# Remote Face Recognition

Rama Chellappa

Johns Hopkins University

# Face recognition from boats and shore: Objectives (2008-2013)

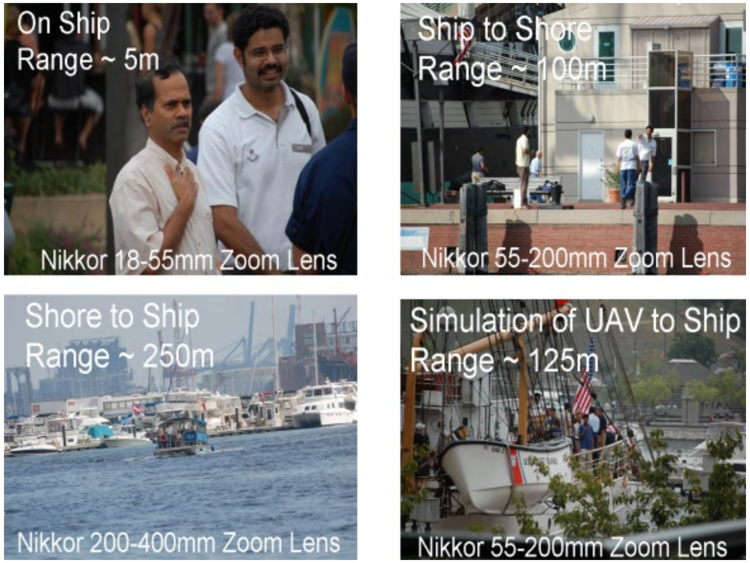
- To understand and mitigate the degradations in the acquisition of biometric signatures in the maritime domain.
- To develop robust algorithms for remote face recognition in the maritime domain.
- Evaluate the effectiveness of remote biometrics algorithms.
- What is the significance and potential scientific impact of the project?
  - Extends the range and operating conditions of object recognition research -- a fundamental goal of computer vision research
- Participants
  - Belhumeur, Boult, Davis, Duraiswami, Jacobs, Kriegman, and Nayar

# Data collection

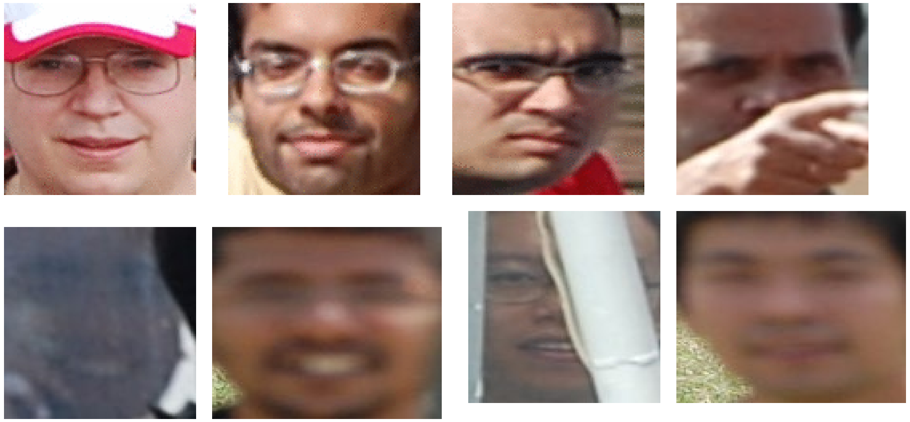
- Baltimore inner harbor
  - Images of subjects in a boat at 25-400m from the camera
  - Maritime conditions
    - Detected faces have blur, occlusion, severe lighting variations, pose and expression
    - Over 2000 frontal (or close to frontal images)
    - Hundreds of video sequences have been collected.
    - Each face was labeled based on identity, pose, illumination, blur and occlusion.
    - Some of the artifacts are unique to maritime conditions.
- Xfinity Center, UMD
  - During winter months

# Some examples of remote data

Baltimore Inner Harbor data



Xfinity Center, UMD data



Atmospheric effects (fog, mist, rain, etc.)
Blur
Jitter due to ship motion
Low-resolution
Illumination, pose variations
Occlusion
Presence of others
Collecting large data sets

# Face recognition in 2008 and 2013

- 2008
  - Frontal, well-illuminated, high-resolution, sharp and occlusion-free face recognition problem was addressed.
  - Face data set typically consisted of a few thousands of faces.
  - Constrained data sets (PIE, FERET,..) were used for evaluation,
  - Video-based face recognition was barely discussed.
  - Face recognition/verification seen as a standalone problem.
- 2013
  - Non-frontal, not so well illuminated, blurry, low-resolution and limited occlusion face recognition problem is being addressed.
  - No one is impressed with near 100% recognition on frontal, well-illuminated and high-resolution face data sets.
  - Millions of faces are included in the data set. Challenging face data sets (LFW, MBGC, UMD MURI) are used for evaluation.
  - Many approaches to video-based recognition are being considered.
  - Face recognition/verification integrated into surveillance and indexing applications

