

Demographic Effects of Race on Face Recognition

Presented by:

Dr. Michael King, Associate Professor

Department of Computer Engineering and Sciences

Florida Institute of Technology

November 27th, 2018

Collaborative Research effort with

Dr. Kevin Bowyer, University of Notre Dame

Growing concerns about use of FR

- There have also been growing expressions of concern over “bias” in facial recognition algorithms from the Privacy and Civil Liberties community as well as several influential federal legislators.
- The Georgetown Law Center on Privacy & Technology has published reports detailing concerns relative to border security and law enforcement and use of face recognition.
- A widely viewed TED talk by Joy Buolamwini: “How I’m fighting bias in algorithms”



Prior Research on the topic

- **Research related to face recognition accuracy varying between demographic groups goes back to at least FRVT 2002.**
- **A number of scholarly research publications through the years.**
- **However, there is relatively little work on this topic, and it more documents that effects exist than explains why they exist.**

Objective (project ongoing)

Investigate root causes of variations in performance accuracy of automated face recognition systems.

Motivating questions:

- 1. Verify what variations exist?**
- 2. Understand why do they exist?**

Experiment Details

- 2 COTS algorithms + 1 open source CNN (ResNet)
- Limited publicly available data ideally suited to support this study
- Utilized the Morph Album2 Dataset collected by UNC-W
- Mugshot quality data used primarily to support academic research on facial aging

Sample Images



Summary of Curated dataset

| | Original Total Images | No Face Visible - Dropped | Duplicate Image - Dropped | Labeling Error - Dropped | Labeling Error - Corrected | Total Images Dropped | Curated Dataset |
|------------------|-----------------------|---------------------------|---------------------------|--------------------------|----------------------------|----------------------|-----------------|
| Album 2 | 53,633 | 259 | 140 | 3 | 18 | 402 | 53,231 |
| African-American | 42,897 | 176 | 105 | 0 | 9 | 281 | 42,616 |
| Caucasian | 10,736 | 83 | 35 | 3 | 9 | 121 | 10,615 |

MORPH Dataset: Curation Issues

- No face / partial face



- Duplicate images



Original Label: Caucasian
New Label: African-American

Original Label: African-American
New Label: Caucasian

- Mis-labeled



RESULTS: Face Detection

Failure to Enroll (FTE) with MORPH Dataset

- COTS A:**

| | Total Images | FTE Cases |
|-------------------------|--------------|----------------|
| Album 2 | 53,231 | 135 (0.25%) |
| African-American subset | 42,616 | 109 (0.26%) |
| Caucasian subset | 10,615 | 26 (0.24%) |



