

How good is my AI System?

Measures and Evaluation for Face Recognition

Dr.-Ing. Alexander Loos

Automatic Face Analysis

Motivation

Active research topic

- Age estimation
 - age groups
 - real age
- Accessories
 - glasses
 - hat
 - beard
 - masks
 - ...
- Gender classification
 - male
 - female
- Identification
 - recognition
 - verification
 - clustering
 - ...



Google Cloud



FAT

—
Face Analysis Technology

Automatic Face Analysis

Face Recognition Technology @ IDMT

Animal identification using face recognition

- **SAISBECO**

- „A Semi-Automated Audiovisual Species and Individual Identification System for Behavioral Research and Conservation“
- Collaboration between Fraunhofer and Max-Planck Institutes
- Identification and facial analysis of great apes
- Estimation and monitoring of population sizes

- **SIMSALACAT**

- Cat flap based on face recognition
- Including detection of prey in mouth
- Implemented on edge device (Jetson Nano)

- **BearID**

- Face recognition, age estimation, and gender classification of grizzly bears
- Collaboration with Applied Conservation Science Lab, Canada



Automatic Face Analysis

Face Recognition Technology @ IDMT

- **SPIRIT**

- EU funded project
- Collaboration with law enforcement agencies

- **Goal:**
Development of novel tools and approaches for resolving digital identities

- **Result:**
Platform that allows to create social graphs to perform social and criminal network analysis

<https://www.spirit-tools.com/>

© www.spirit-tools.com

SPIRIT

Scalable Privacy Preserving Intelligence Analysis for Resolving Identities

Automatic Face Analysis

Face Recognition Technology @ IDMT

- **FaceMatch – Climate Change in German TV Program**
 - *How salient are climate change & biodiversity in the German TV programme?*
 - (a) Which **representatives from politics, science, economy, law and culture** are visible within the climate discourse?
 - (b) How is the **share of men and women** in the climate discourse?
 - Recording of 20 German TV stations for two months (01.09.2022 – 31.10.2022), 19h/day (05:30 am – 0:30 am)
 - **37,476** recorded programs
 - **65 TB** video data
 - **23,181.7 h** video
 - **4,655,706** detected face-tracks
 - **806,390** distinct persons
 - **146** reference persons



FaceMatch

TV Program Study

Showing data from 2022-09-01 to 2022-10-31

Set date range:



Select Stations Display per (selected) station

Programme Summary

Property	Value
Number of stations	20
Number of recorded programmes	37,476
Video hours analysed	77 h
Faces detected	655
Overall in-record distinct persons	0
Female vs. Male percentage for all faces	72.5%
Minimal persons occurrence time	3.20 s
Maximal persons occurrence time	1635.60
Average persons occurrence time	05 s

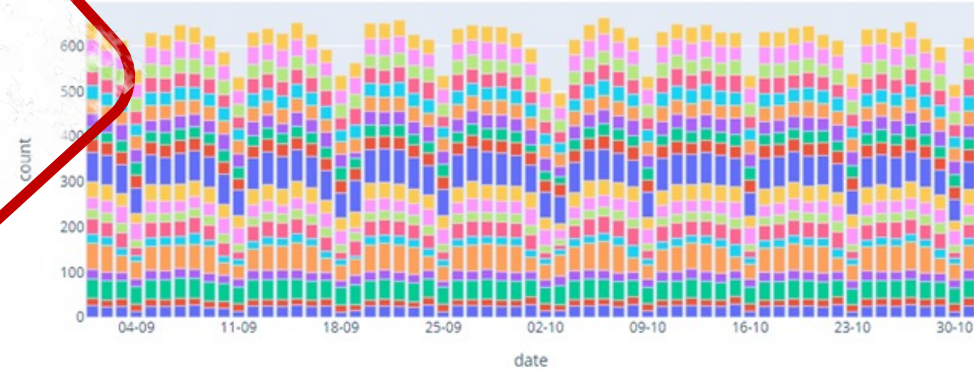


Match Summary

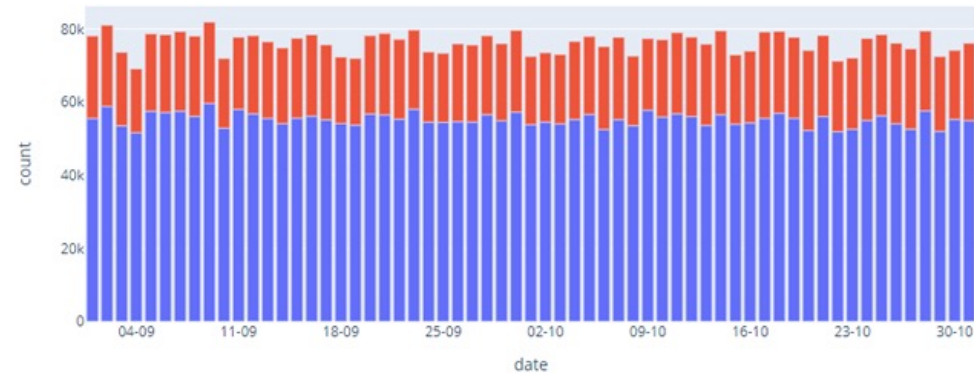
<input type="checkbox"/>	Reference person	<input type="checkbox"/>	Appearance time (min)	<input type="checkbox"/>	Number of programs with appearance
<input type="checkbox"/>	filter data...	<input type="checkbox"/>		<input type="checkbox"/>	
<input type="checkbox"/>	Olaf Scholz		1336.94		2227
<input type="checkbox"/>	Robert Habeck		713.47		1266
<input type="checkbox"/>	Christian Lindner		571.22		899
<input type="checkbox"/>	Karl Lauterbach		458.70		545
<input type="checkbox"/>	Friedrich Merz		373.83		335
<input type="checkbox"/>	Markus Söder		358.82		469
<input type="checkbox"/>	Stefan Weil		341.65		413
<input type="checkbox"/>	Annalena Baerbock		274.10		331
<input type="checkbox"/>	Ursula Von der Leyen		191.99		338
<input type="checkbox"/>	Christian Dürr		190.19		204

<< < 1 / 10 > >>

Number of recorded programmes



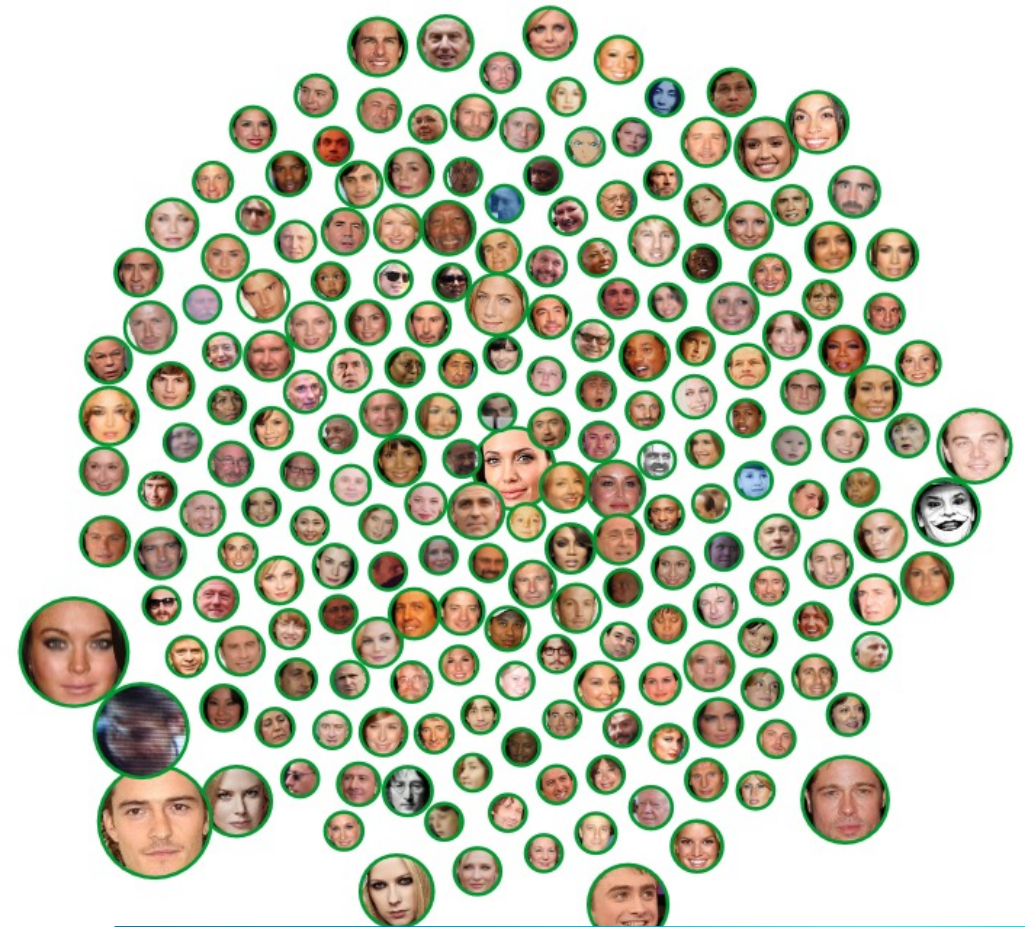
Number of Faces (Male/Female)



Automatic Face Analysis

Face Recognition Technology @ IDMT

- **FaceMatch – Unsupervised Face Clustering**
 - Grouping, recognition and search of persons in private image collections
 - Extracted information stored in local database
 - Data remains on host system: 100% compliance with GDPR regulations
 - Demo available for Windows 10 and as Docker Image for Linux



In-Cluster Accuracy

accurate

FaceMatch

Unsupervised Face Clustering

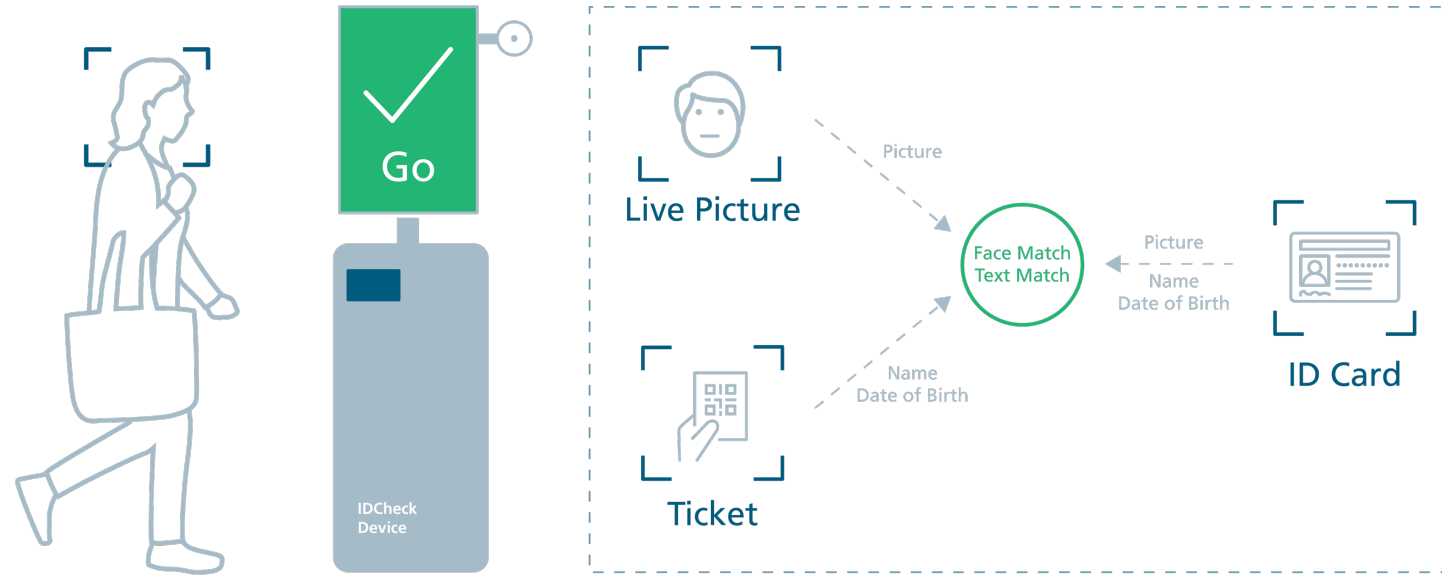
Automatic Face Analysis

Face Recognition Technology @ IDMT

▪ IDCheck

- Robust and privacy-compliant verification
- Extraction of personal data from any identity document
- Extraction of personal data from ticket or certificate
- Extraction of facial data via live camera on site
- Confirmation of an „identity match“ when facial and personal data correspond
- Implemented on edge device (Jetson Nano)

