RISE RESEARCH INNOVATION SCHOLARSHIP ENTREPRENEURSHIP

Abstract ID#: 3383
Category: Engineering and Technology
Undergraduate

Northeastern University

Measuring Fairness In Automatic Face Recognition

Alice Loukianova†, Dylan Dasgupta†, Rohan Krishnamurthi†, William Cutler†, Joseph P. Robinson

Opportunity

- Facial recognition (FR) is all around us.
- Acquiring transparent and fair assessment of systems that we so heavily rely on is crucial.

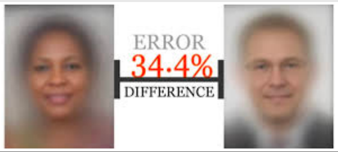

Fig 1. On average, performance gaps across gender are imbalanced.

- Recent studies in FR favors certain subgroups, i.e., demographics (**Fig. 1**)– a problem concerning researchers and mainstream alike.
- Furthermore, there exists no convention for quantitative or qualitative analysis and ratings for bias (aka “fair AI”)
- Lacks labeled data for bias studies in FR.
- Recent work at NEU on this problem [1].

The Solution

Fig 4. Dashboard developed as an easy-to-use FR bias assessment tool generates various plots, stats, and automatic report generation.

Data and Analysis

- Balanced Faces in the Wild (BFW) consists of many faces and even more pairs with balance across ethnicity, gender, and identity (**Table 1**).
- Verification is the task: pairs are *genuine* if a true match; else, *imposter*.
- A global threshold yields imbalance accuracies across subgroups - different levels of sensitivity in scores (**Fig 2**).
- Applying subgroup-specific thresholds minimizes the error rate **(Fig 3)**.
- We built a dash as an open-source tool for researchers and practitioners to measure FR systems for bias (**Fig 4**).

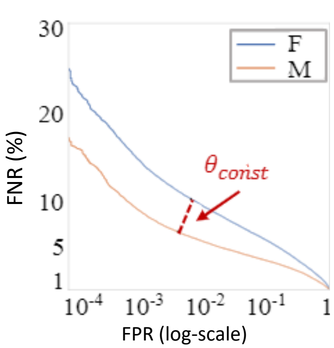

Fig 2. Gender DET Curve. Same FNR obtained with a constant θ [1].

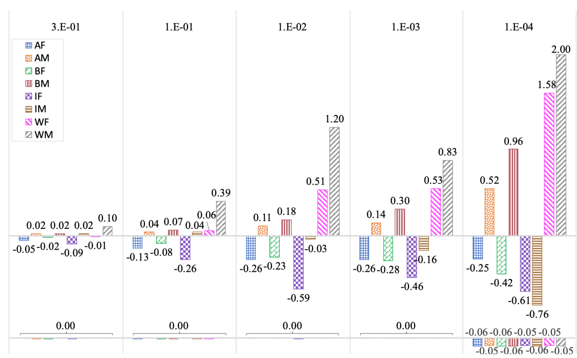

Fig 3. Percent difference from intended FPR. *Top:* single, optimal threshold. *Bottom:* Subgroup-specific thresholds reduces this difference to near zero [1].

Table 1. The BFW dataset statistics across each race & gender subgroup.

	Asian (A)		Black (B)		Indian (I)		White (W)		
	Female (AF)	Male (AM)	BF	BM	IF	IM	WF	WM	Aggregated
# Faces	2,500	2,500	2,500	2,500	2,500	2,500	2,500	2,500	20,000
# Subjects	100	100	100	100	100	100	100	100	800
# Faces / Subject	25	25	25	25	25	25	25	25	25
# Positive Pairs	30,000	30,000	30,000	30,000	30,000	30,000	30,000	30,000	240,000
# Negative Pairs	85,135	85,232	85,016	85,141	85,287	85,152	85,223	85,193	681,379
# Pairs (Total)	115,135	115,232	115016	115,141	115287	115,152	115,223	115193	921,379

Impact

Fairness

Promote fairness in facial recognition models via convention for measuring bias and encouraging such evaluations in other AI subfields.

Transparency

Provide non-technical users a method for understanding the limits of their model via interactive dashboard **(Fig. 4)** and auto-report generation.

Extensibility

Allow efficient inclusion of additional soft attributes of interest (e.g. age, hair color) to tailor bias evaluation.

Awareness

Further important ongoing conversations in AI regarding the equity and fairness of progress made.

† Equal contribution (ordered alphabetically)

[1] Joseph P. Robinson, et al. "Face Recognition: Too Bias, or Not Too Bias?." in *arXiv preprint arXiv:*2002.06483 (2020).