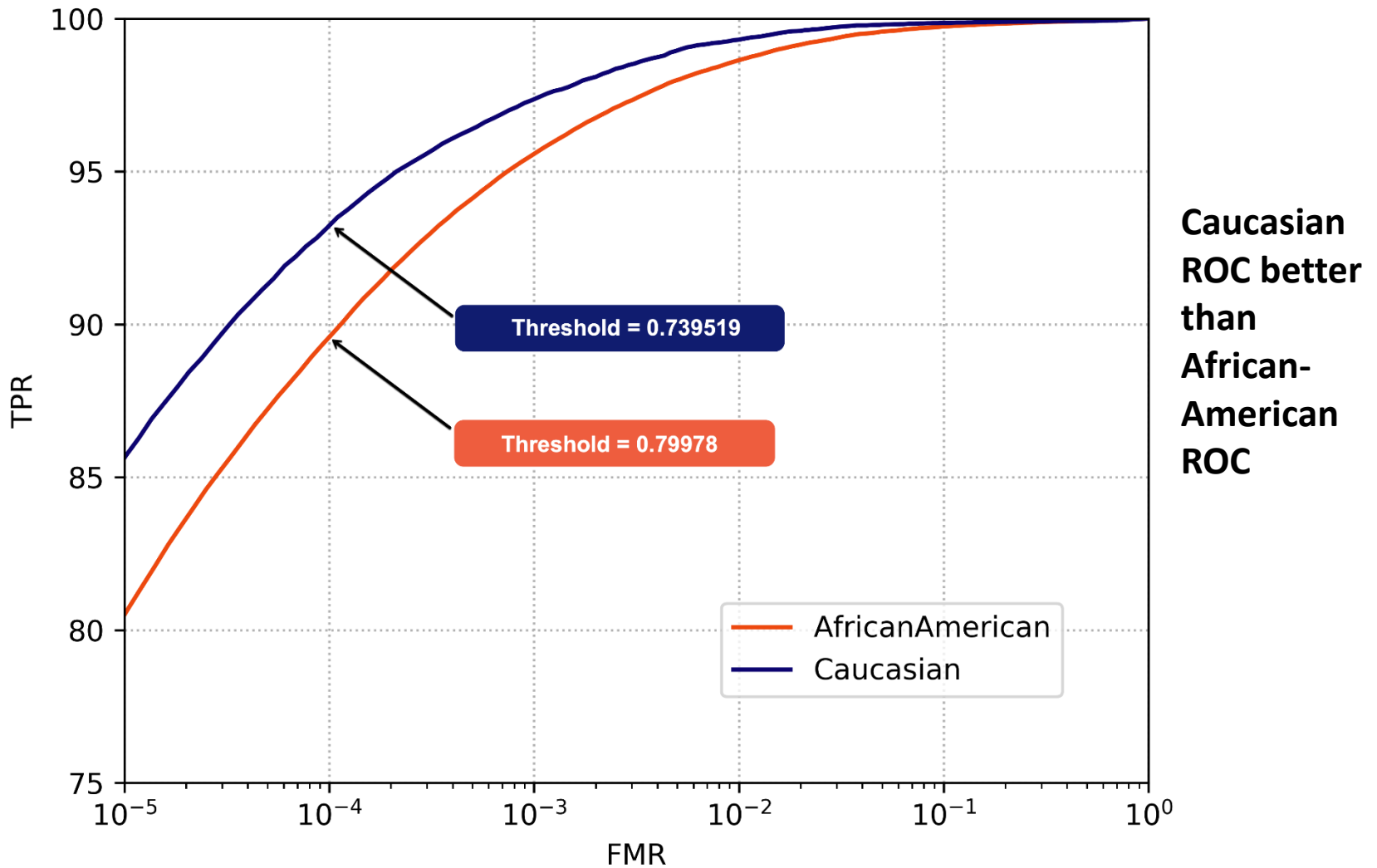
- COTS B:**

| | Total Images | FTE Cases |
|-------------------------|--------------|----------------|
| Album 2 | 53,231 | 5 (0.0093%) |
| African-American subset | 42,616 | 4 (0.0093%) |
| Caucasian subset | 10,615 | 1 (0.0094%) |

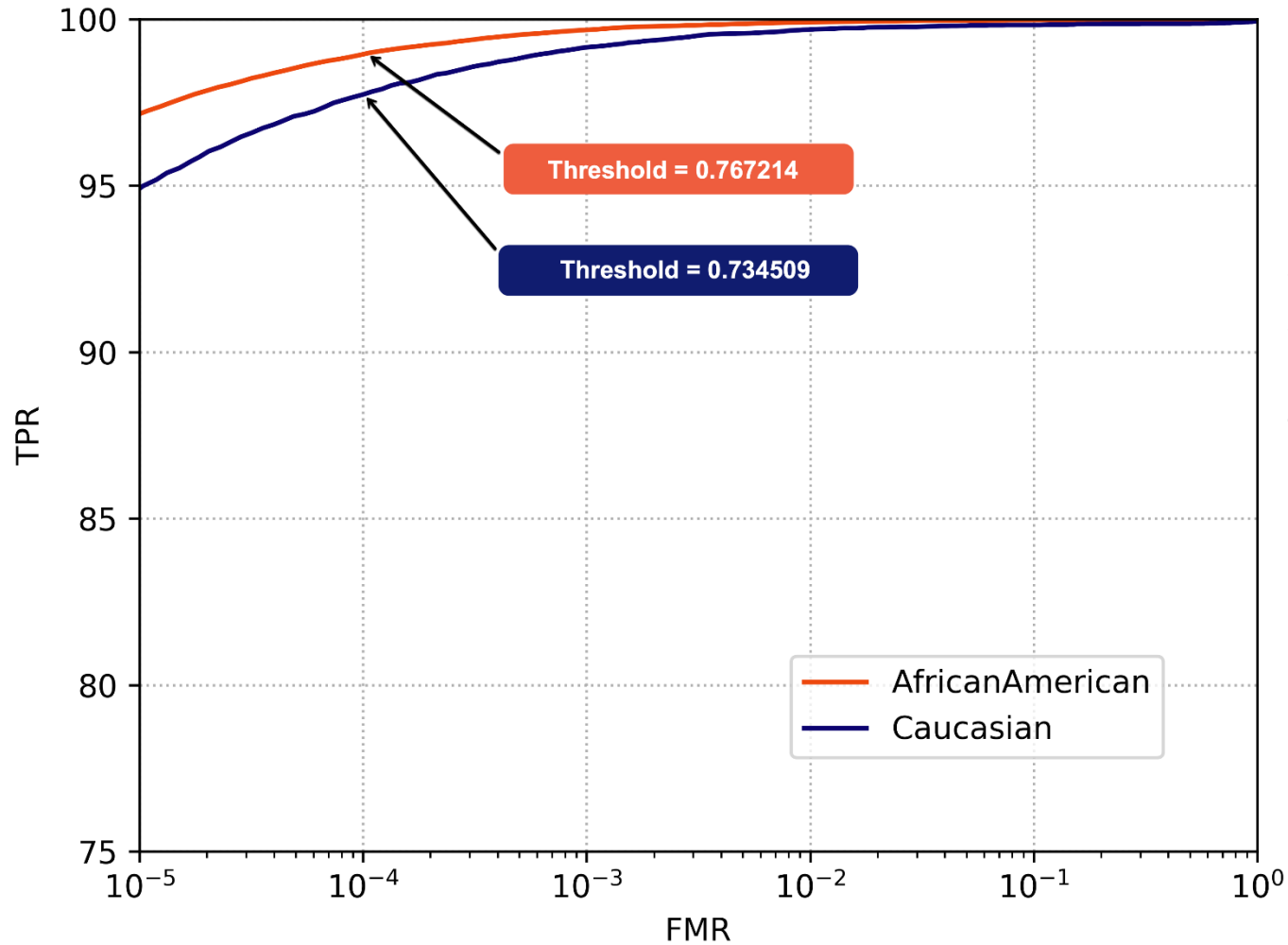


FTE rates are very similar

COTS-A ROC Curves (Race)