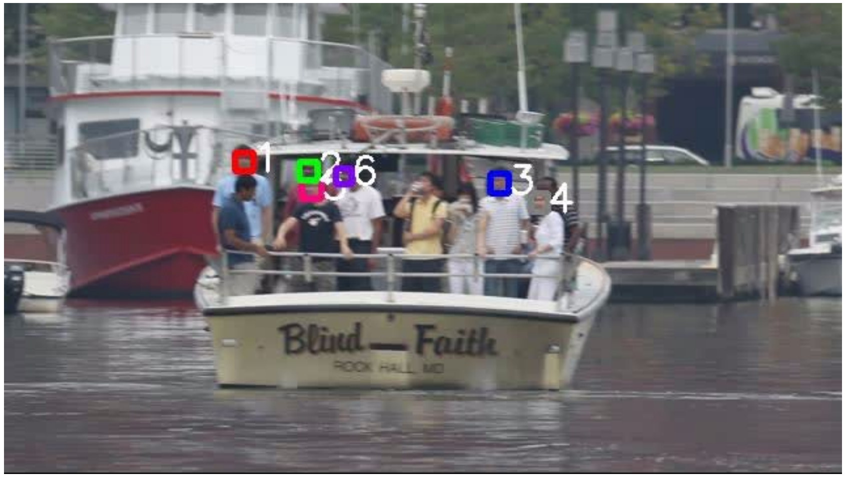
# Preprocessing – Face detection

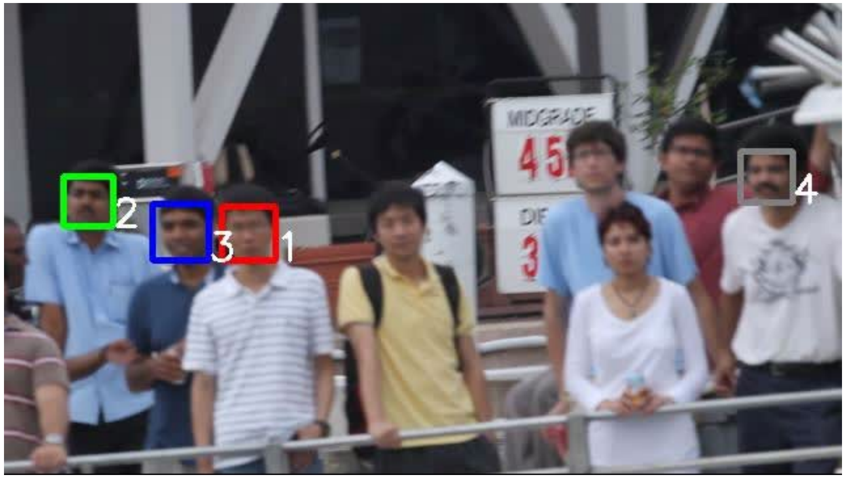
- Before MURI
  - Viola Jones face detector
  - Video stabilization and face tracking using particle filters
- MURI
  - PLS method
  - Transitioned to DARPA VMR program
- Video stabilization and face tracking (UMD)
  - Association of frame-based detections using conditional random field models

**Examples of face detection in shore-to-ship and simulated UAV-to-ship scenarios**

# Acquisition of faces in motion

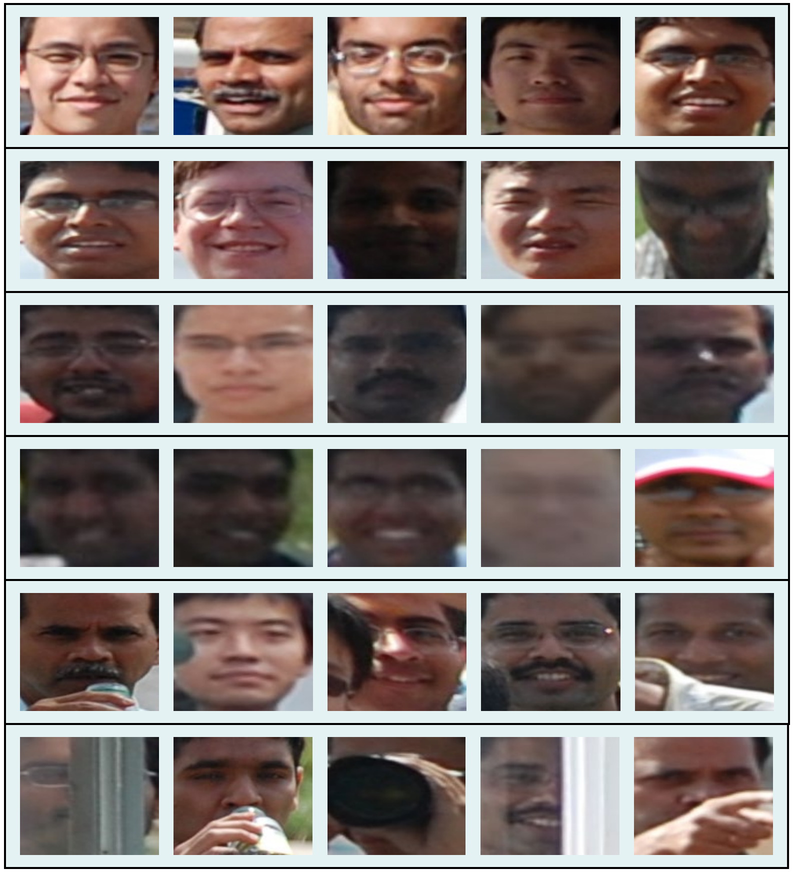
- We developed a Bayesian, scene-adaptive approach that is effective for scenarios involving sensor motion (ship to ship, ship to shore etc).
- Prior models tuned to scenes
- Online estimation of conditional random field models

