Bias in Facial Recognition

Daniel Alexander Dittlbacher-Hupf
Rodrigo Leeb
Louis Jörg Gerhard Rösch
Tomasz Bronislaw Wisniewski

Januar 2022

Inhalt

Face Detection

- High-level Überblick
- Object detection
- Suche nach bestimmten Merkmalen
- Zuerst nach Augen suchen
- Erfolg hängt vom Bild ab
- Machine learning

Facial Recognition

- Face detection
- Normalisierung
- Merkmale aufstellen
- Mit der Datenbank abgleichen

Verwendungsmöglichkeiten

- Authentifizierung
- Videoüberwachung
- Photographie

Bias – Definition

- Bezeichnet die variierende Erfolgsrate eines Face Recognition Systems unter bestimmten Bedingungen
- Wir betrachten Bias in verschiedenen demographischen Gruppen

Bias – Definition

- Entsteht u. a. durch
  - Einsatz eines Systems in Bedingungen, für die das System nicht entwickelt wurde
  - Enorme Komplexität und Variation des Problems
  - Große homogene Anteile in Trainingsdatensätzen

Bias in Ausgangsdaten

- Lerndaten entscheiden über Genauigkeit
- Rassismus bzw. Ungleiche Repräsentation
- Sexismus
- In weiterer Folge unfaire Behandlung

Bias in Ausgangsdaten

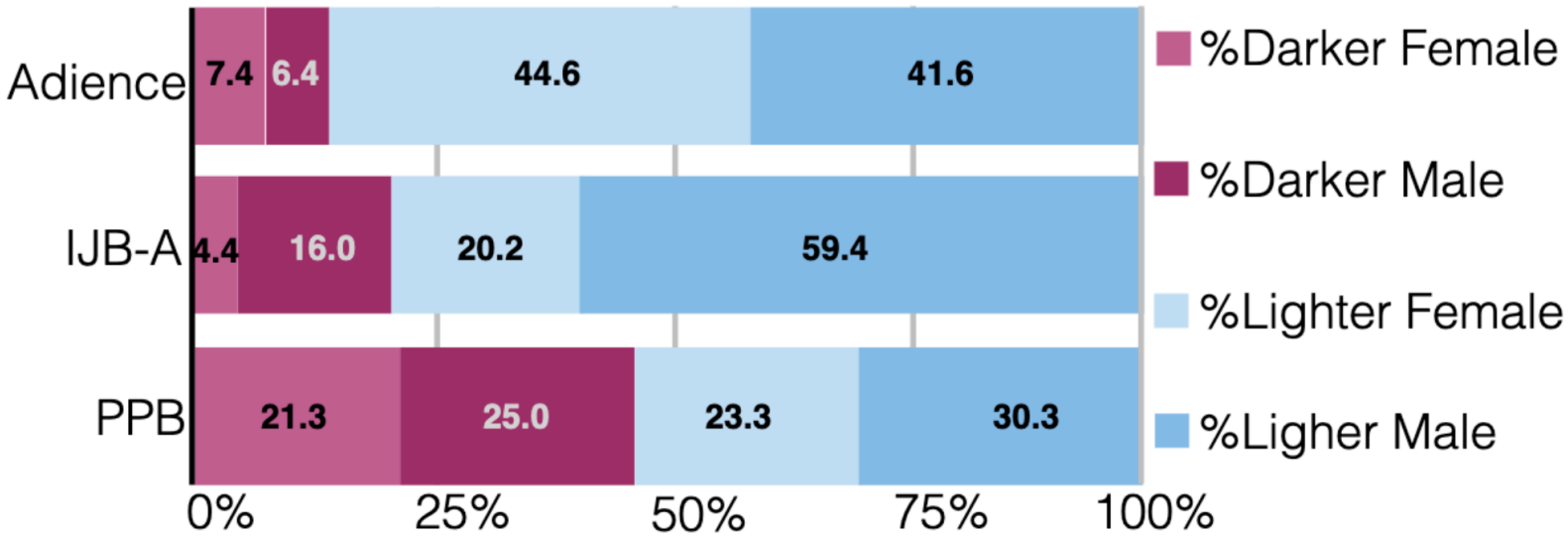

Abbildung: Bias in Ausgangsdaten

Folgen

- Ungleiche Erkennung
- Falsch positive Resultate
- Einstellung von Gesichtserkennungssystemen

Psychologische Faktoren

- Algorithmen wurden per Hand erstellt
- Man orientiert sich am eigenen Aussehen
- Studie erkannte Unterschiede bei Algorithmen aus verschiedenen Nationen

Joy Buolamwini

- Eine der führenden Forscherinnen in diesem Gebiet
- War selbst betroffen

Regulierung?

- Öffentliche Forderung nach Regulierung durch
  - Menschenrechtsorganisationen
  - Forscher
  - Technologiekonzerne

Regulierung?