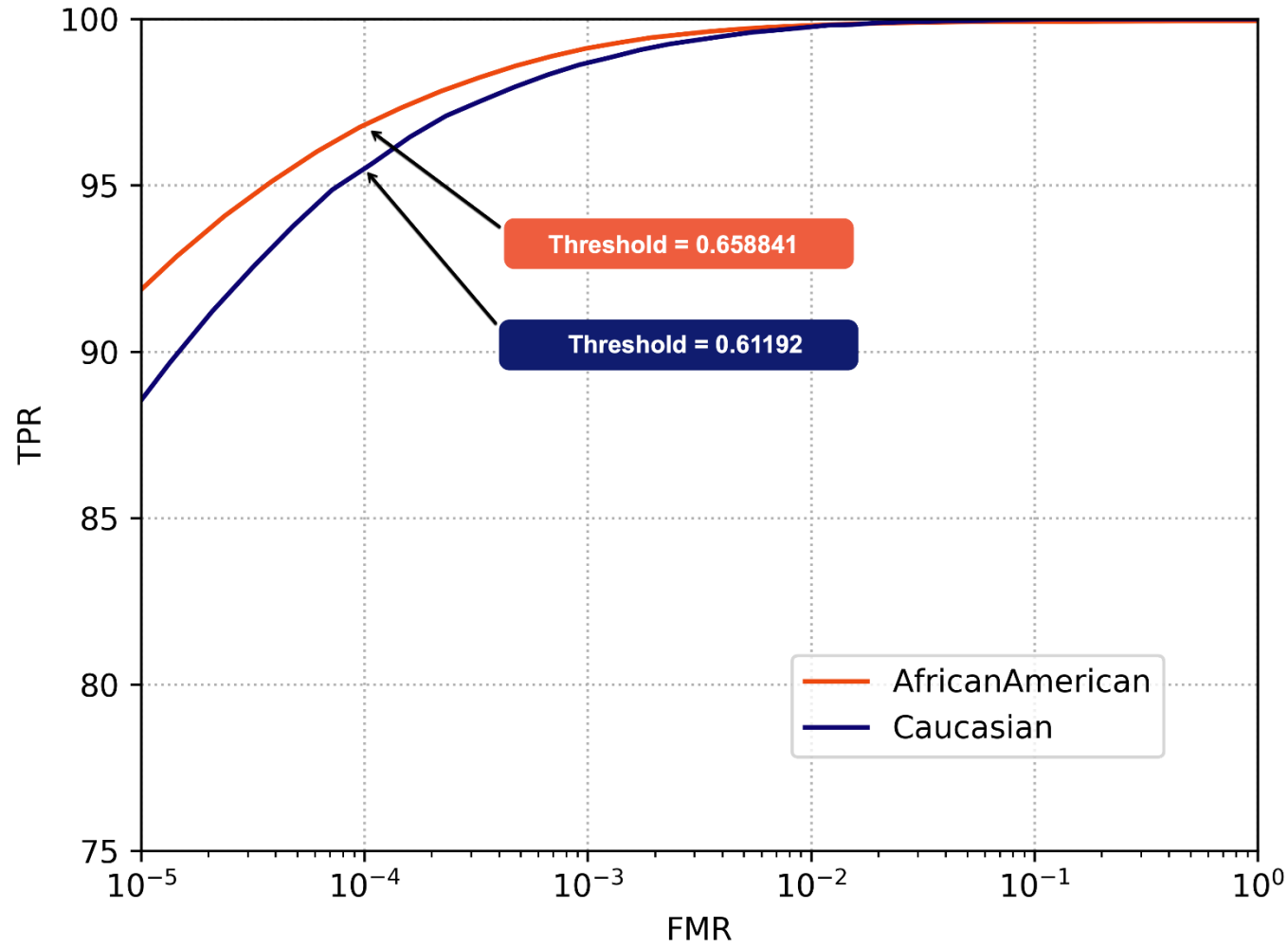


COTS-B ROC Curves (Race)