<https://www.idmt.fraunhofer.de/idcheck>



IDCHECK

IDCheck

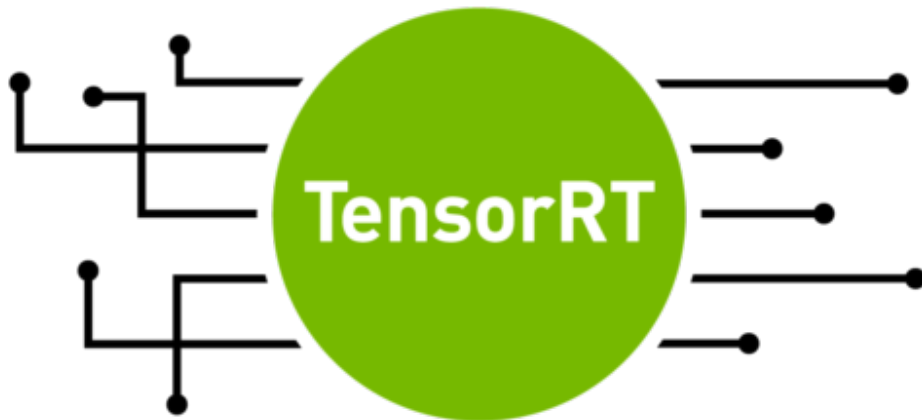
AI-based identity verification

Automatic Face Analysis

Face Recognition Technology @ IDMT

- **Nvidia Jetson Nano**

- Quad-core 64-bit ARM CPU
- 128-core NVIDIA GPU
- 4GB or 2GB LPDDR4 memory