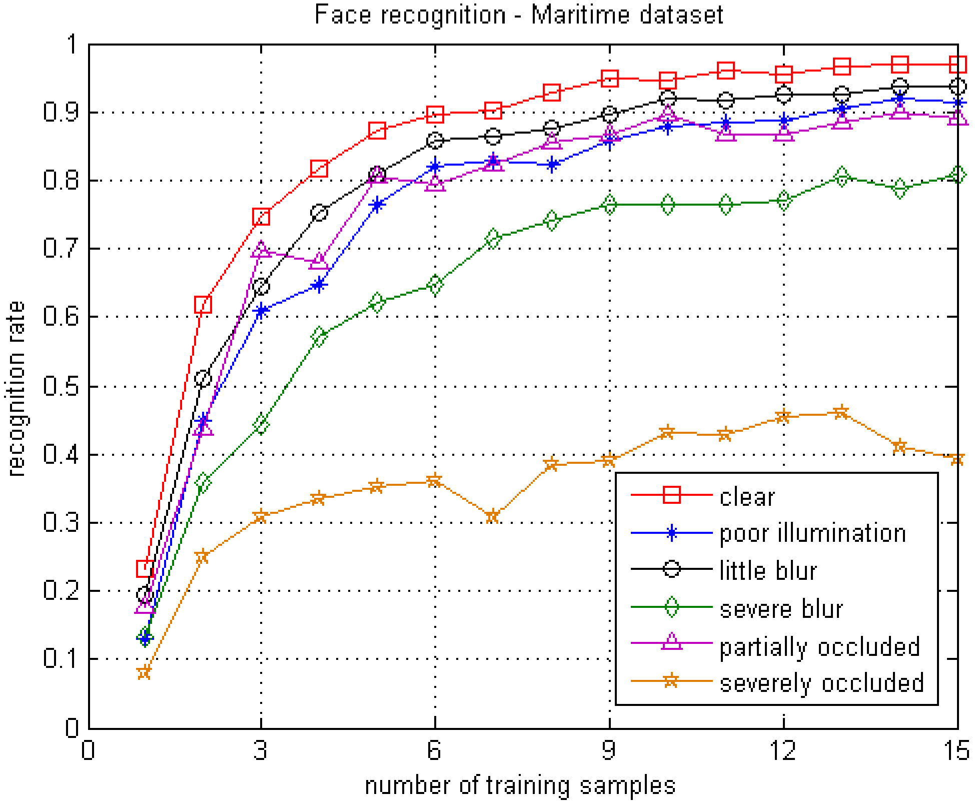




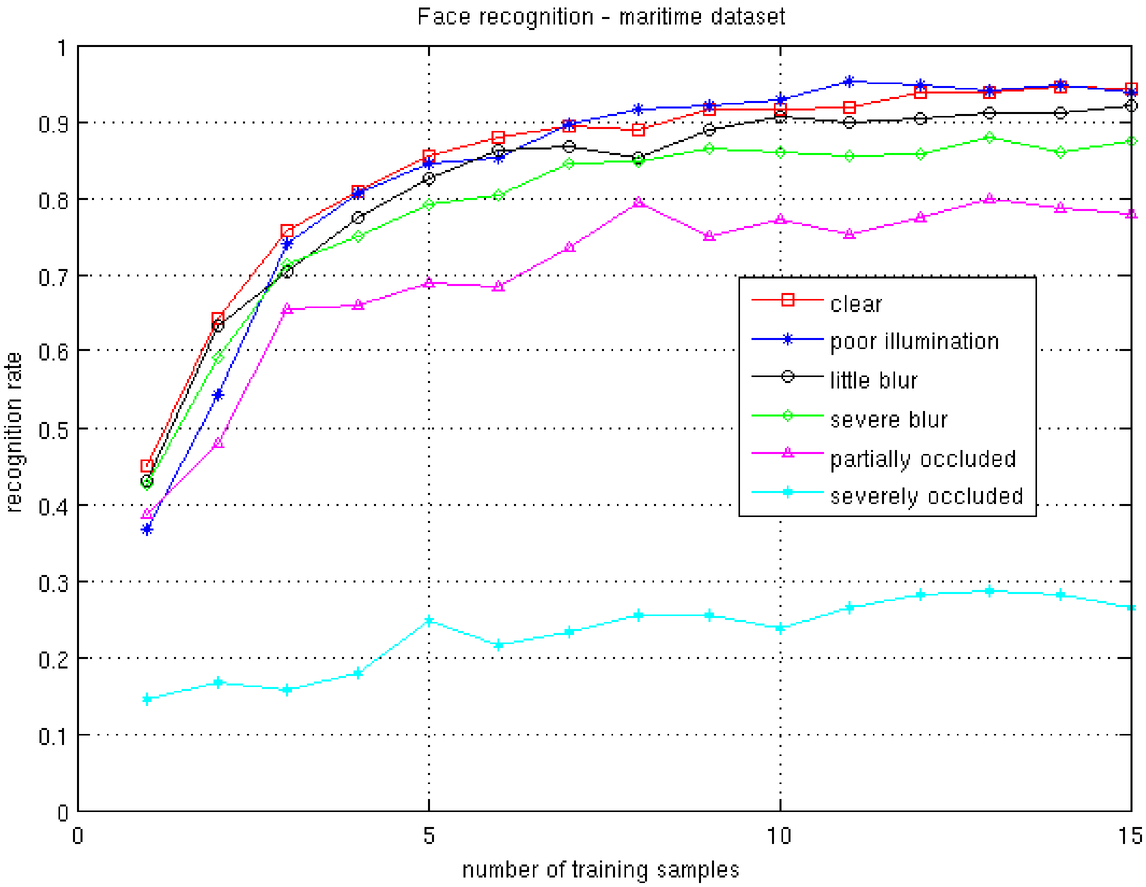
# Face recognition on maritime data - partial least squares method