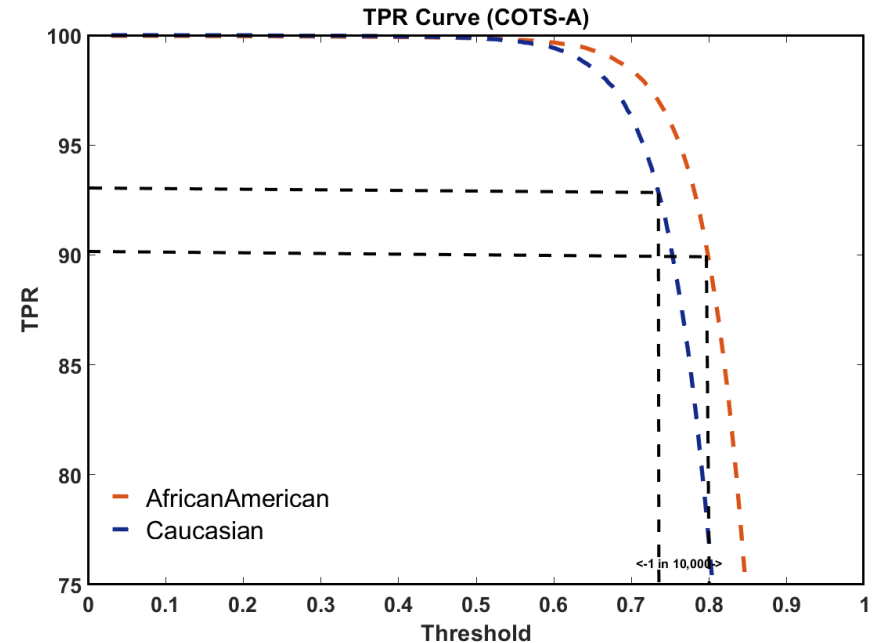
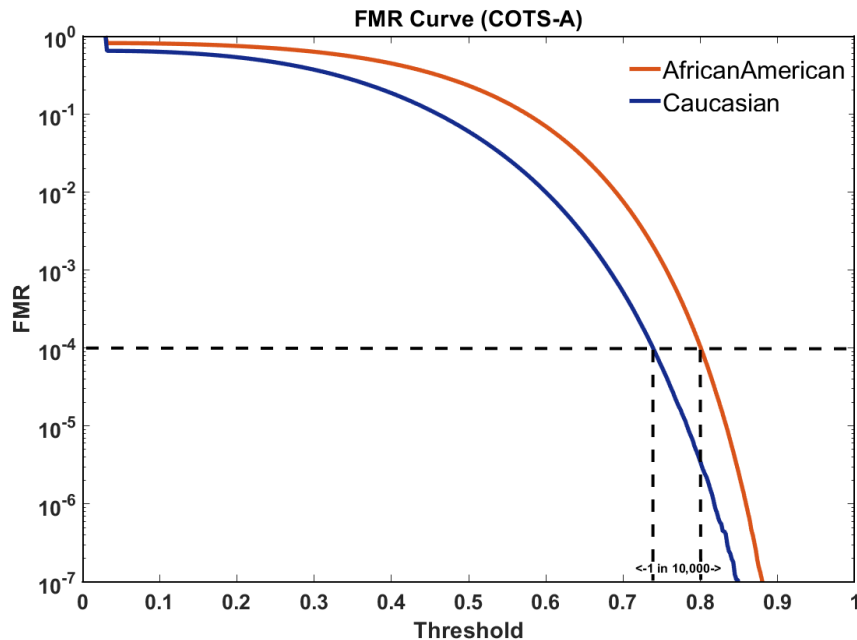
African-American ROC better than Caucasian ROC

ResNet ROC Curves (Race)



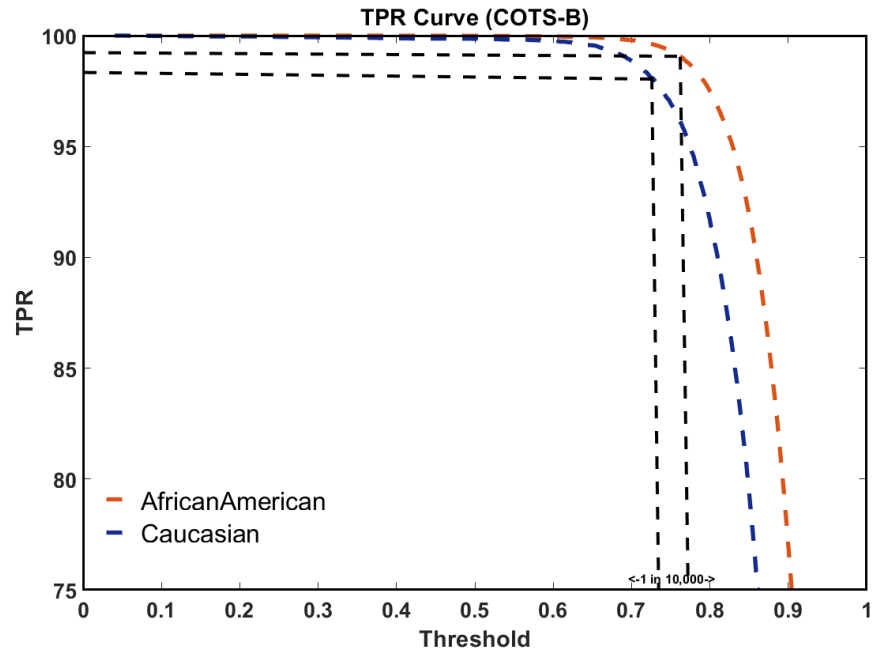
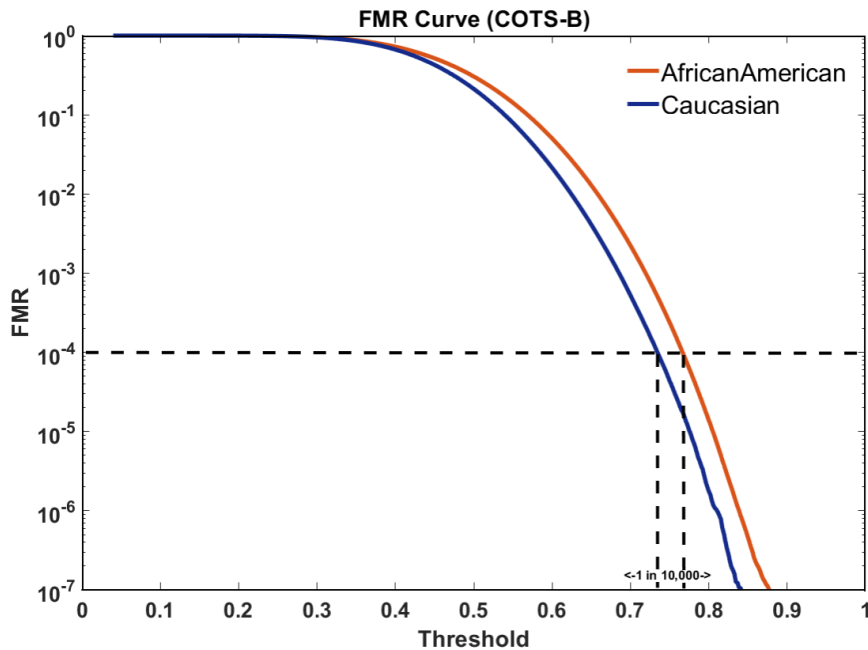
African-American ROC better than Caucasian ROC