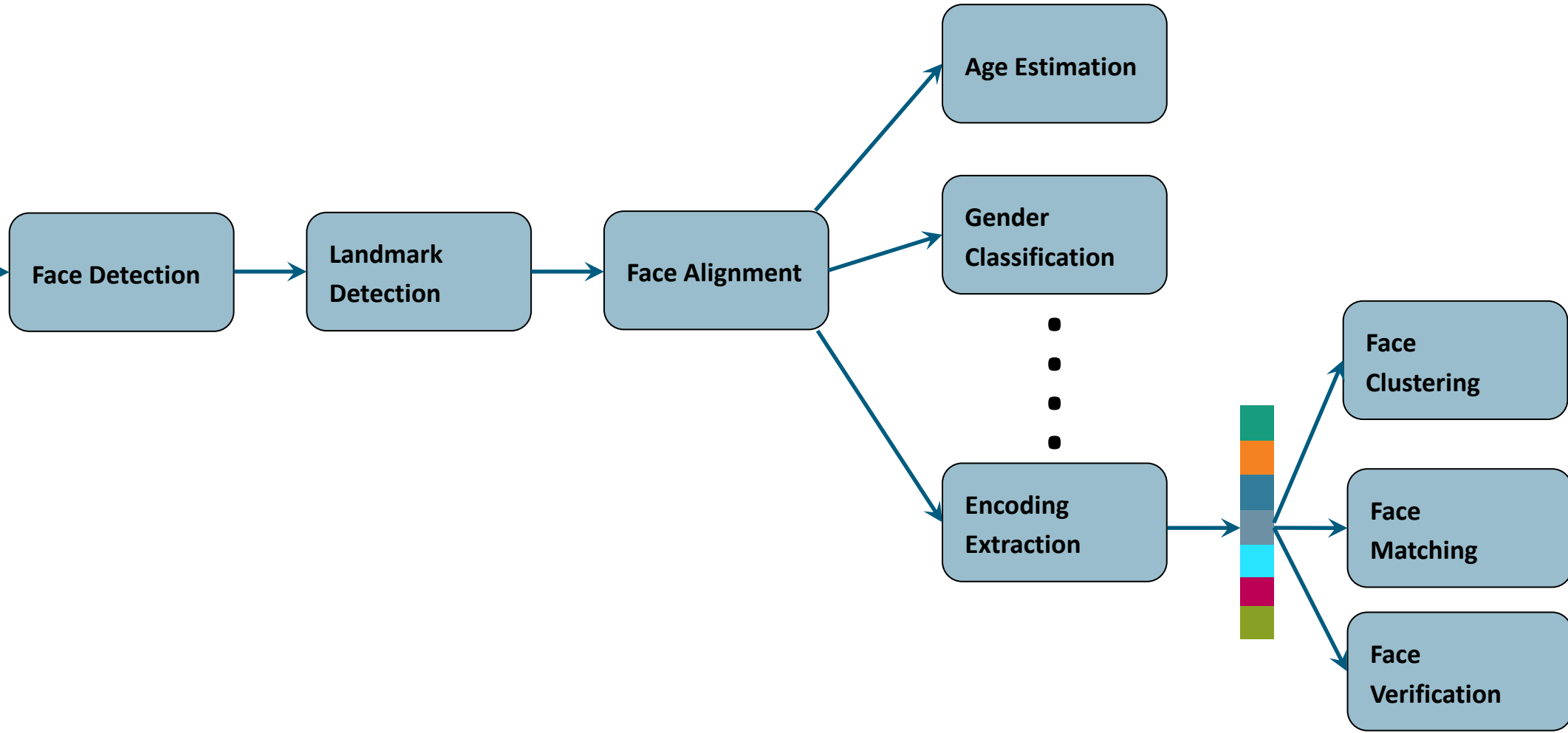


© <https://developer.nvidia.com/embedded/jetson-nano-developer-kit>

Automatic Face Analysis

Face Recognition Technology @ IDMT

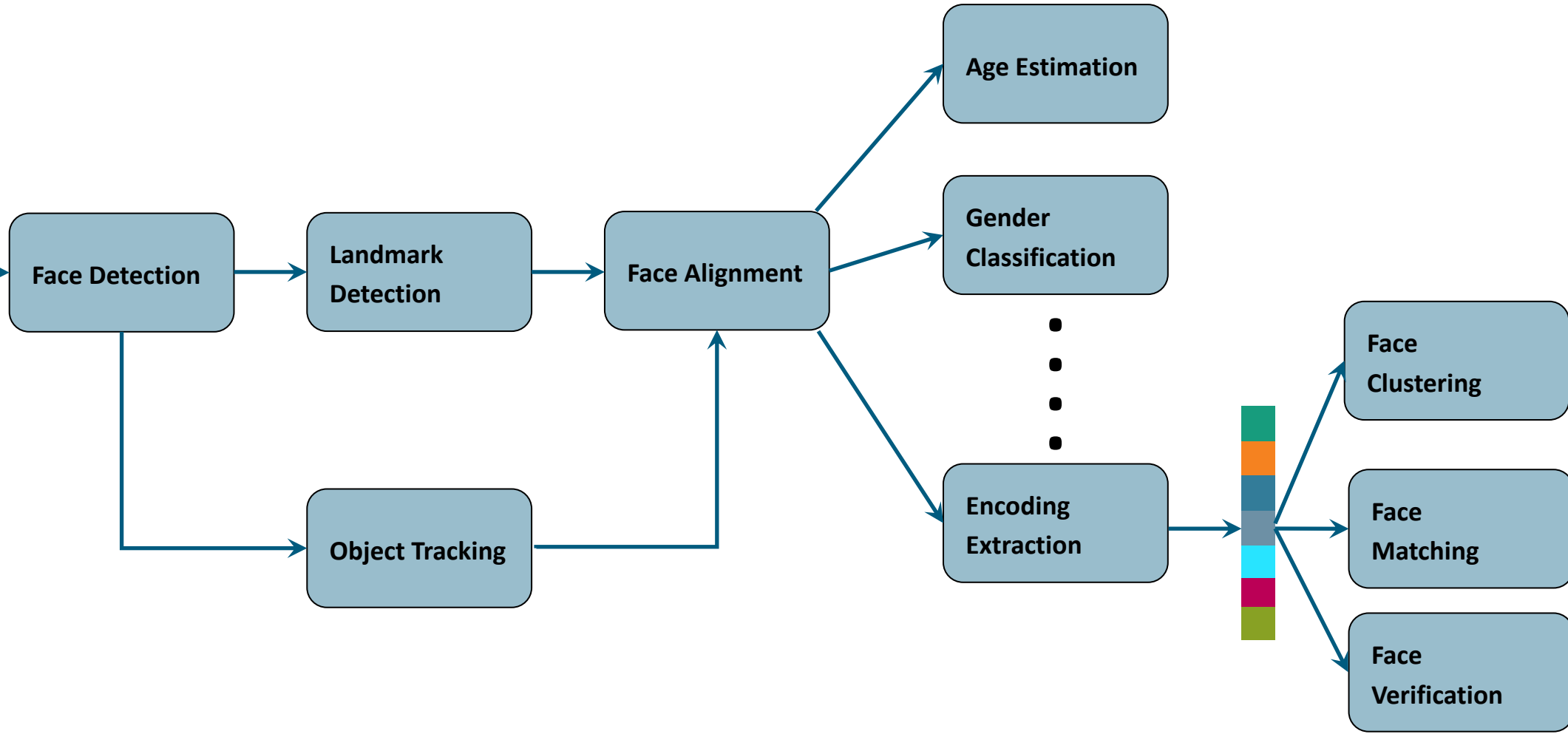
© Ellen DeGeneres/Twitter, Keystone

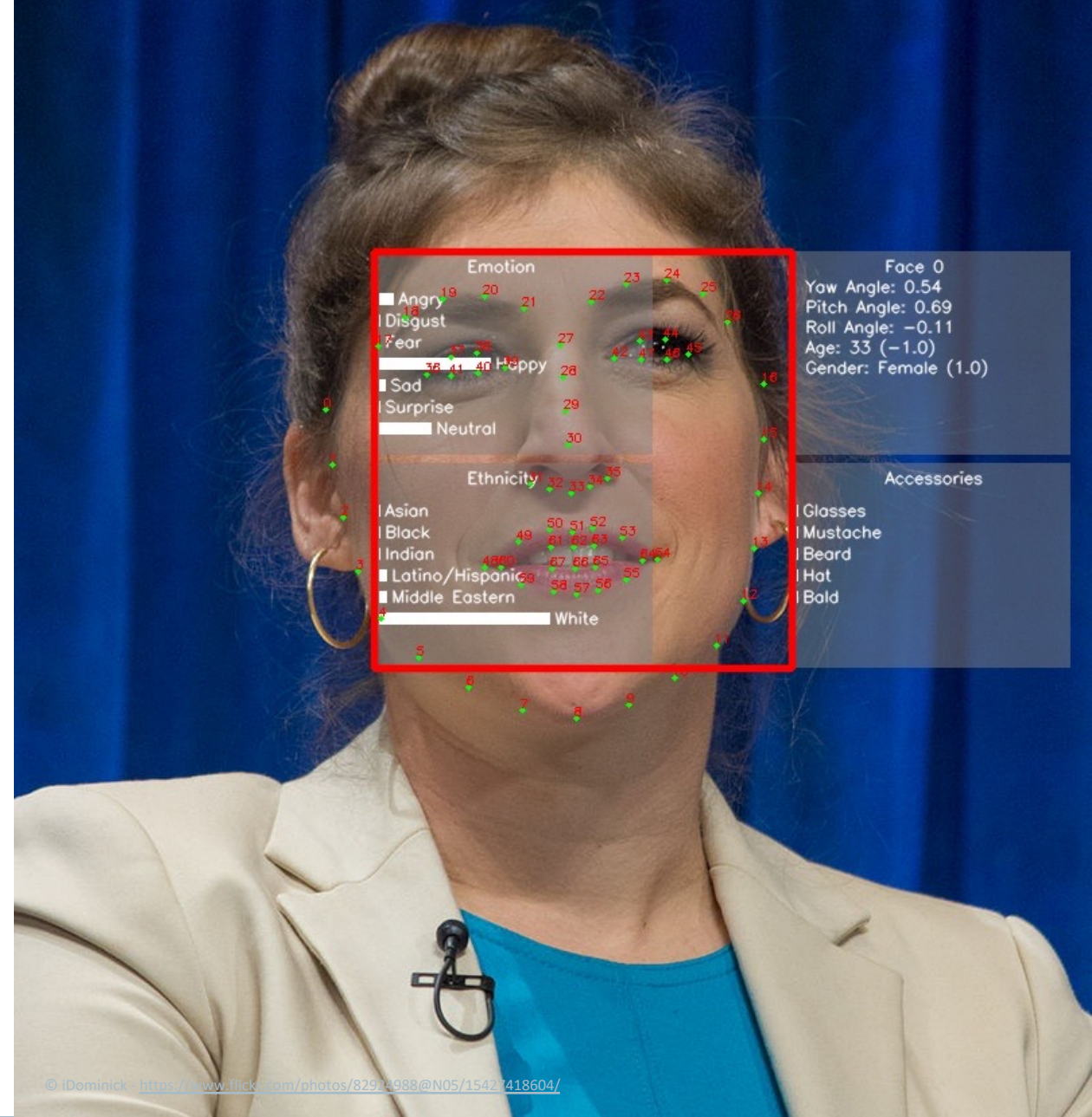
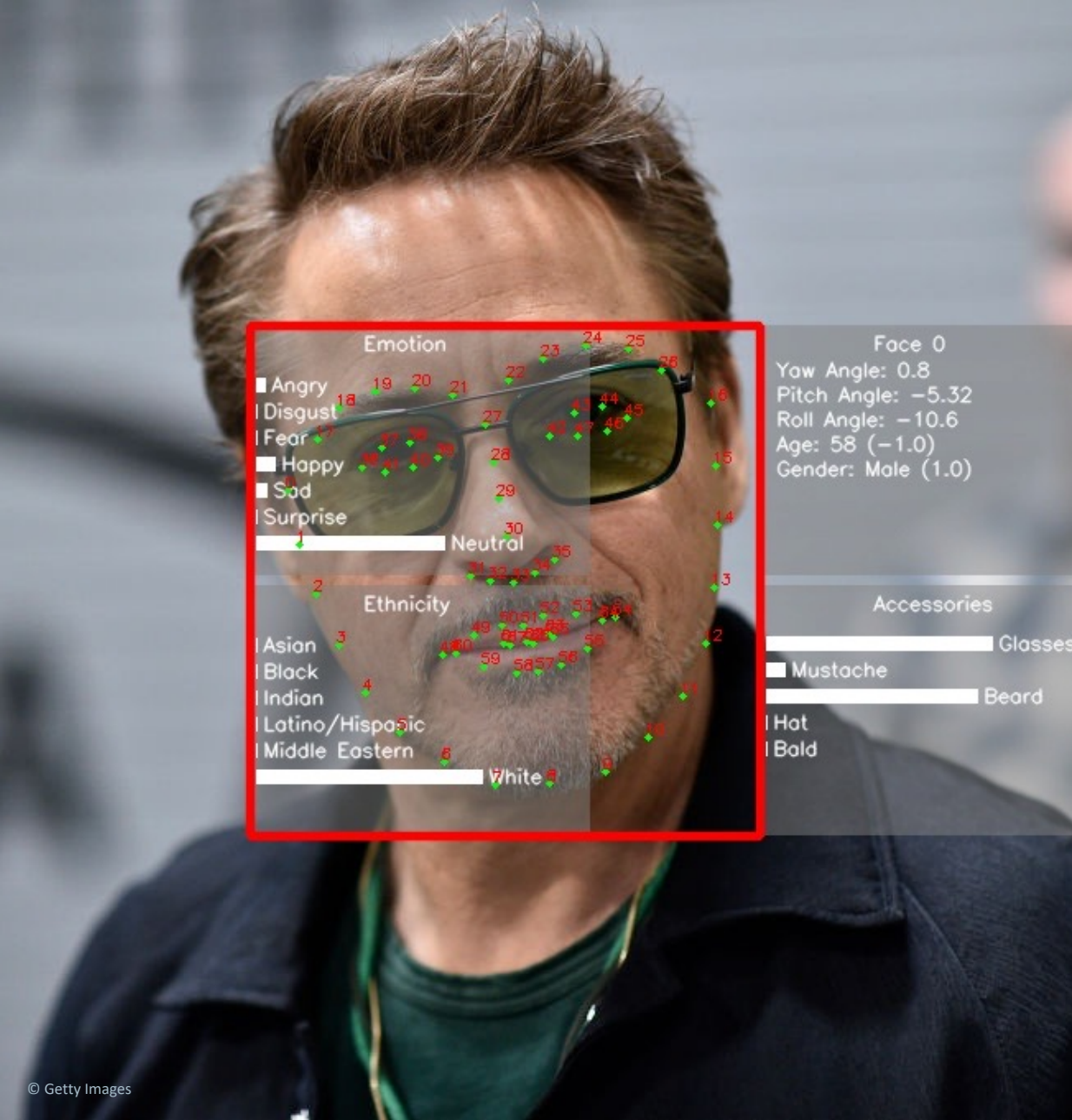