Feature dimensionality reduction 50,000 to 20.



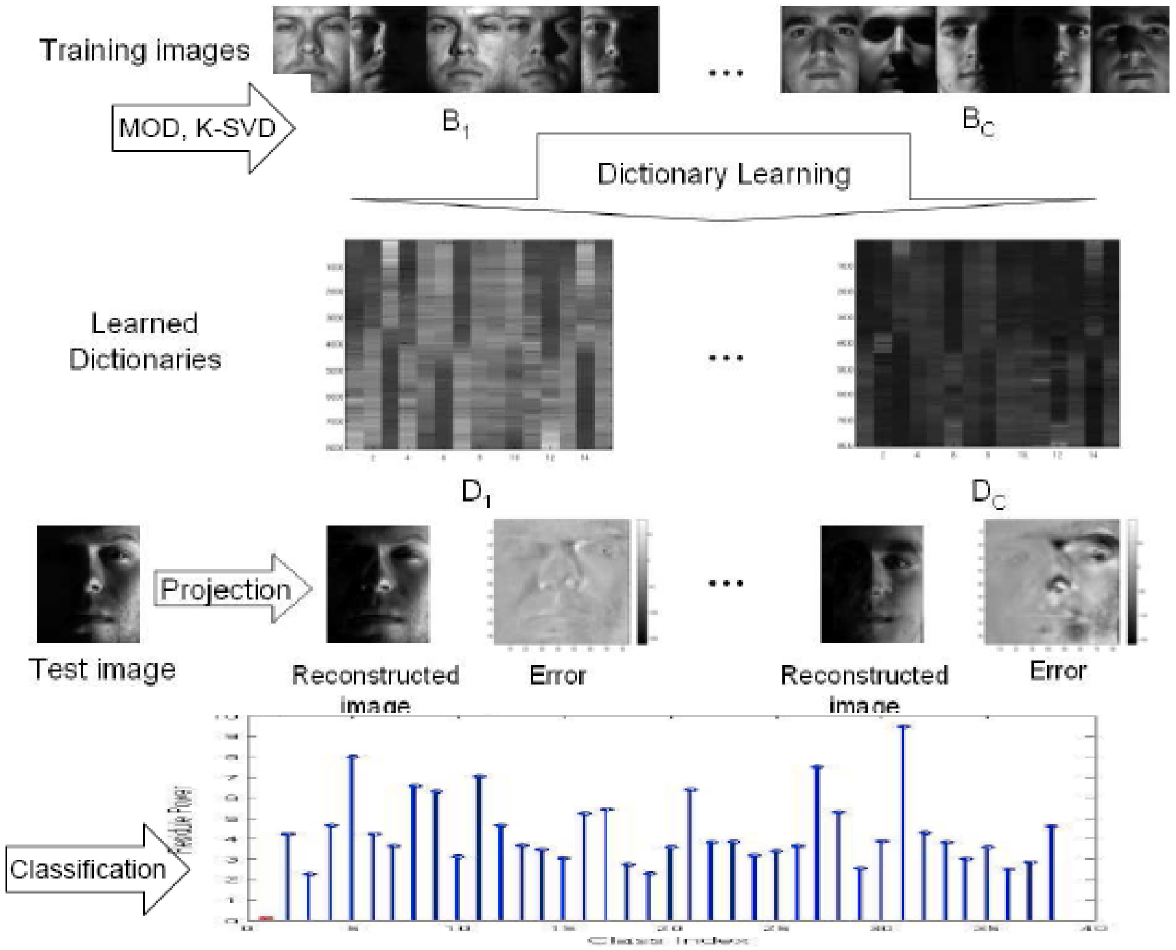


# Face recognition on maritime intensity data - SVM



Kernel PCA
Regularized LDA
SVM

# Dictionary-based face recognition



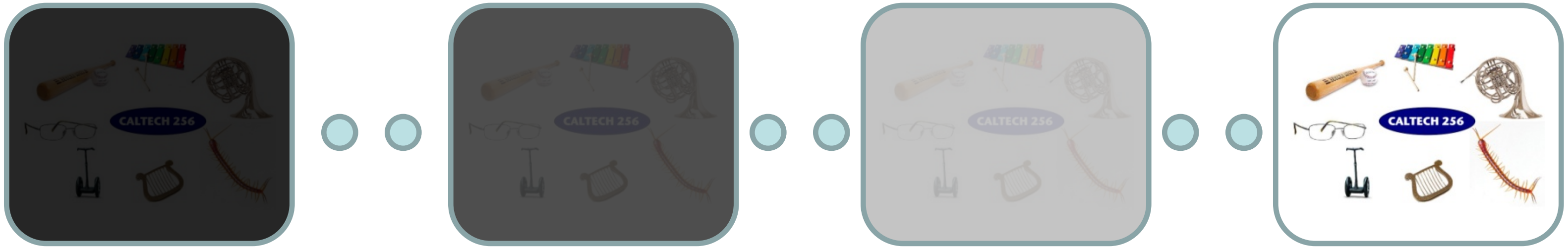
Classification result among several classes

How to learn Dictionaries?

K-SVD

M. Aharon, et al. 2006

# Re-Identification of faces



Domain 1
(labeled
Baltimore)

Intermediate domains
(Incremental learning)

Domain 2
(unlabelled
UMD
Comcast
center)

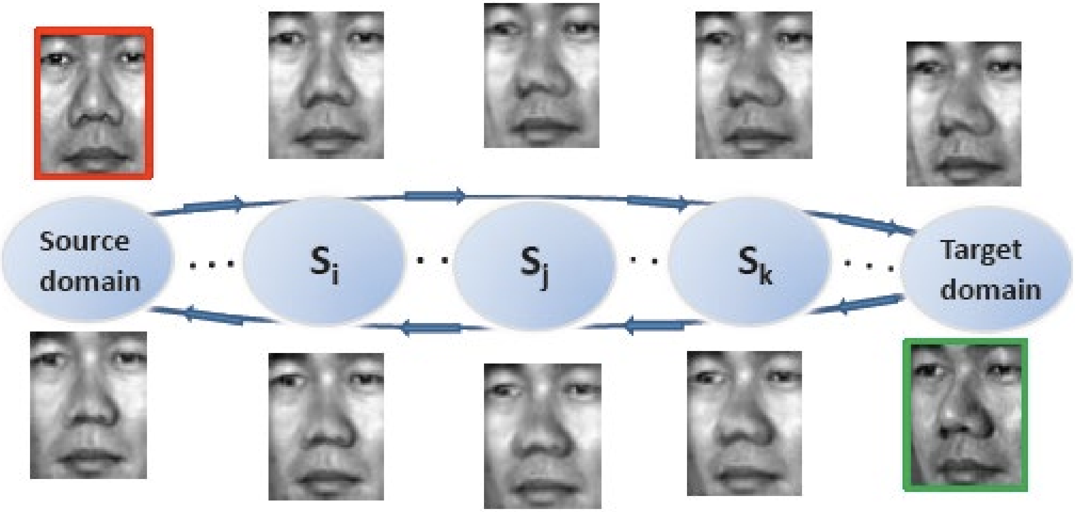
## *Questions*

❖ How to obtain meaningful intermediate domains?

❖ How to characterize incremental domain shift information to perform recognition?

❖Variations due to pose, illumination, background,..

# Domain adaptation via dictionaries

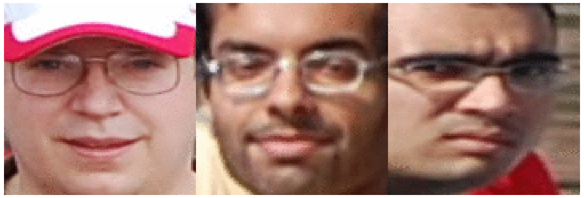


The top half of the figure shows some intermediate images synthesized from a given source image of frontal view (in red box). The bottom half shows the intermediate images generated from a given target image of side view (in green box).