COTS-A FMR and TPR Curves (Race)



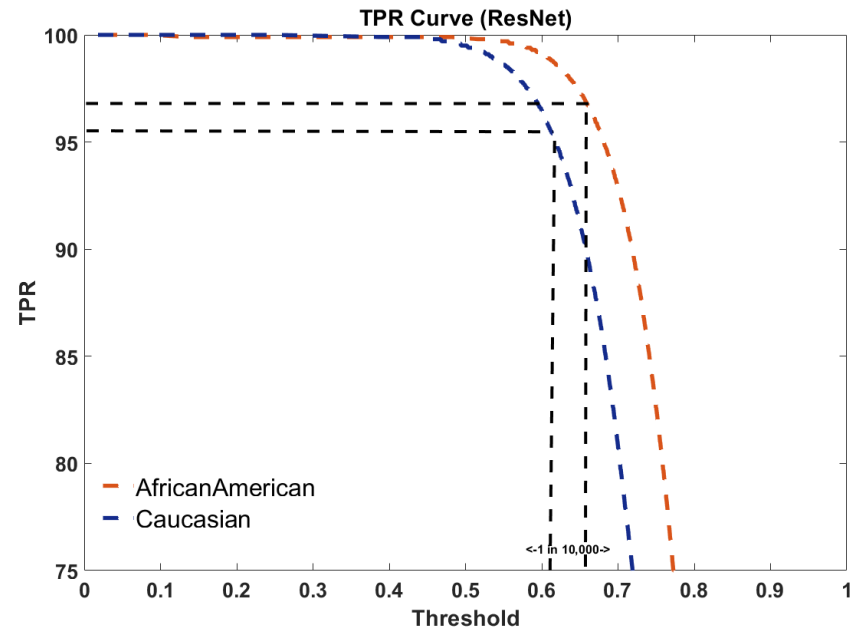
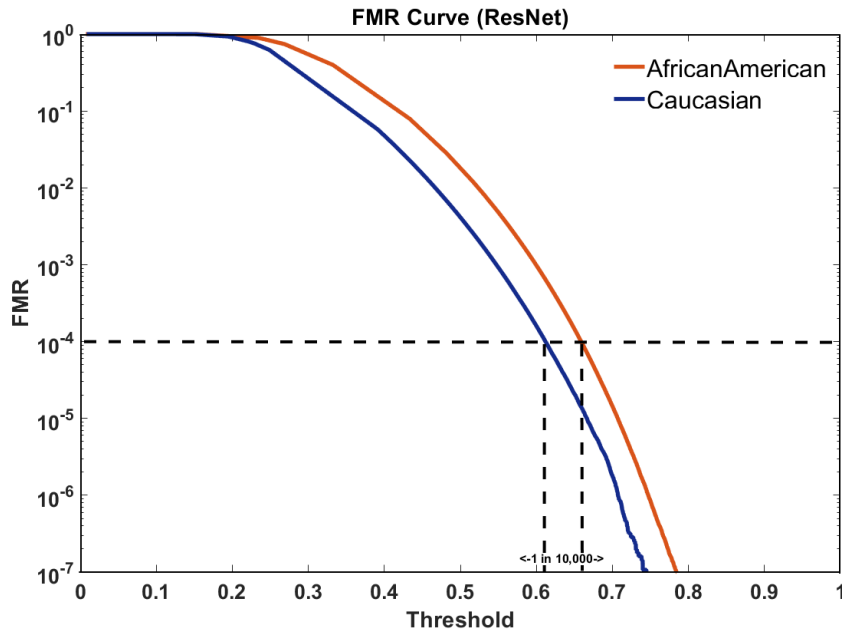
- For a desired FMR, the threshold setting would need to be different for each demographic group.
- On Morph, at a fixed decision threshold value, we find that the African-American scores generally result in a higher FMR and TPR than the Caucasian scores.

COTS-B FMR and TPR Curves (Race)