Automatic Face Analysis

Face Recognition Technology @ IDMT

© Ellen DeGeneres/Twitter, Keystone

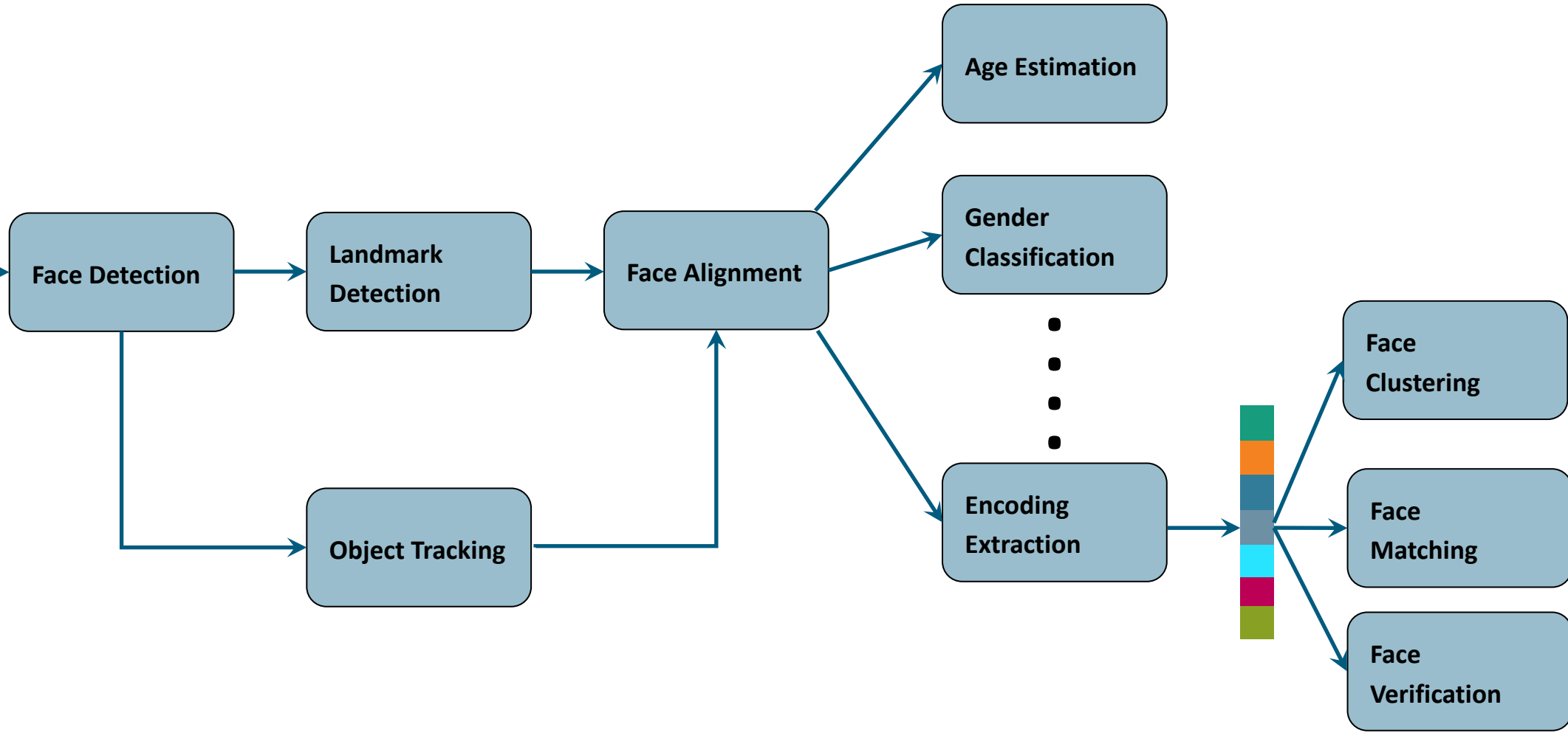




Evaluation of Face Analysis Algorithms

Performance Metrics

© Ellen DeGeneres/Twitter, Keystone

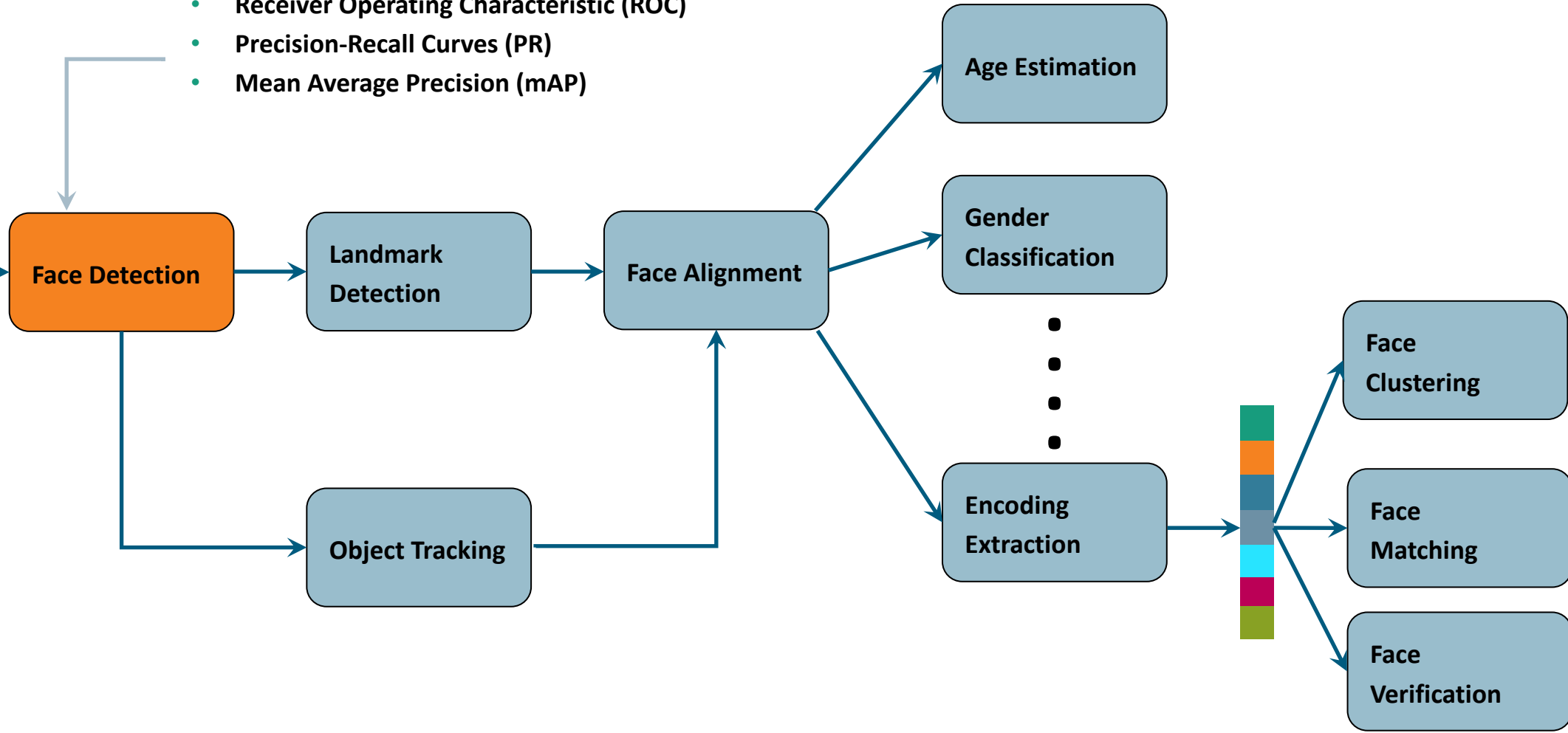


Evaluation of Face Analysis Algorithms

Performance Metrics

- Receiver Operating Characteristic (ROC)
- Precision-Recall Curves (PR)
- Mean Average Precision (mAP)

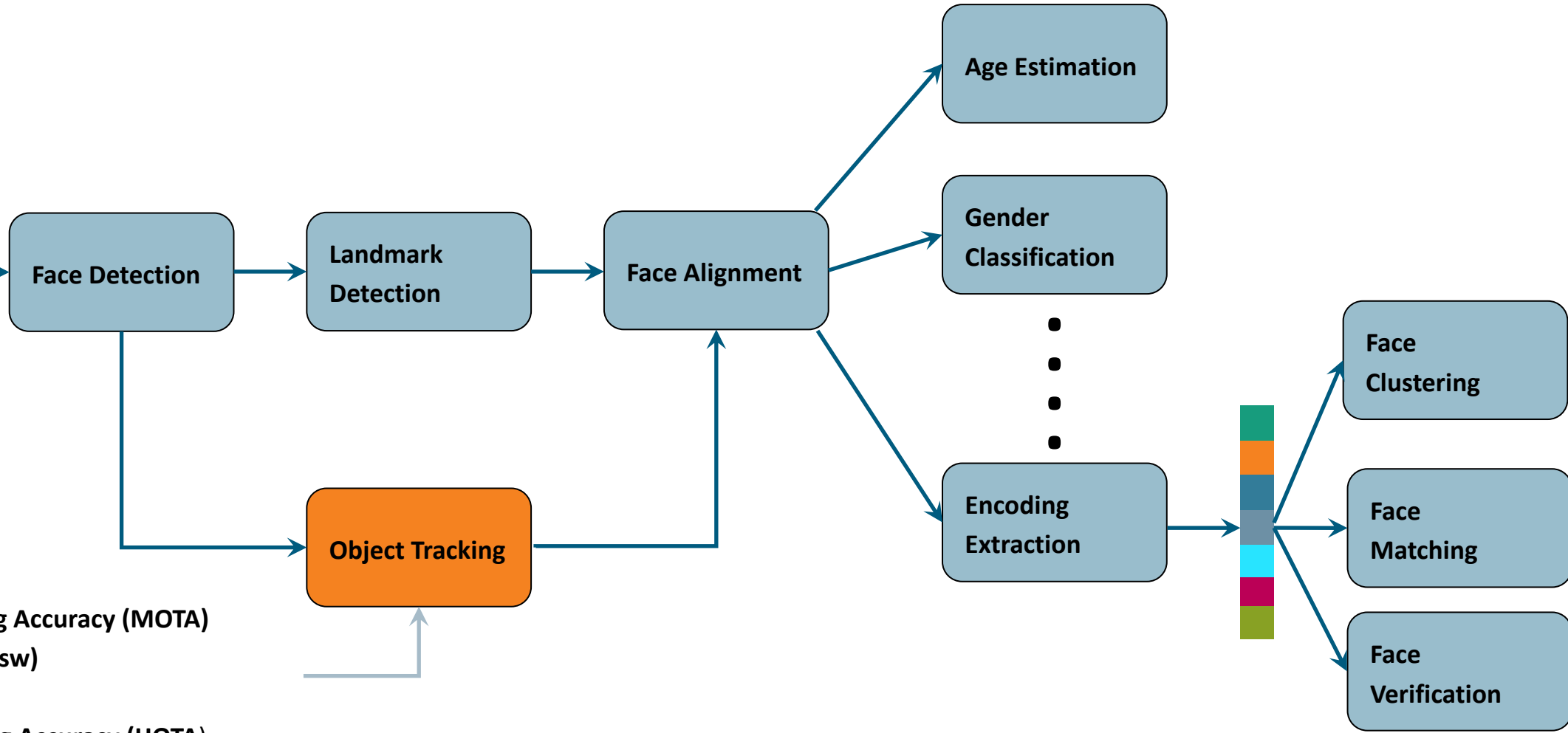
© Ellen DeGeneres/Twitter, Keystone



Evaluation of Face Analysis Algorithms

Performance Metrics

© Ellen DeGeneres/Twitter, Keystone

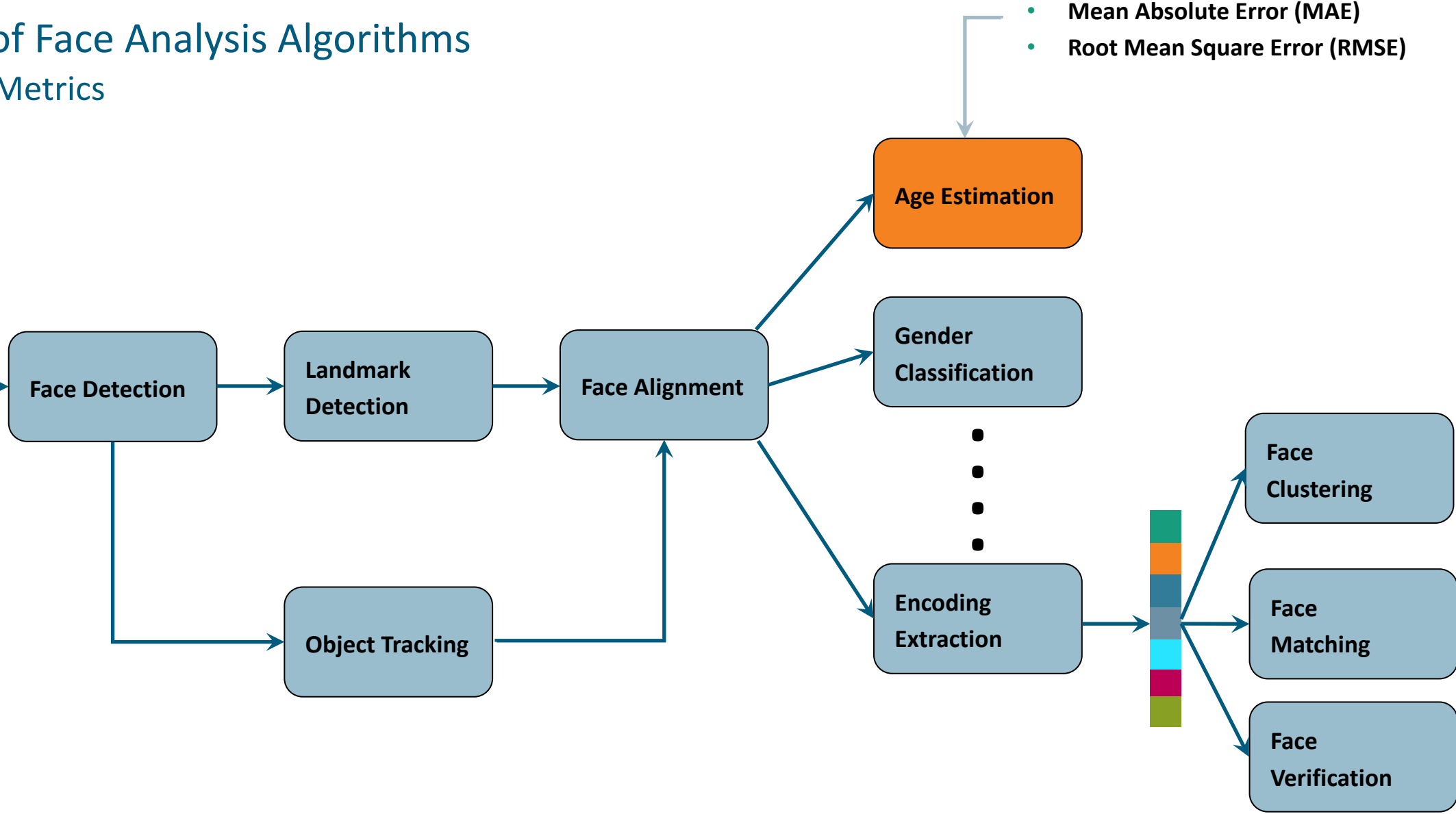


- Multi Object Tracking Accuracy (MOTA)
- Identity Switches (IDsw)
- Fragmentation (FM)
- Higher Order Tracking Accuracy (HOTA)