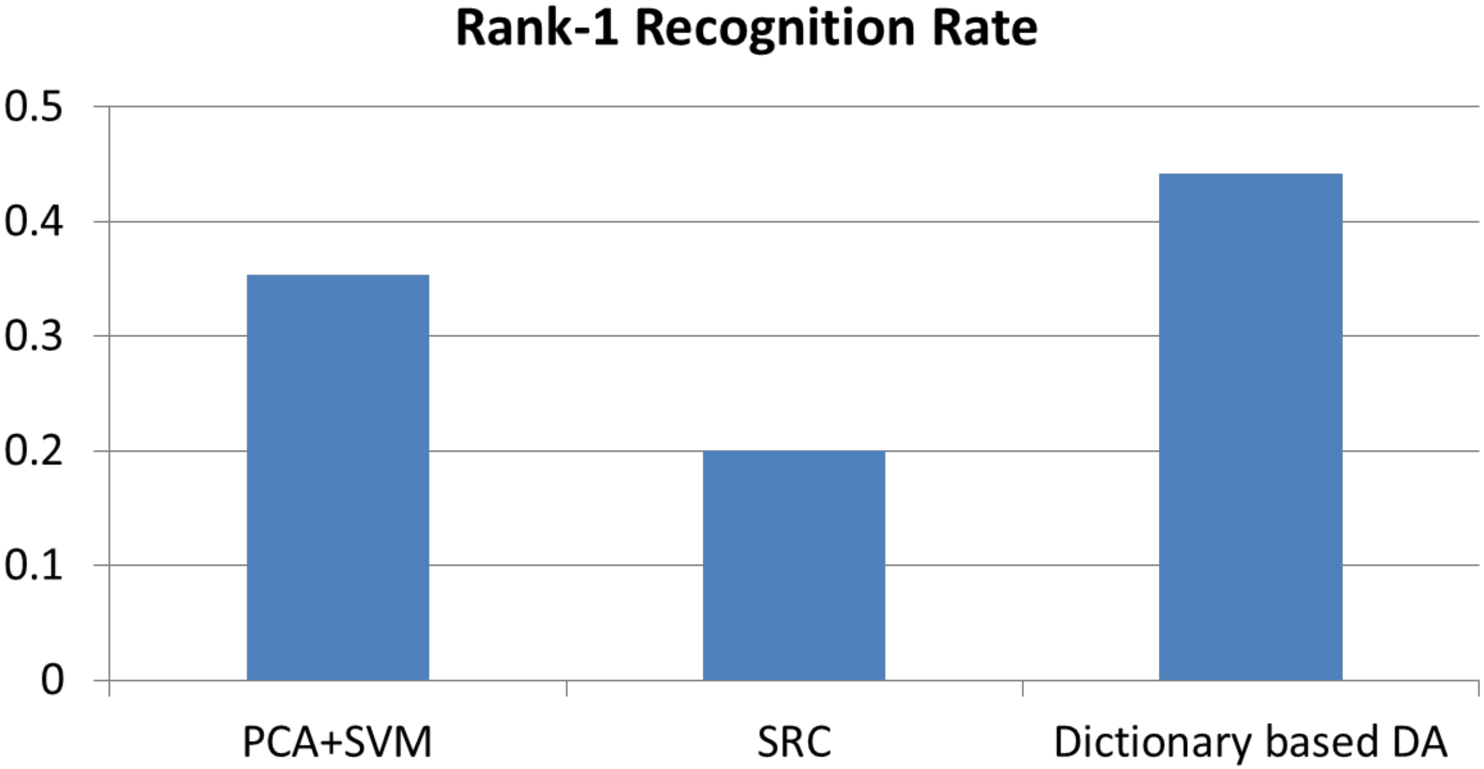
- Assume there exist K intermediate domains $\{S_k\}_{k=1}^{K}$ which smoothly bridge the information gap between the source and target domain. A domain dependent dictionary $D_k$ is learned for each intermediate domain $S_k$.