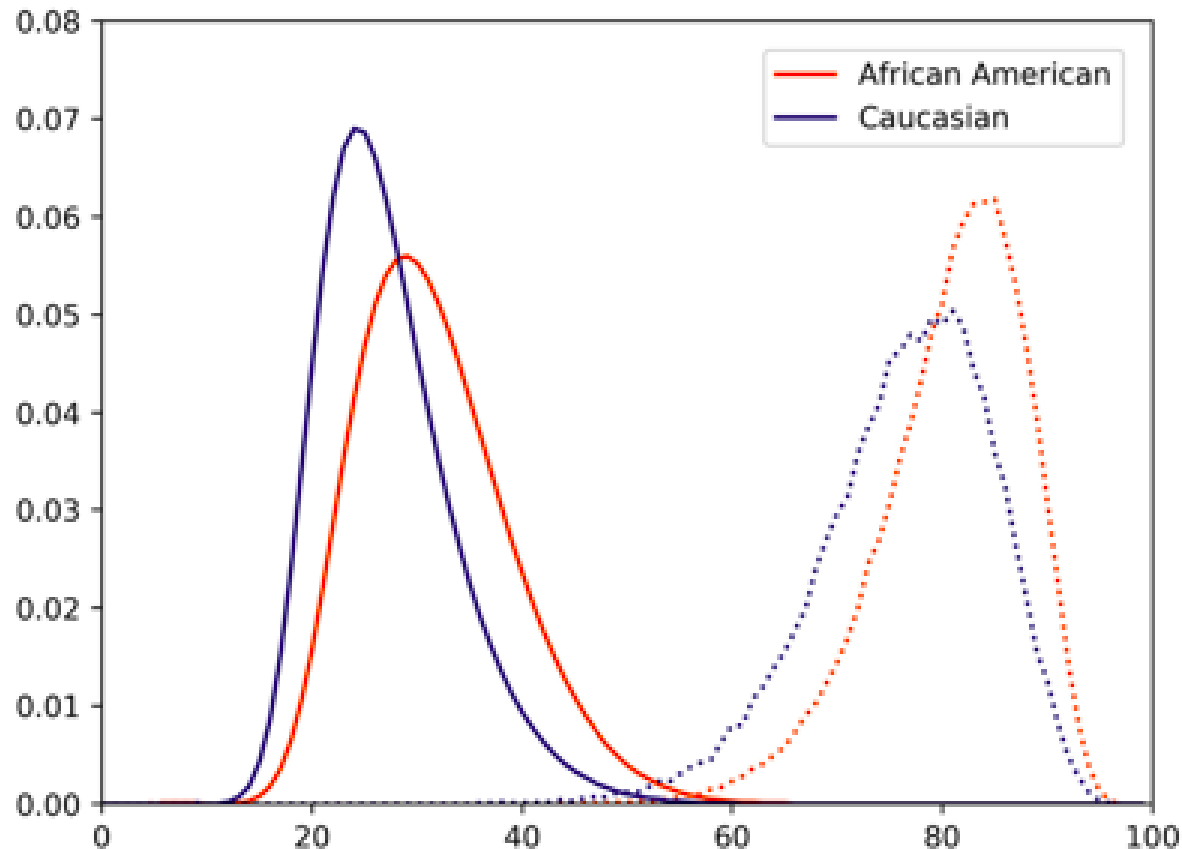
- For a desired FMR, the threshold setting would need to be different for each demographic group.
- On Morph, at a fixed decision threshold value, we find that the African-American scores generally result in a higher FMR and TPR than the Caucasian scores.

ResNet FMR and TPR Curves (Race)



- For a desired FMR, the threshold setting would need to be different for each demographic group.
- On Morph, at a fixed decision threshold value, we find that the African-American scores generally result in a higher FMR and TPR than the Caucasian scores.

ResNet Authentic & Imposter Distributions (Race)



African-American Impostor and Authentic both shifted to higher similarity scores.

Image Quality and ICAO Compliance

- Used IFace SDK to check for ICAO compliance
- 48% of the African-American and 57% of Caucasian images are rated as ICAO compliant.