Evaluation of Face Analysis Algorithms

Performance Metrics

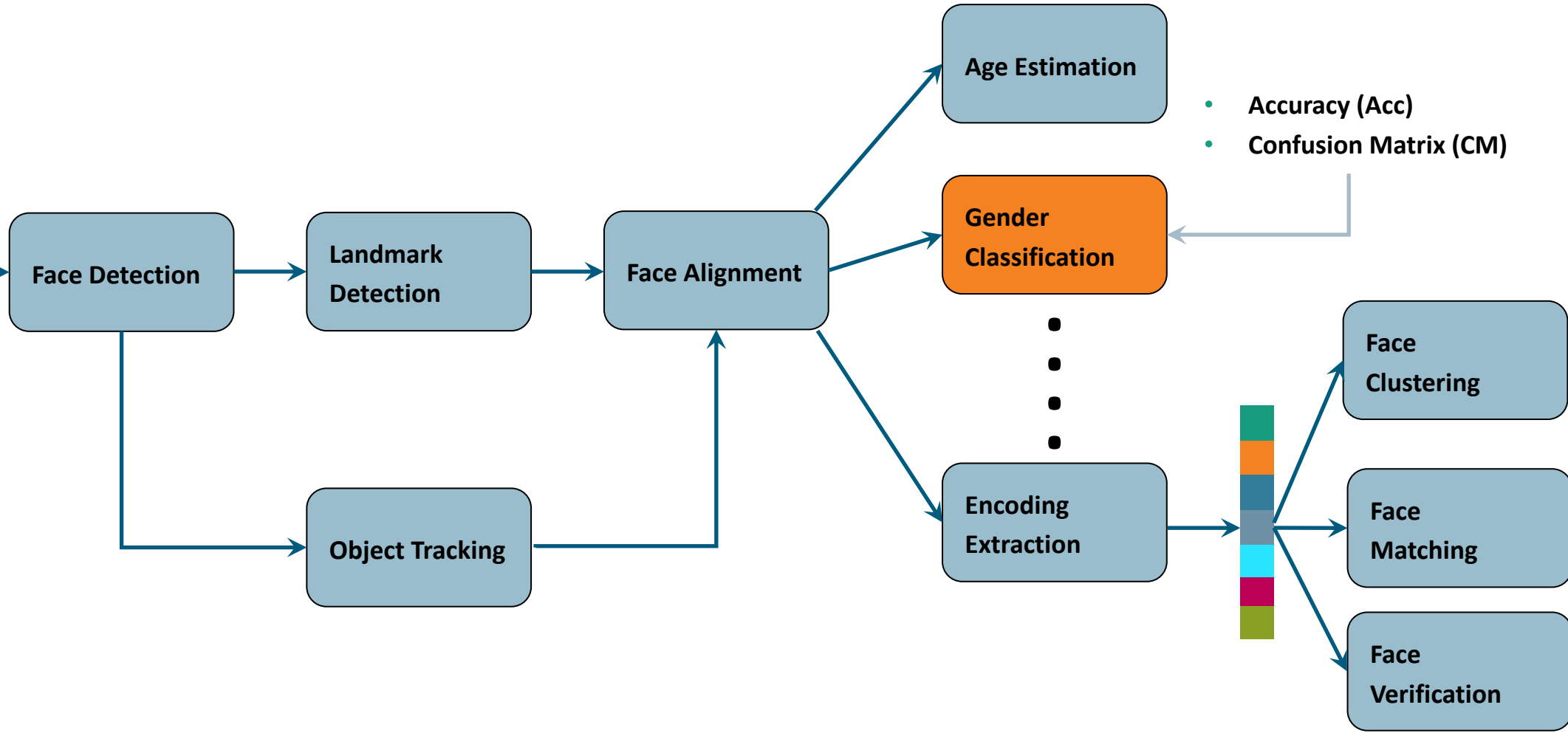
© Ellen DeGeneres/Twitter, Keystone



Evaluation of Face Analysis Algorithms

Performance Metrics

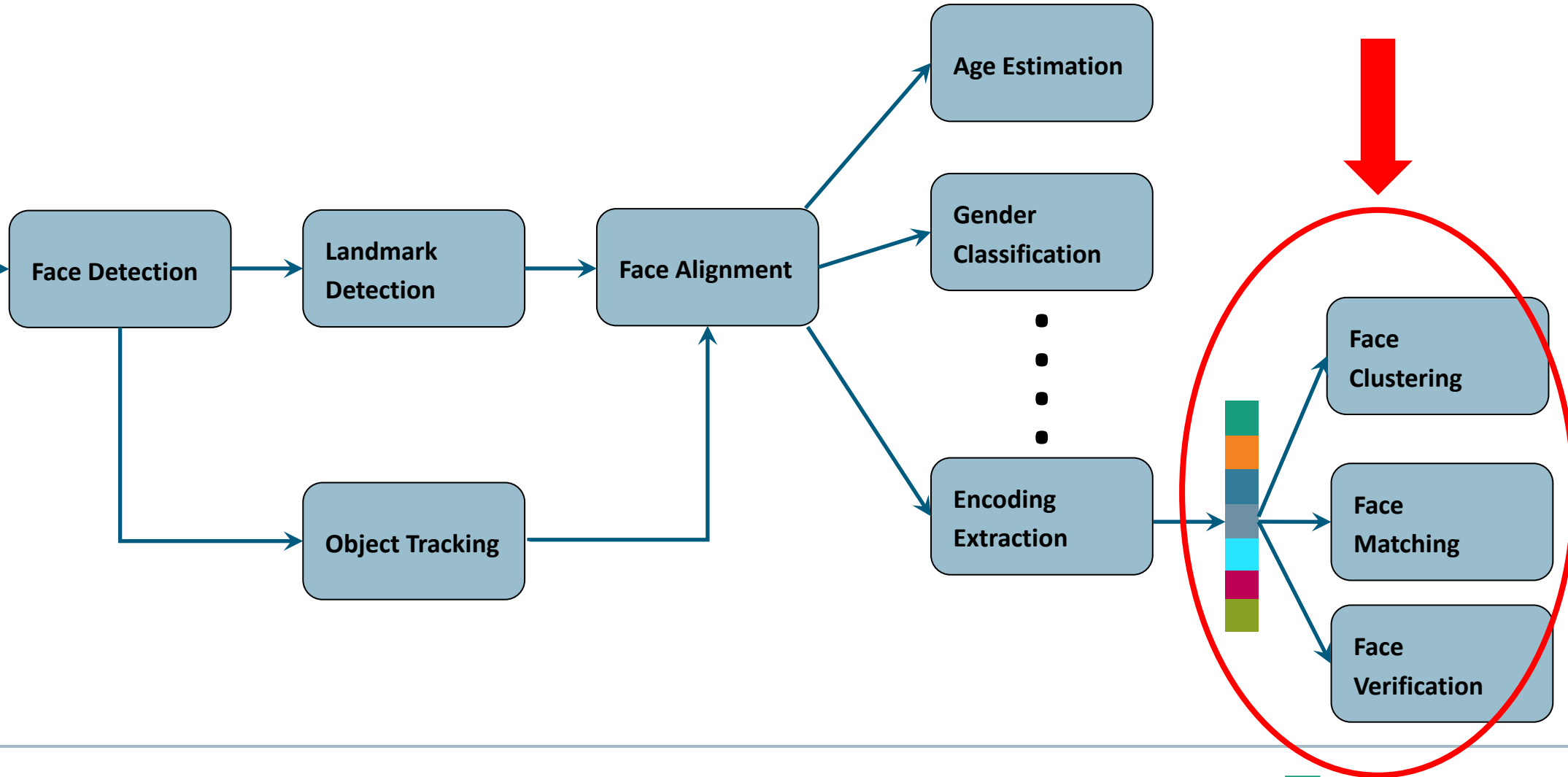
© Ellen DeGeneres/Twitter, Keystone



Evaluation of Face Analysis Algorithms

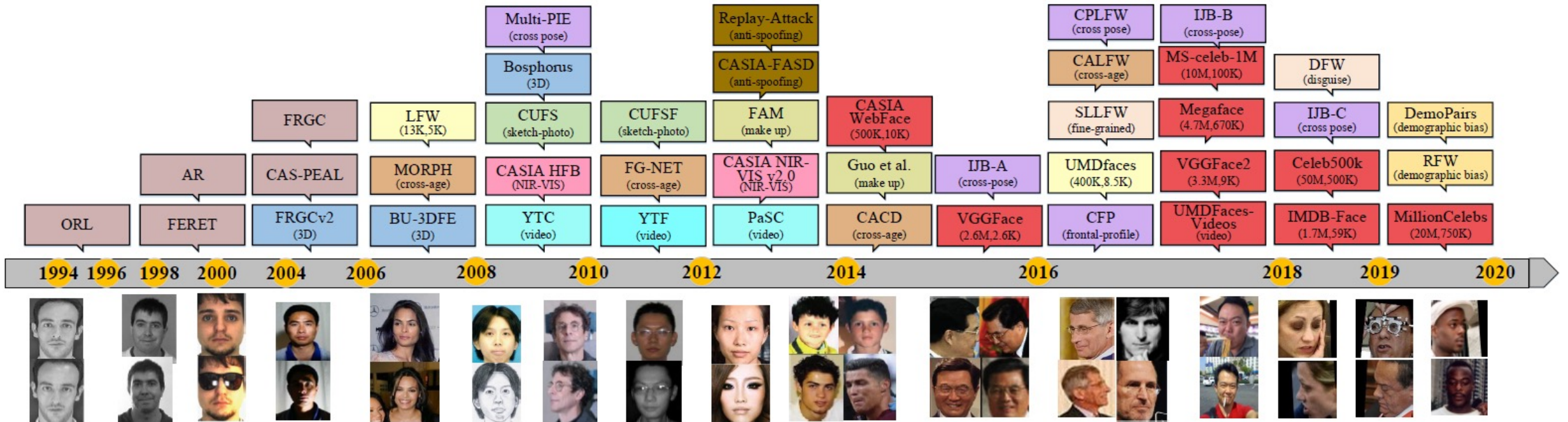
Performance Metrics

© Ellen DeGeneres/Twitter, Keystone



Evaluation of Face Recognition Algorithms

Large Scale Training Datasets

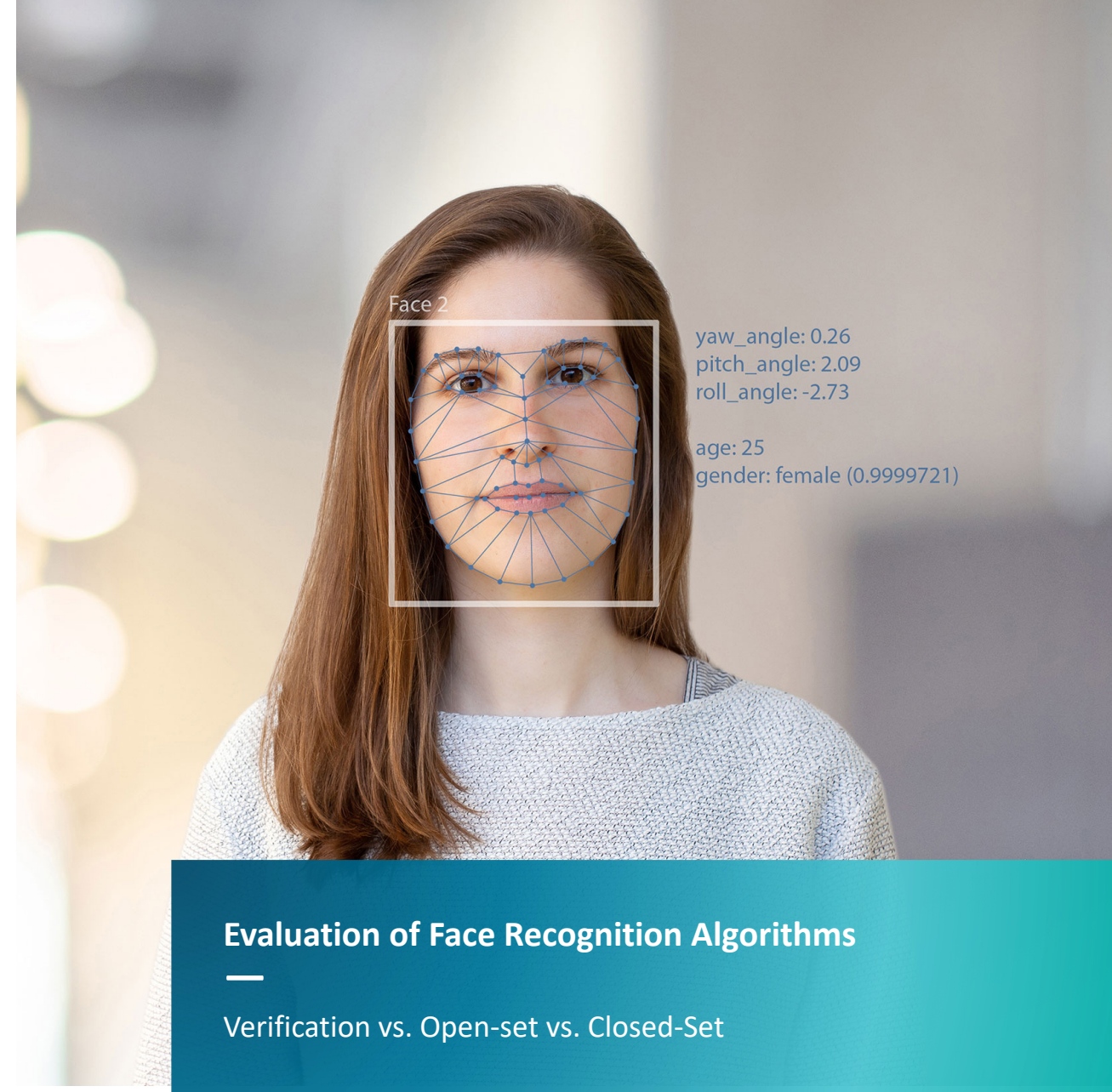


Source: M. Wang, W. Deng, „Deep Face Recognition: A Survey“, ArXiv preprint, arXiv:1804.06655v9

Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics

- **Gallery Set:**
 - Set of images of known subjects initially enrolled in the system
- **Probe Set:**
 - Set of images of known (or unknown) subjects different from the gallery set
- **Face Verification:**
 - Computes one-to-one similarity between the gallery and the probe image to determine if the two images are from the same subject
 - Relevant to access control systems and re-identification
- **Face Identification:**
 - Computes one-to-many similarity to determine the specific identity
 - **Closed-Set Identification:**