- Denn FRT ist:
  - fehlerhaft
  - intransparent
  - diskriminierend

‘It’s dangerous, racializing, and has few legitimate uses; facial recognition needs regulation and control on par with nuclear waste.’

Ziele

- Gewährleistung des Datenschutzes
- Diskriminierung (durch Bias) verhindern
- ‘Automation Bias’ verhindern

Situation in Europa

- 'AI-Act'
- Risikoorientierter Ansatz
- Klassifizierung der AI-Anwendungen in 4 Klassen anhand von
  - Sensibilität des betroffenen Lebensbereichs
  - Impact der AI-Entscheidung
- **Zusätzlicher Faktor Bias-Potential**

Situation in Europa

$\longrightarrow$ FRT ist ein ‘Hochrisiko-System’

Anforderungen für FRT

- Nutzung durch Strafverfolgungsbehörden nur in Ausnahmefällen
- Konformitätsprüfung für das Produkt samt Zertifizierung
  - $\longrightarrow$ CE-Label für ‘AI-Software’

Anforderungen für FRT

- Anforderungen
  - Technische Dokumentation
  - Risikomanagement
  - Protokollierung
  - Gute Qualität der verwendeten Daten
  - Mitigation Systems
  - Transparenz
  - Angemessene Genauigkeit

Regulierung in den USA

- Keine Gesetze auf Ebene des Bundes
- Stattdessen Urteile und Gesetze für einzelne Staaten/Städte
- Illinois, Washington & Texas
  - Verbot der Nutzung ohne Einwilligung
- **San Francisco**
  - Nutzung durch Polizei untersagt (2019)
  - CCPA (2020)

Bias – Mitigation Methods

- Annahme: nur eine Bildkomposition in Datensätzen
- Offensichtlicher Ansatz: uniforme Trainingdatensätze
  - es ist schwer zu bewerten, was uniform bedeutet

Bias – Mitigation Methods

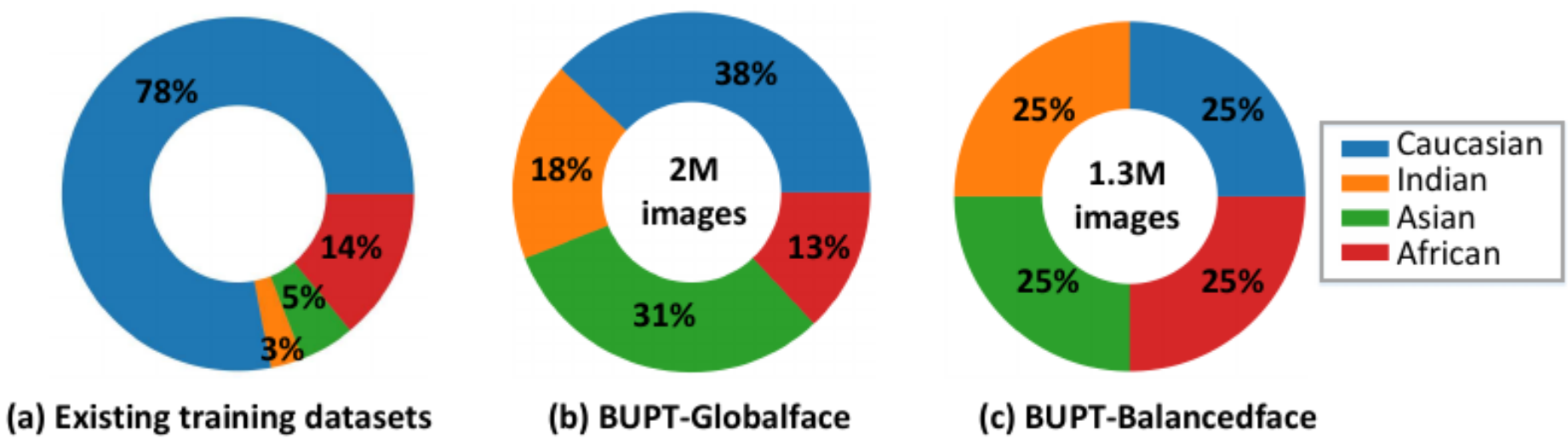

Abbildung: Trainingsdatensätze ('Neural Learning from Unbalanced Data'; Sixue Gong, et al.)