072933_0F37



271018_03M21



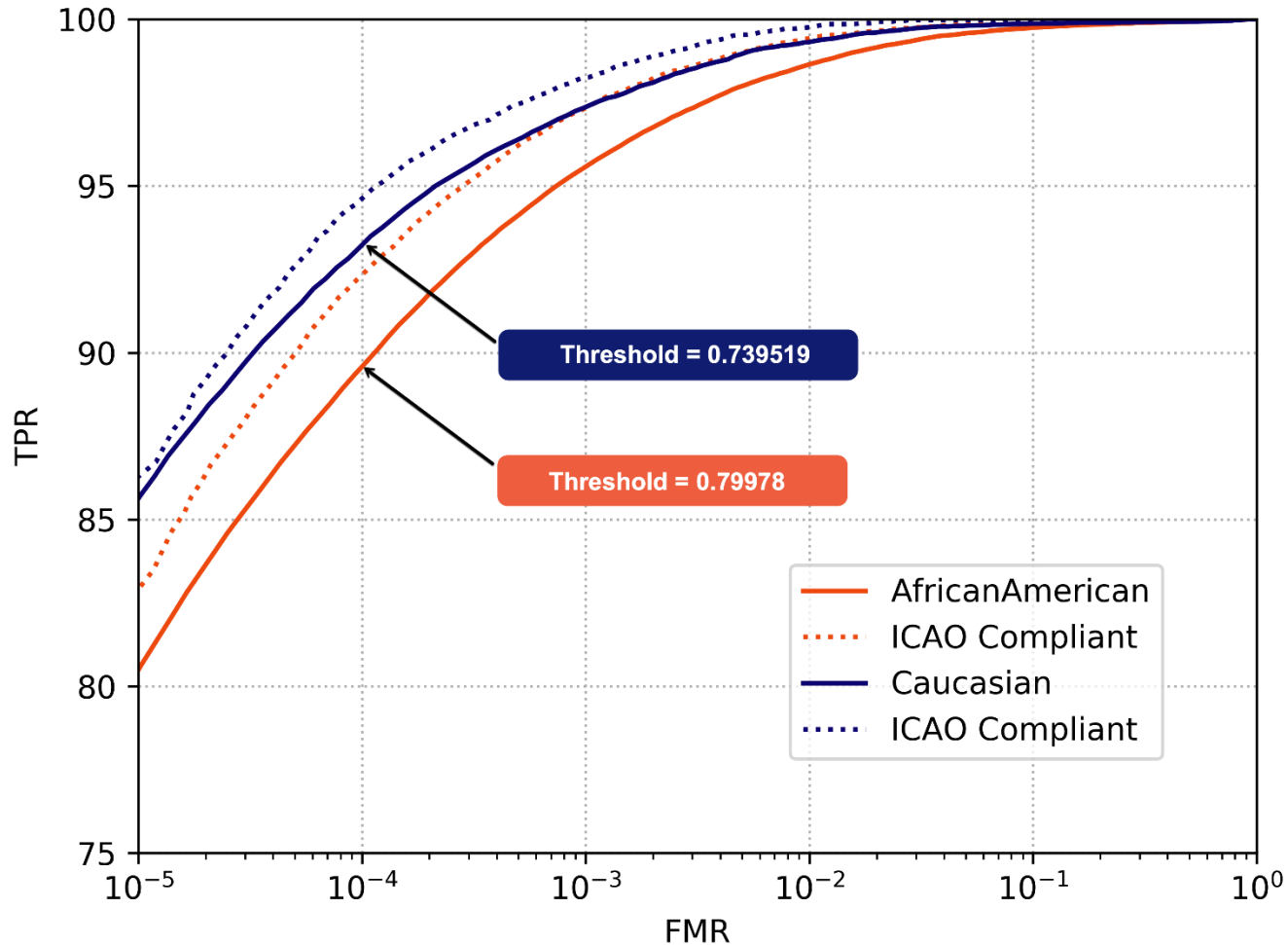
311448_01F31



042384_2M49

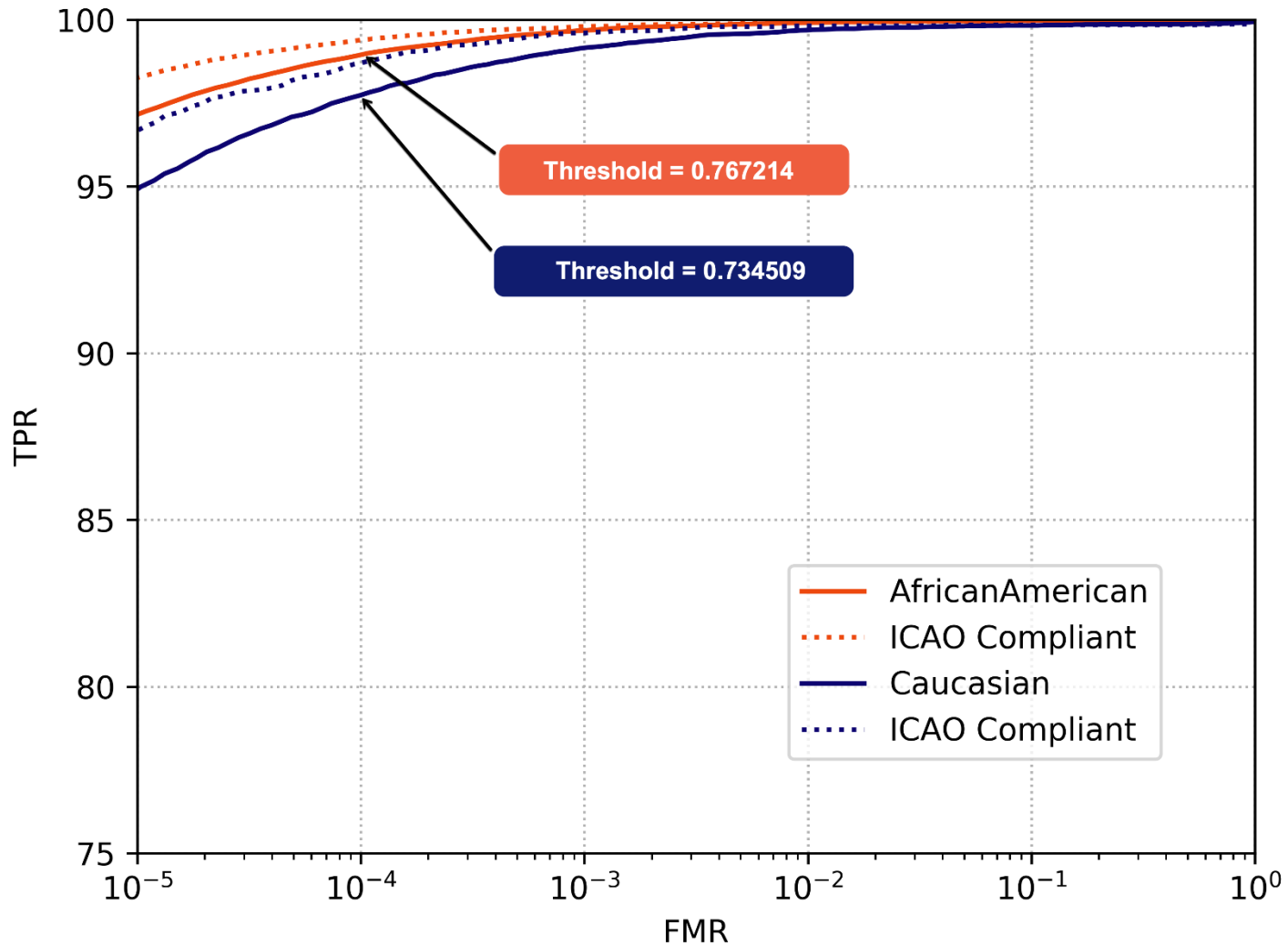
**Example Images In Curated MORPH Dataset Rated Not ICAO-Compliant
By IFace SDK (extended)**