 - Person in probe image is **known** to the system
 - **Open-Set Identification:**
 - Person in probe image is **known** or **unknown** to the system



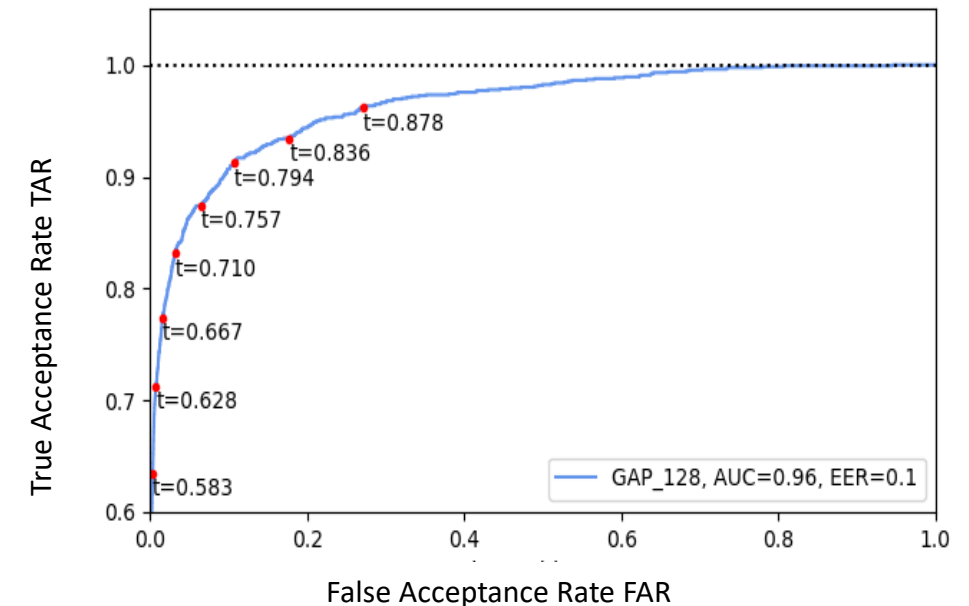
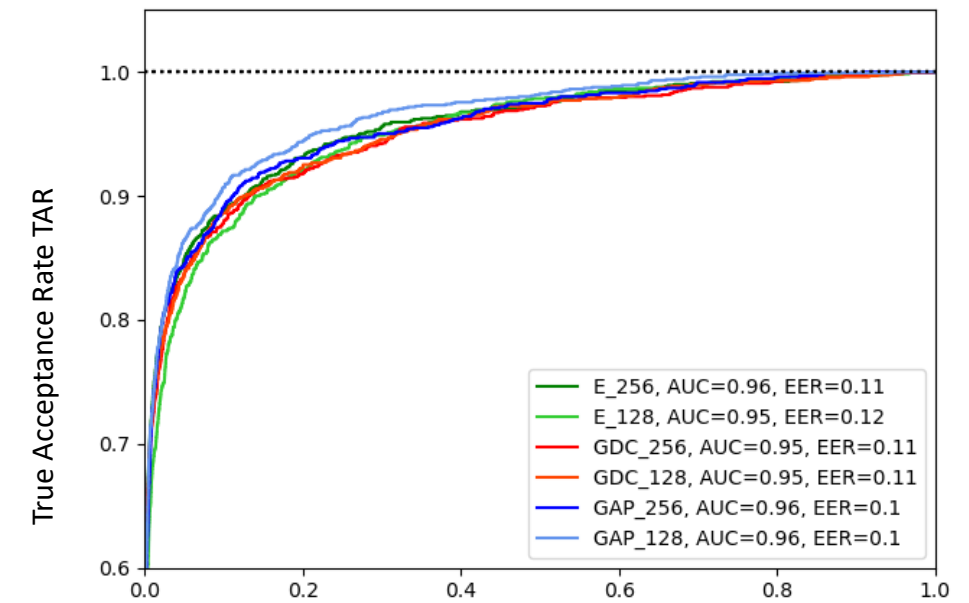
Evaluation of Face Recognition Algorithms

Verification vs. Open-set vs. Closed-Set

Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics (Verification)

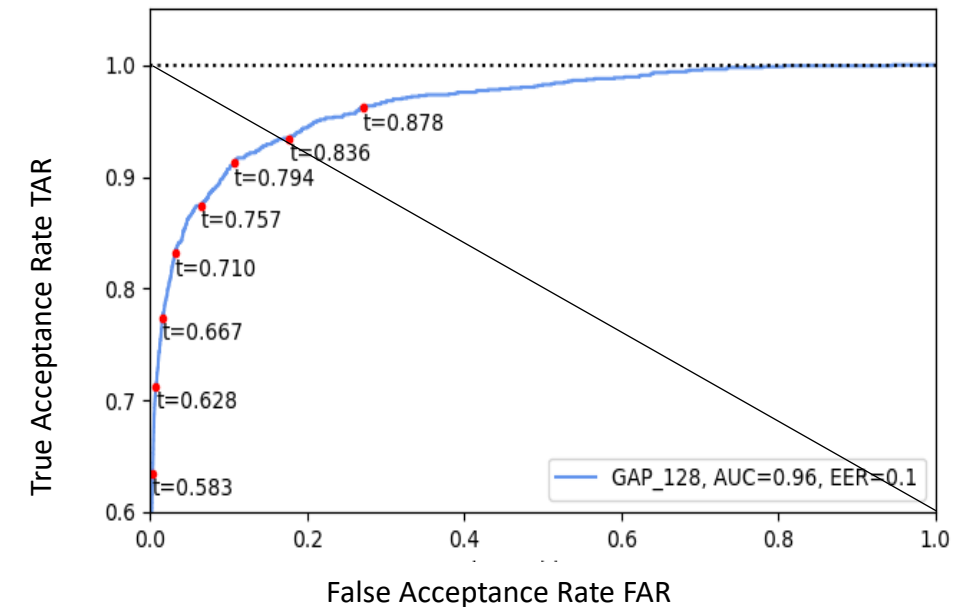
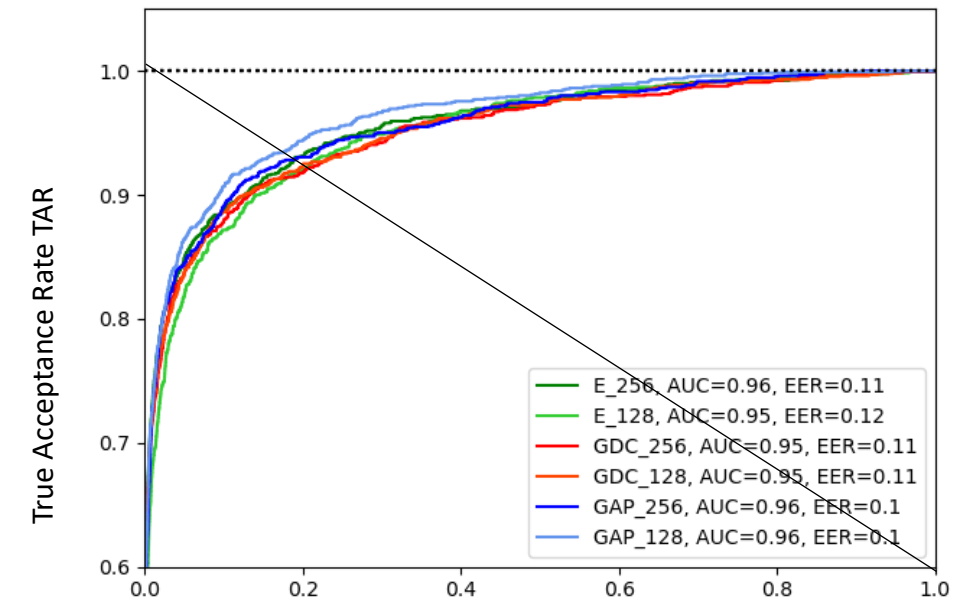
- **Face Verification:**
 - Computes one-to-one similarity between the gallery and the probe image to determine if the two images are from the same object
 - Relevant to access control systems and re-identification
- **Receiver Operating Characteristic (ROC) curve**
 - True Acceptance Rate (TAR) vs. False Acceptance Rate (FAR) at different thresholds
 - TAR: fraction of genuine individuals that exceed acceptance threshold
 - FAR: fraction of impostor individuals that exceed acceptance threshold
- **Comparison of systems**
 - Area Under Curve (AUC)
 - Equal Error Rate (EER)
 - PaSC: $TAR@10^{-2}FAR$
 - IJB-A: $TAR@10^{-3}FAR$
 - Megaface: $TAR@10^{-6}FAR$
 - MS-celeb-1M: $TAR@10^{-9}FAR$



Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics (Verification)