- We learn the intermediate data to approximate the observations in the corresponding intermediate domains. The intermediate data is then utilized to build classifiers.

# Results

- 75 images from Baltimore dataset as gallery (source domain)



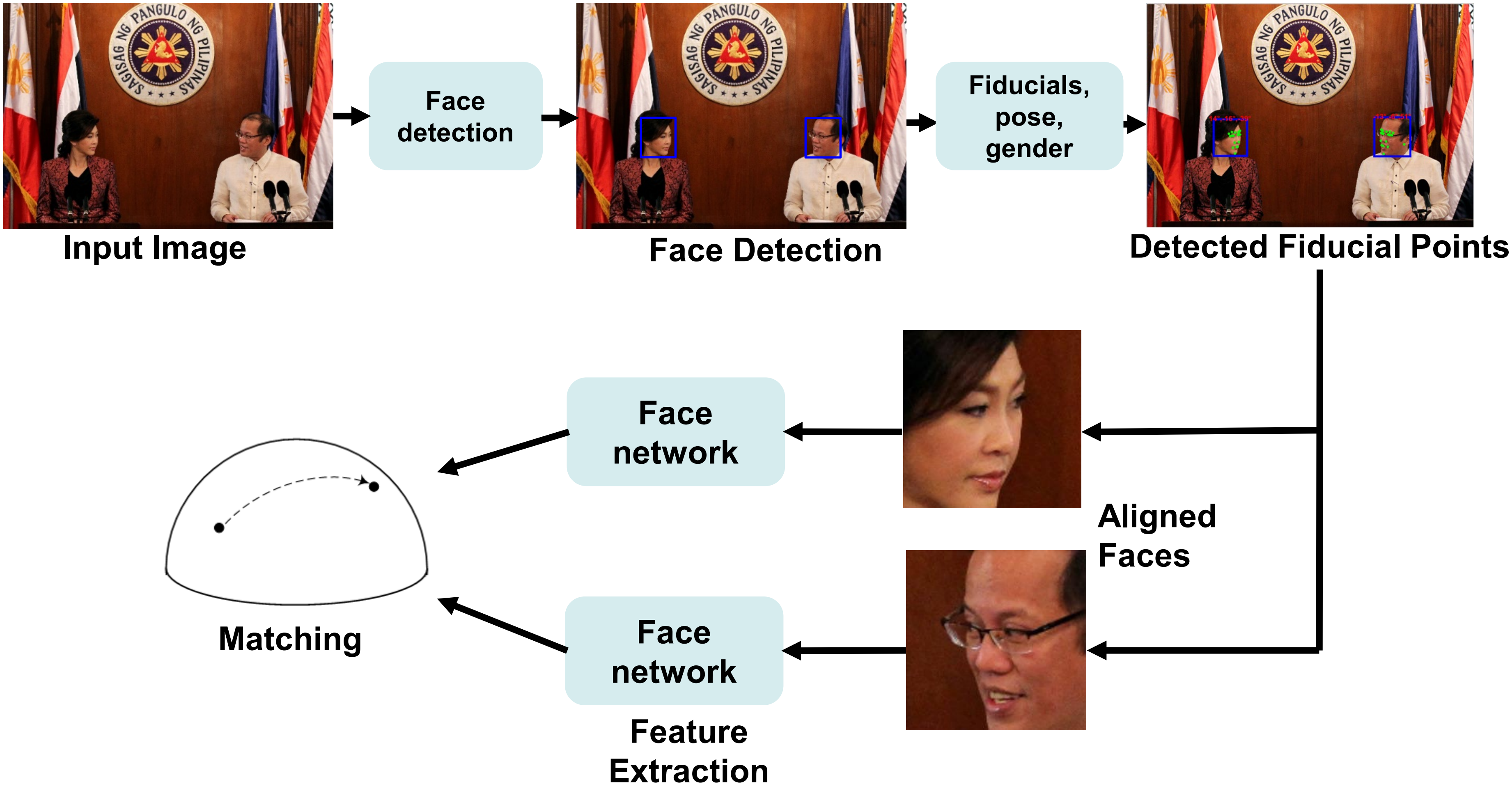
150 images from Comcast dataset as probe (target domain).