Bias – Mitigation Methods

- Adaptive Margin Loss
  - zusätzliche Variablen werden eingeführt und beim Training zur Laufzeit angepasst
- Feature Disentanglement
  - es wird versucht, aus dem Output des Systems Features zu erheben

Mitigation Methods

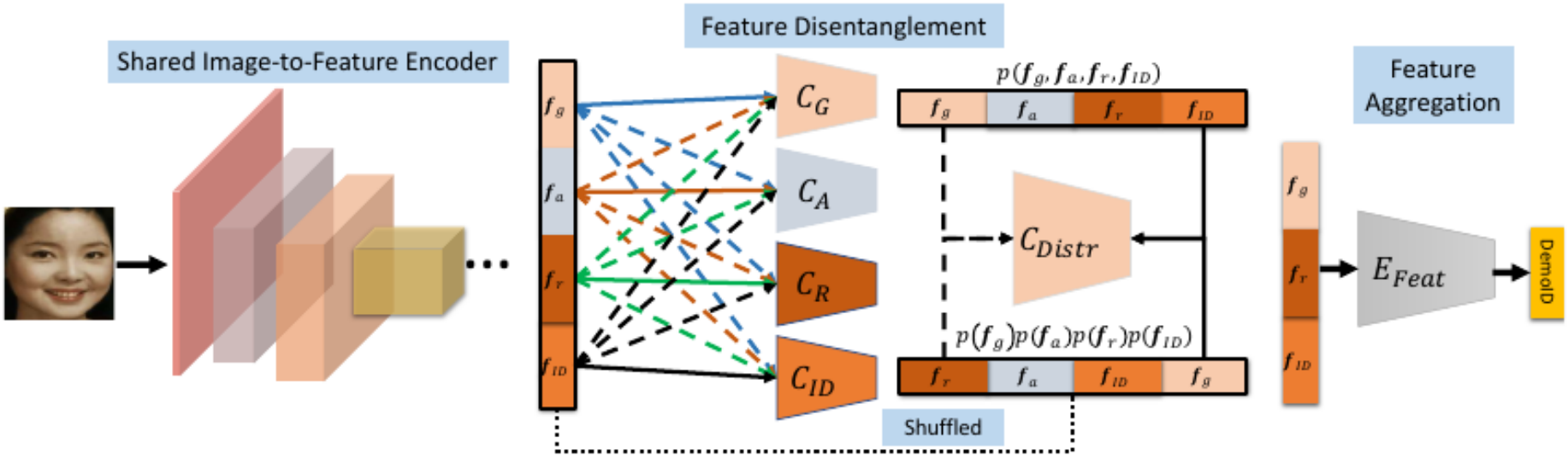

Abbildung: Feature Disentanglement ('Exploring Disentangled Feature Representation Beyond Face Identification'; Yu Liu et al.)

Quellen

1. Mitigating Bias in Face Recognition using Skewness-Aware Reinforcement Learning, Sixue Gong, et al. http://cvlab.cse.msu.edu/pdfs/gong_liu_jain_cvpr2021.pdf

2. Neural Learning from Unbalanced Data, Sixue Gong, et al. https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123740324.pdf

3. A Gentle Introduction to Deep Learning for Face Recognition, Jason Brownlee on machinelearningmastery.com https://machinelearningmastery.com/introduction-to-deep-learning-for-face-recognition/

4. Exploring Disentangled Feature Representation Beyond Face Identification, Yu Liu, et al. https://openaccess.thecvf.com/content_cvpr_2018/papers/Liu_Exploring_Disentangled_Feature_CVPR_2018_paper.pdf

Quellen

5 Bericht zur Gesichtserkennungstechnologie in Österreich, Amnesty International, https://www.amnesty.at/media/8397/amnesty_gesichtserkennungstechnologie-in-oesterreich_bericht-mai-2021.pdf

6 Open letter calling for a global ban on biometric recognition technologies that enable mass and discriminatory surveillance, Accessnow https://www.accessnow.org/ban-biometric-surveillance/

7 Silicon Valley has admitted facial recognition technology is toxic, John Naughton https://www.theguardian.com/commentisfree/2020/jun/13/silicon-valley-has-admitted-facial-recognition-technology-is-toxic-about-time

Quellen

8 Facial Recognition is the Plutonium of AI, Prof. Luke Stark, https://static1.squarespace.com/static/59a34512c534a5fe6721d2b1/t/5cb0bf02eef1a16e422015f8/1555087116086/Facial+Recognition+is+Plutonium+-+Stark.pdf

9 Vorschlag für eine VERORDNUNG DES EUROPÄISCHEN PARLAMENTS UND DES RATES ZUR FESTLEGUNG HARMONISIERTER VORSCHRIFTEN FÜR KÜNSTLICHE INTELLIGENZ (GESETZ ÜBER KÜNSTLICHE INTELLIGENZ) UND ZUR ÄNDERUNG BESTIMMTER RECHTSAKTE DER UNION. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX\%3A52021PC0206

10 BAN FACIAL RECOGNITION TECHNOLOGIES FOR CHILDREN AND FOR EVERYONE ELSE, LINDSEY BARRETT https://www.bu.edu/jostl/files/2020/08/1-Barrett.pdf

Quellen

11 Examining The San Francisco Facial-Recognition Ban, Tony Raval, Forbes
https://www.forbes.com/sites/forbestechcouncil/2019/06/21/examining-the-san-francisco-facial-recognition-ban/