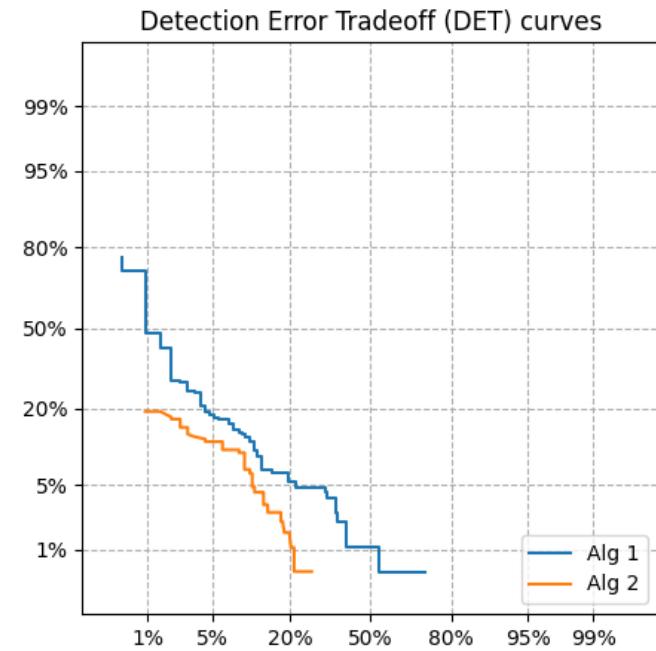
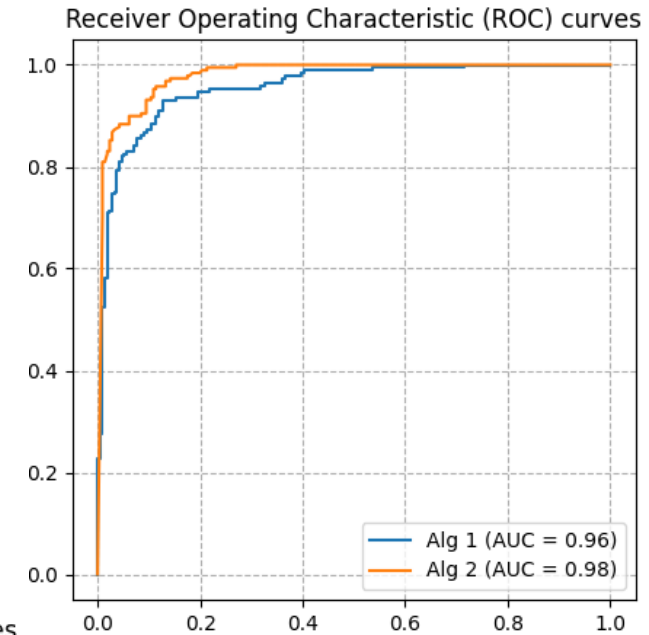
- **Face Verification:**
 - Computes one-to-one similarity between the gallery and the probe image to determine if the two images are from the same object
 - Relevant to access control systems and re-identification
- **Receiver Operating Characteristic (ROC) curve**
 - True Acceptance Rate (TAR) vs. False Acceptance Rate (FAR) at different thresholds
 - TAR: fraction of genuine individuals that exceed acceptance threshold
 - FAR: fraction of impostor individuals that exceed acceptance threshold
- **Comparison of systems**
 - Area Under Curve (AUC)
 - Equal Error Rate (EER)
 - PaSC: $TAR@10^{-2}FAR$
 - IJB-A: $TAR@10^{-3}FAR$
 - Megaface: $TAR@10^{-6}FAR$
 - MS-celeb-1M: $TAR@10^{-9}FAR$



Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics (Verification)

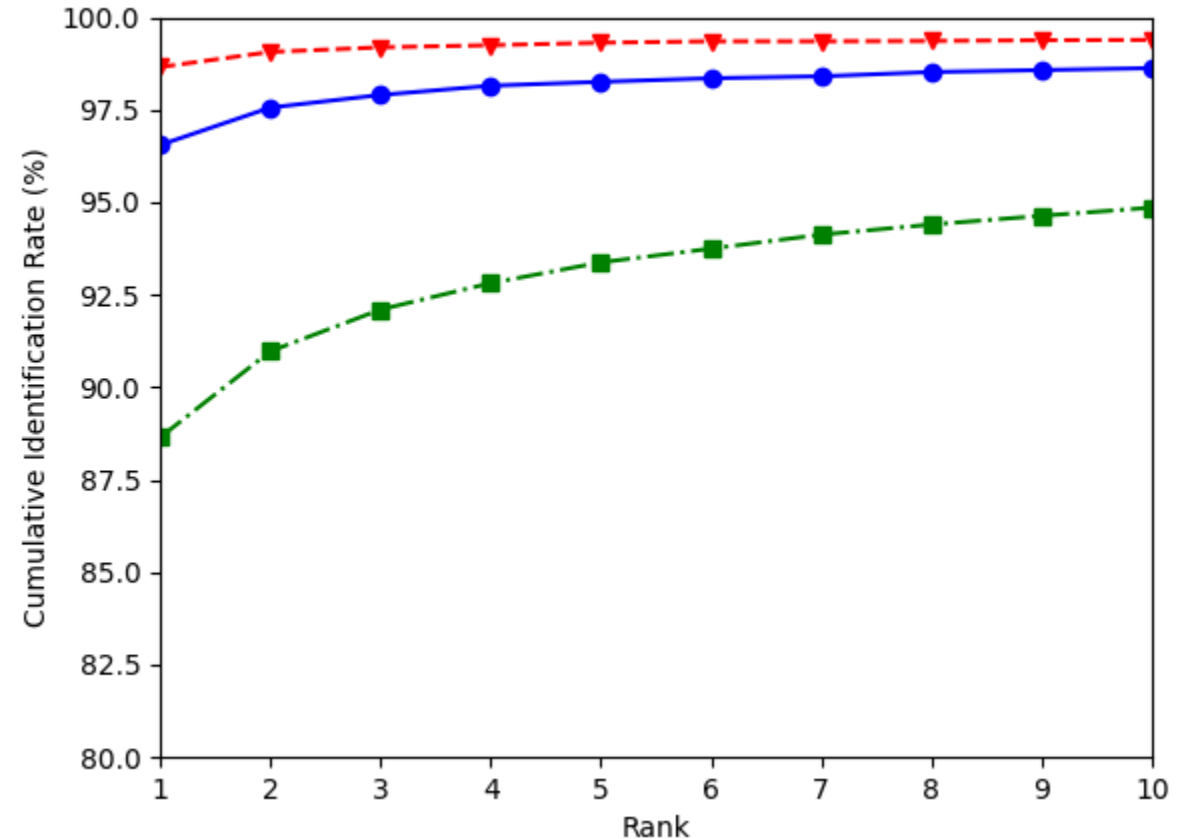
- **Face Verification:**
 - Computes one-to-one similarity between the gallery and the probe image to determine if the two images are from the same object
 - Relevant to access control systems and re-identification
- **Detection (Decision) Error Trade-Off Curve (DET)**
 - False Rejection Rate (FRR) vs. False Acceptance Rate (FAR) at different thresholds
 - FRR : fraction of genuine individuals classified as unknown
 - FAR: fraction of impostor individuals that exceed acceptance threshold
- Might be easier to visually compare different algorithms
- The user can deduce directly from the DET-curve plot at which rate FRR will improve when willing to accept an increase in FAR (or vice-versa)



Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics (Identification – Closed Set)

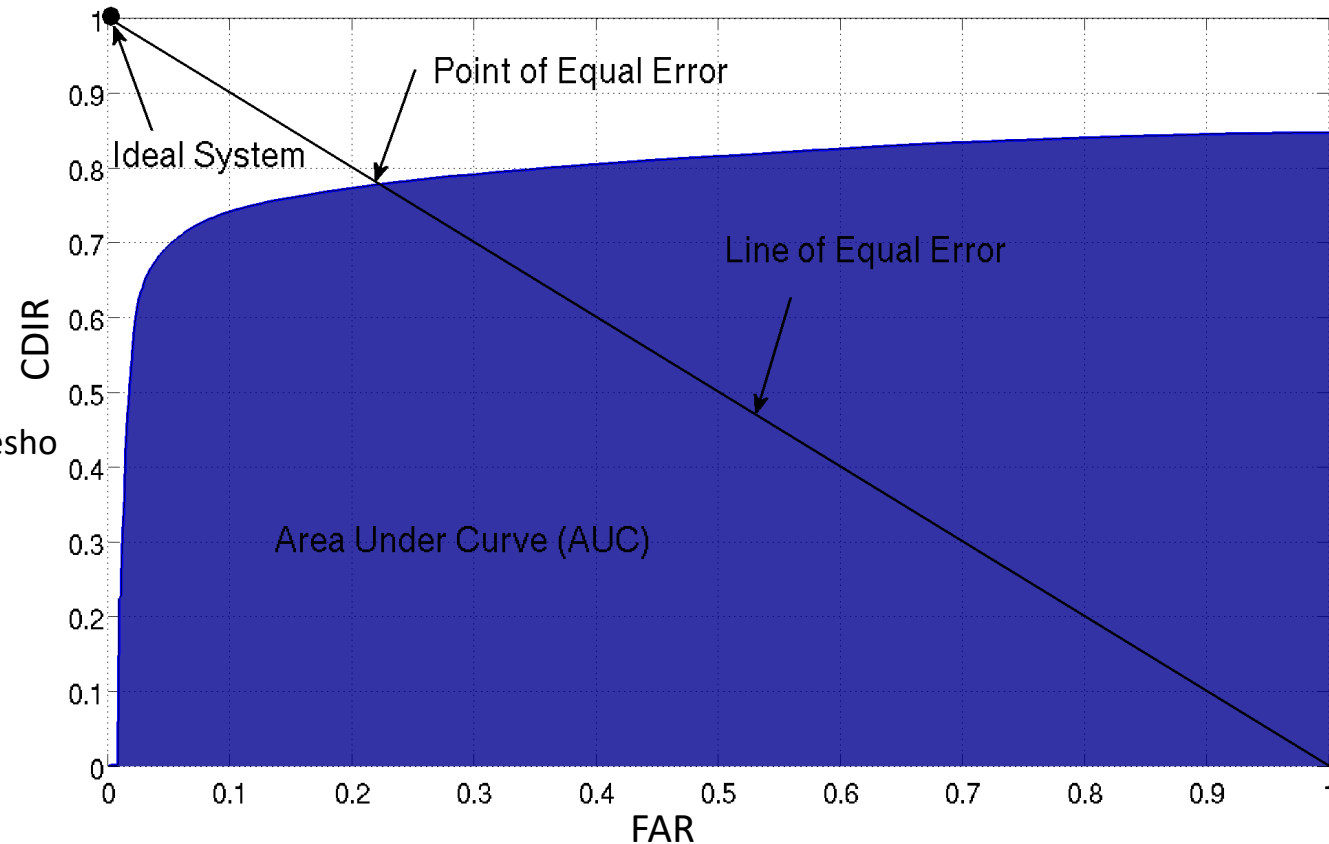
- **Face Identification: Closed-Set Identification:**
 - Computes *one-to-many* similarity to determine the specific identity
 - Person in probe image is *known* to the system
- **Cumulative Match Characteristic (CMC)**
 - Reports the percentage of probes identified within a given rank
 - Cumulative Identification Rate vs. Rank
 - Most frequent metric: Rank-1 Accuracy



Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics (Identification – Open Set)