COTS-A ROC Curves (ICAO)



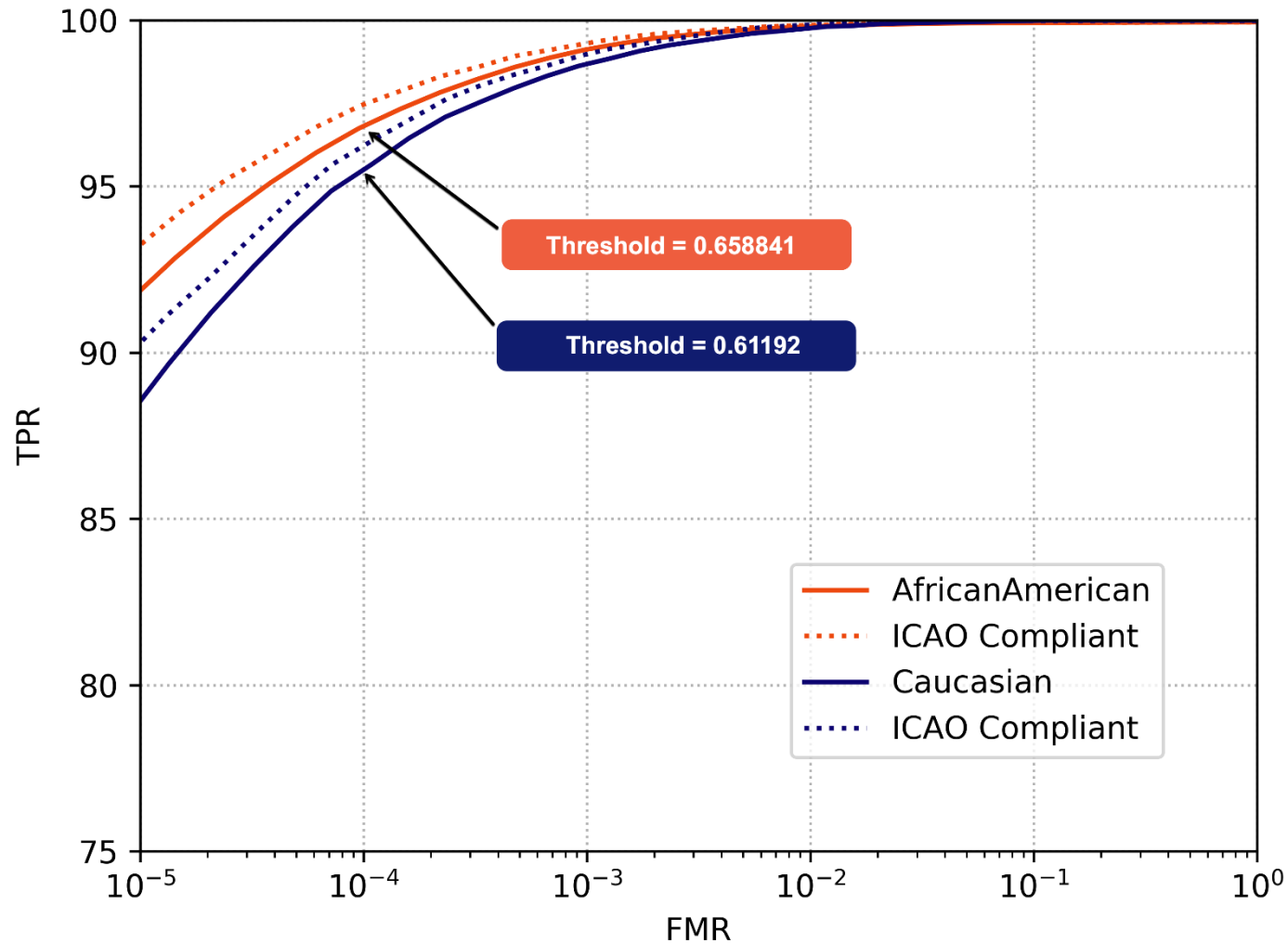
A simple ICAO compliance check improves performance and decreases gap between curves.

COTS-B ROC Curves (ICAO)



A simple ICAO compliance check improves performance

ResNet ROC Curves (ICAO)



A simple ICAO compliance check improves performance

Takeaways:

On the MORPH dataset,

- **COTS A results in a better ROC curve for Caucasians than for African-Americans.**
- **COTS B and ResNet result in a better ROC curve for African-Americans than for Caucasians.**
- **ROC curves hide that comparison between groups is not achieved at the same decision threshold. A given TPR and FMR line drawn across the plot intersects the curves at different decision threshold values.**
- **A simple quality gate improves accuracy.**



Thank You!!!

Michael King, Ph.D.
Florida Institute of Technology
Melbourne, Florida, USA

E: michaelking@fit.edu

O: (321) 674-8636

C: (703) 598-6991