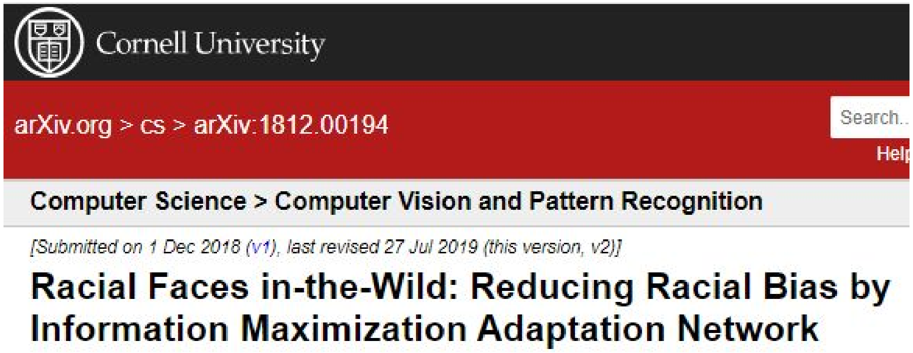
[4] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: A database for studying face recognition in unconstrained environments.

AIH FaceAlgo performance on racial biasness testing

AIH Technology is fully committed to addressing the issue of racial biasness in facial recognition. In this report, we provide benchmarking tests results from a public database that is specifically designed to study racial biasness, as well as a glimpse into the algorithm engineering measures that AIH engineers have taken to minimize racial biasness.

Benchmarking tests result from RFW Databases

Racial faces in the Wild (RFW) is a recently launched public benchmark database designed for studying racial biasness in facial recognition algorithms with unconstrained face images.[5] Unlike NIST's FRVT, which uses standards-compliant photos (i.e. passport photos), RFW uses facial images with a large range of the variation seen in everyday life. This includes variation in pose, lighting, expression, background, race, ethnicity, hairstyles, camera quality, color saturation, etc. Such approach in designing databases has effectively minimized performance saturation, which is regularly demonstrated in databases containing standards-compliant photos (i.e. passport photos), with most algorithms scoring above 95% in accuracy across the board.



arXiv.org > cs > arXiv:1812.00194

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 1 Dec 2018 (v1), last revised 27 Jul 2019 (this version, v2)]

Racial Faces in-the-Wild: Reducing Racial Bias by Information Maximization Adaptation Network

This particular public benchmark test, RFW, provides "real-world" test for studying the performance of facial recognition algorithms with a major focus on examining racial biasness. It contains four groups of testing subjects from the following categories: Caucasians, South Asians, East Asians, and Africans. The number of individuals and sample size of each database were kept at a uniform level to provide a fair and accurate comparison of racial biasness.[6]

[5] Wang, M., Deng, W., Hu, J., Tao, X., & Huang, Y. (2019). Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In Proceedings of the IEEE International Conference on Computer Vision (pp. 692-702).

[6] *Ibid*.

]

Figure 1. *Snapshot of face images used in the RFW database*. In order to avoid performance saturation, difficult-to-recognize face images were purposely selected to test the robustness of facial recognition algorithms, using challenging face image variations of the same individual subjects and similar appearance of different individual subjects.

For the purpose of truly evaluating the racial biasness of AIH Technology's facial recognition algorithm (FaceAlgo), we chose the publicly available RFW Database as benchmark.

AIH Technology's Score on the RFW Database

AIH's Facial Recognition as a Service (FRaaS) - FaceAlgo API deployed on Amazon AWS cloud was used to compute RFW's four databases, including Caucasian, African, South Asian, and East Asian. The results were compared with data obtained by Wang *et al*. from commercially available facial recognition APIs, including Amazon, Microsoft, and Face++.[7] The results are shown below:

	African	East Asian	Caucasian	South Asian
AIH	97.60%	96.07%	97.28%	97.25%
Amazon	86.27%	84.87%	90.45%	87.20%
Microsoft	75.83%	79.67%	87.60%	82.83%
Face++	87.50%	92.47%	93.90%	88.55%

Table 1. *AIH Technology's accuracy score obtained from the RFW database in comparison to the major commercial facial recognition algorithms offered by Amazon, Microsoft, and Face++*.