- **Face Identification: Open-Set Identification:**
 - Computes *one-to-many* similarity to determine the specific identity
 - Person in probe image is *known* or *unknown* to the system
- **Receiver Operating Characteristic (ROC)**
 - Correct Acceptance and Identification Rate (CDIR) vs. False Acceptance Rate (FAR) at different thresholds
 - CDIR: fraction of genuine individuals that exceed acceptance threshold and are correctly identified
 - FAR: fraction of impostor individuals that exceed acceptance threshold
- **Comparison of Systems**
 - Area under Curve
 - Equal Error Rate



Evaluation of Face Recognition

Evaluation Tasks and Performance Metrics (Identification – Open Set)

- **Open-Set Identification:**

- Computes *one-to-many* similarity to determine the specific identity
- Person in probe image is *known* or *unknown* to the system

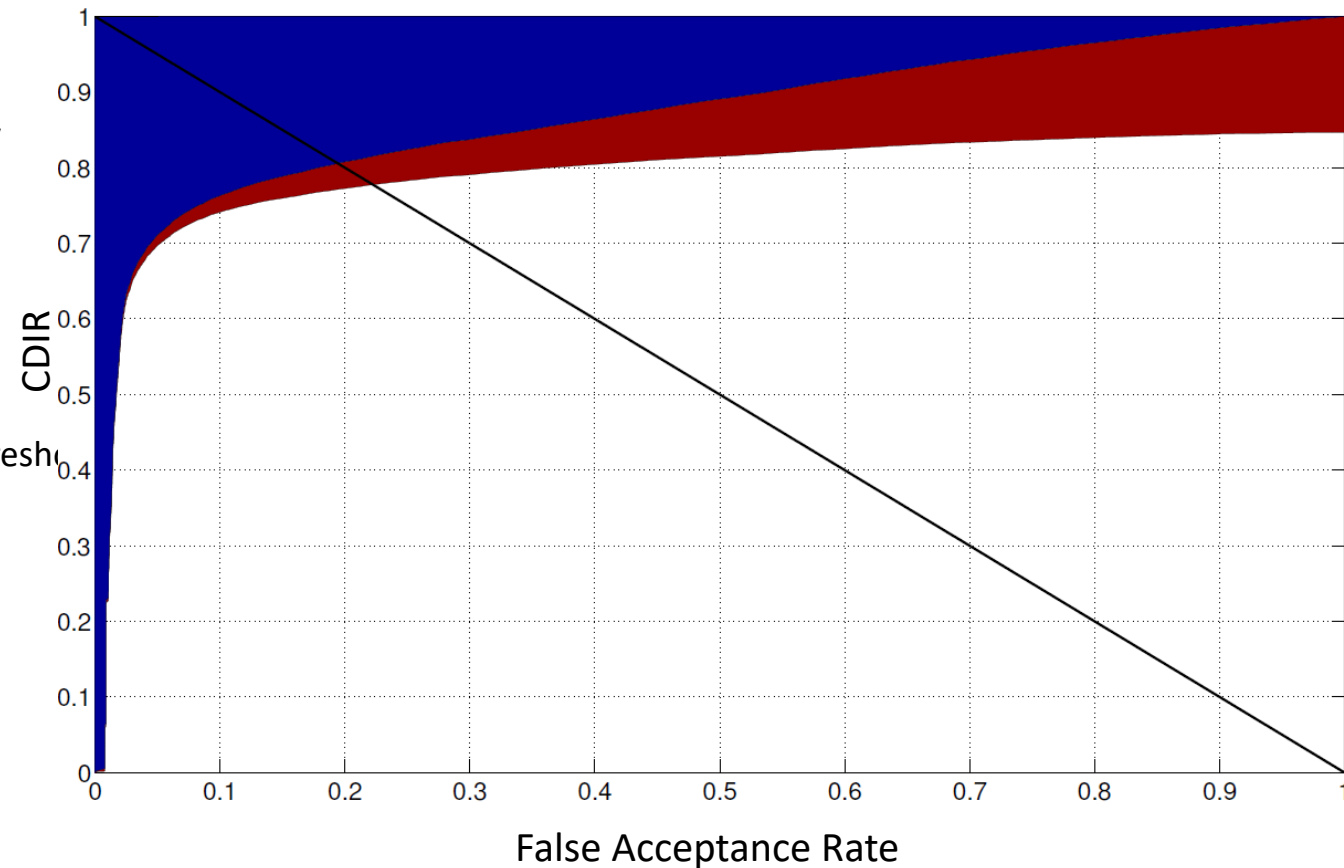
- **Receiver Operating Characteristic (ROC)**

- Correct Acceptance and Identification Rate (CDIR) vs. False Acceptance Rate (FAR) at different thresholds
- CDIR: fraction of genuine individuals that exceed acceptance threshold and are correctly identified
- FAR: fraction of impostor individuals that exceed acceptance
- Threshold

- **Comparison of Systems**

- Area under Curve
- Equal Error Rate

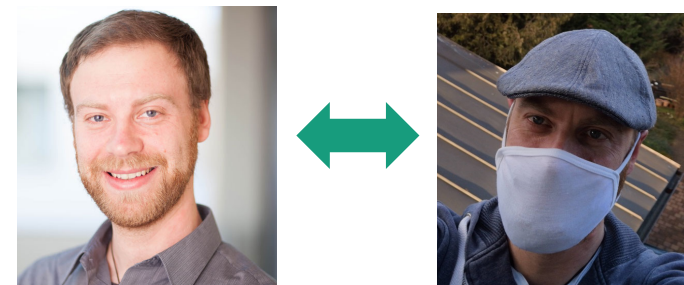
- **Red: Error from False Classification**
- **Blue: Error from False Rejection**



Evaluation of Face Recognition

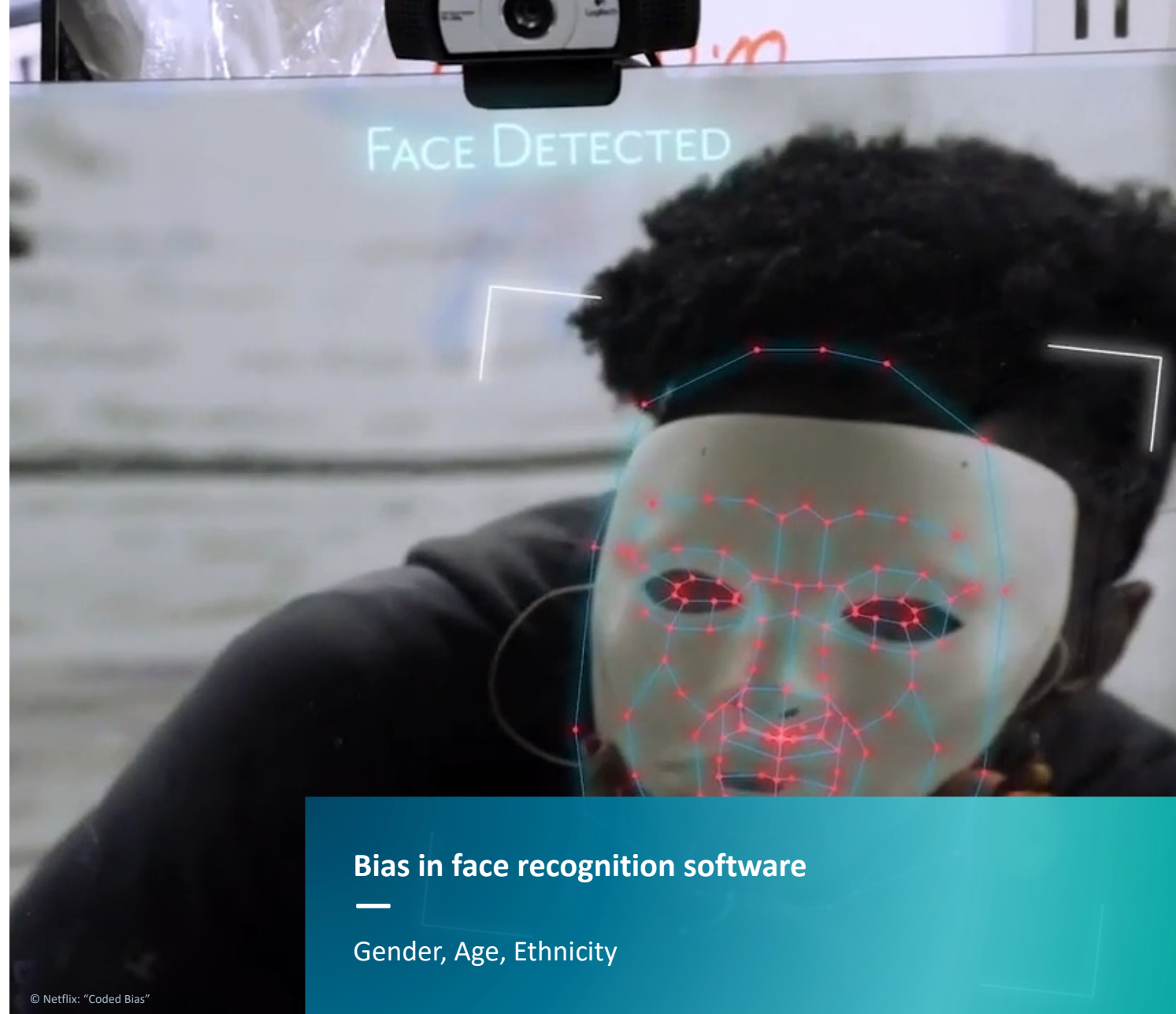
Cross-factor face recognition

- **Cross-Pose Face Recognition**
 - Severe decrease in accuracy from full-frontal to profile face verification
 - Face frontalization in image and feature space
 - Pose Invariant Models (PIM) to learn pose invariant features
- **Cross-Age Face Recognition**
 - Extremely challenging due to changes in facial appearance over time
 - Synthesize desired image with target age using GANs
 - Methods that decompose ageing and identity components and thus extract age invariant representations
- **Cross-Makeup Face Recognition**
 - Increases intra-class variations
 - Synthesize non-makeup versions from makeup images
- **Masked Face Recognition**
 - Face Recognition for masked faces
 - Algorithms trained on eye region only



Evaluation of Face Recognition Biases in face recognition software

- **Gender bias**
 - Male
 - Female
 - Diverse
- **Age bias**
 - Children
 - Juvenile
 - Adult
 - Elderly people
- **Ethnicity bias**
 - Caucasian
 - Asian
 - Indian
 - African
 - ...



How good is my AI System?

Summary

- **> 10 years of experience in the field of face analysis**
 - Face recognition for media applications
 - Automatic annotation of images and videos
 - Unsupervised clustering of similar looking faces
 - TV program study
 - Face recognition on edge devices
 - IDCheck
 - Maskcognizer
 - Special use cases
 - Face recognition for animals
 - Evaluation of face recognition algorithms as a service
- **Evaluation**
 - The whole face recognition pipeline matters!
 - What use-cases are you interested in?
 - Use the correct evaluation metric for your use case!
 - Interpret the results correctly, and look at the errors in particular!
 - What biases and cross-matching scenarios can occur?



How good is my AI system?

Lessons learned

Contact

Dr.-Ing Alexander Loos
Fraunhofer IDMT
Audio-visual Systems

Tel. +49 3677 467-285
Fax +49 3677 467-467
alexander.loos@idmt.fraunhofer.de

Fraunhofer IDMT
Ehrenbergstraße 31
98693 Ilmenau
www.idmt.fraunhofer.de



Thank you for your
attention!