# Unconstrained face recognition

- 2014 – 2018, Supported by IARPA
- UMD (Lead) with CMU, Columbia, JHU, UB, UCCS, UTD.
- Multi-task learning in deep networks
  - Face and gender detection, pose and age estimation, fiducial extraction
- Network of networks
  - Fusion of short and tall networks
- Current template size is 384 floats (1536 bytes or 12288 bits)
  - Hashing reduces size to 3072 bits
- State-of-the art performance on face verification, search, clustering tasks using relatively small training data set.
- Implications to forensics (Collaborations with Jonathon Phillips, and Alice O'Toole) – Proc. National Academy of Sciences, May 28, 2018.
- 2019-2020, transition phase with Columbia, JHU and UT Dallas.

# An end-to-end system for unconstrained face verification



Input Image

Face detection

Face Detection

Fiducials, pose, gender

Detected Fiducial Points

Aligned Faces

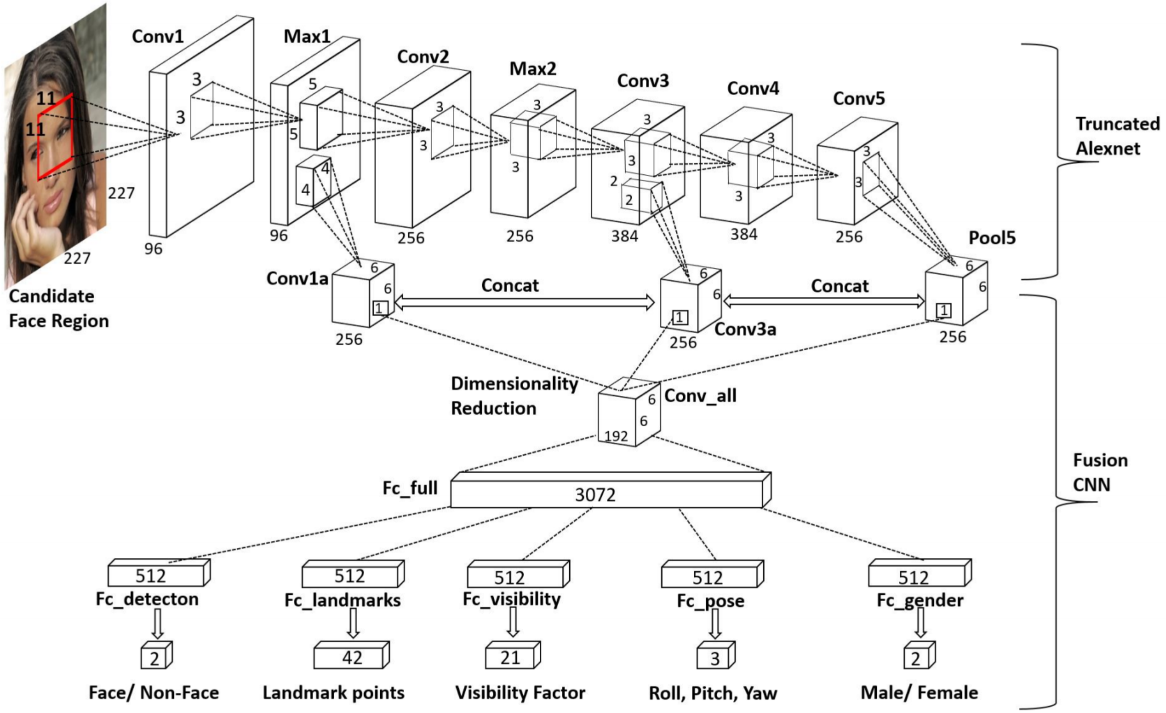
Face network

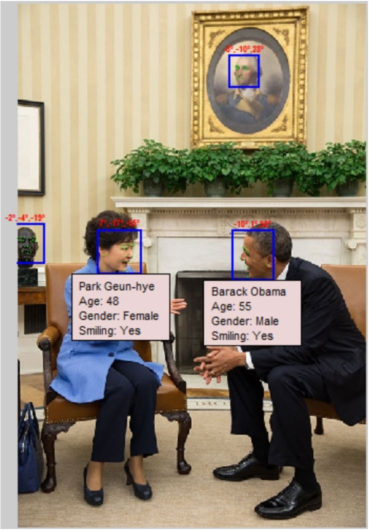
Face network

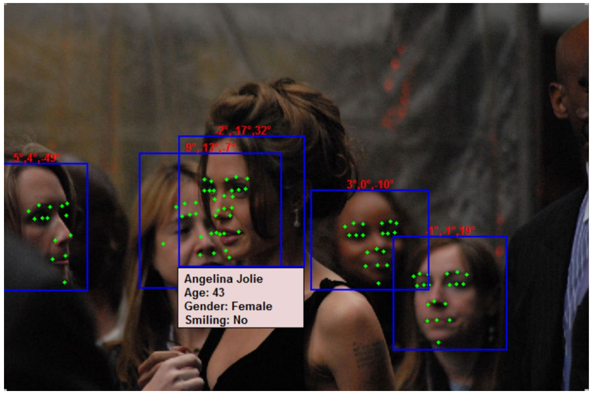
Feature Extraction

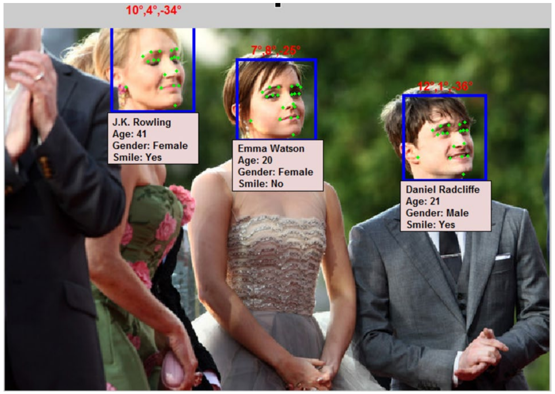
Matching

# Hyperface architecture









# Performance: IJB-C datasets

- The IJB-C evaluation dataset [2] further extends IJB-B. It contains 31, 334 still images and 117, 542 video frames of 3,531 subjects. In addition to the evaluations from IJB-B, this dataset evaluates end-to-end recognition which is the 1:N wild probe. There are about 20, 000 genuine comparisons, and about 15.6 million impostor pairs in the verification protocol. For the 1:N mixed search protocol, there are about 20, 000 probe templates.