[7] *Ibid*.

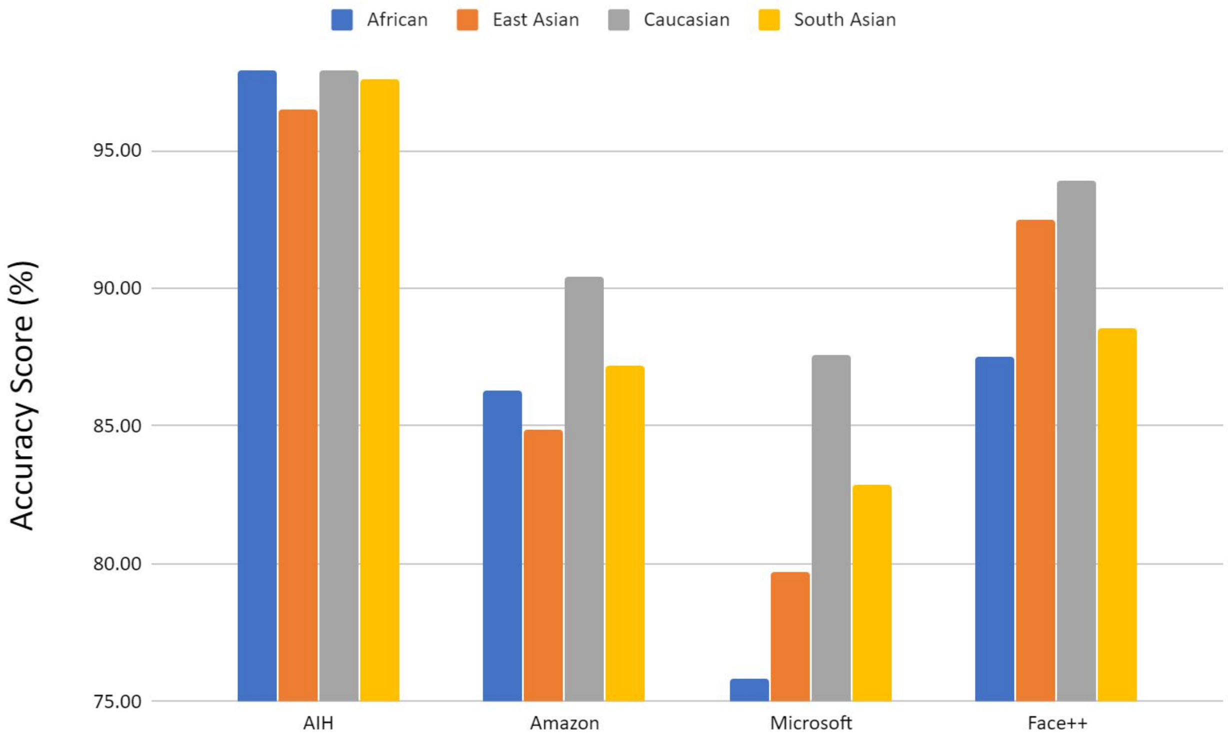


Figure 2. *Visualization of accuracy score comparison between AIH Technology's facial recognition and those offered by other commercial vendors, including Amazon, Microsoft, and Face++.*

As shown in both Table 1. and Figure 2., AIH Technology's FaceAlgo demonstrated the following characteristics

1. RFW results shown a significant reduction in performance saturation in comparison other public benchmarks, with most commercial algorithms scoring below 95% in accuracy;
2. FaceAlgo's overall accuracy performance across four ethnic groups scored consistently above Amazon, Microsoft, and Face++; and
3. FaceAlgo exhibited minimal accuracy performance differentials amongst the four ethnic groups tested, and its accuracy score is the highest in the African group.