[1] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. C. Adams, T. Miller, N. D. Kalka, A. K. Jain, J. A. Duncan, K. Allen et al., "IARPA Janus Benchmark-B face dataset," in CVPR Workshops, 2017, pp. 592–600.

[2] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney et al., "IARPA Janus Benchmark–C: Face dataset and protocol," in 11th IAPR International Conference on

Biometrics, 2018.

# UMD-Janus: Results (IJB-C 1:1 Verification)

IJB-C dataset contains 3,548 subjects with 21,295 still images and 117542 video frames sampled from 11,799 videos in addition to 10,044 non-face images as distractors.

| | (True Accept Rate % @ False Accept Rate) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $10^{-8}$ | $10^{-7}$ | $10^{-6}$ | $10^{-5}$ | $10^{-4}$ | $10^{-3}$ | $10^{-2}$ | $10^{-1}$ |
| Center Loss | 36.0 | 37.6 | 66.1 | 78.1 | 85.3 | 91.2 | 95.3 | 98.2 |
| MN-vc | - | - | - | - | 86.2 | 92.7 | 96.8 | 98.9 |
| SENet50+DCN | - | - | - | - | 88.5 | 94.7 | **98.3** | **99.8** |
| ArcFace | - | - | **85.4** | **92.8** | **95.6** | **97.2** | 98.0 | 98.8 |
| $\text{UMD}_A$ | 16.5 | 19.5 | 43.6 | 77.6 | 91.9 | 95.6 | 97.8 | 99.0 |
| $\text{UMD}_R$ | **60.6** | **67.4** | 76.4 | 86.2 | 91.9 | 95.7 | 97.9 | 99.2 |
| $\text{UMD}_{\text{(Fused)}}$ | 54.1 | 55.9 | 69.5 | 86.9 | 92.5 | 95.9 | 97.9 | 99.2 |

# UMD-Janus: Results (IJB-C 1:N Identification)

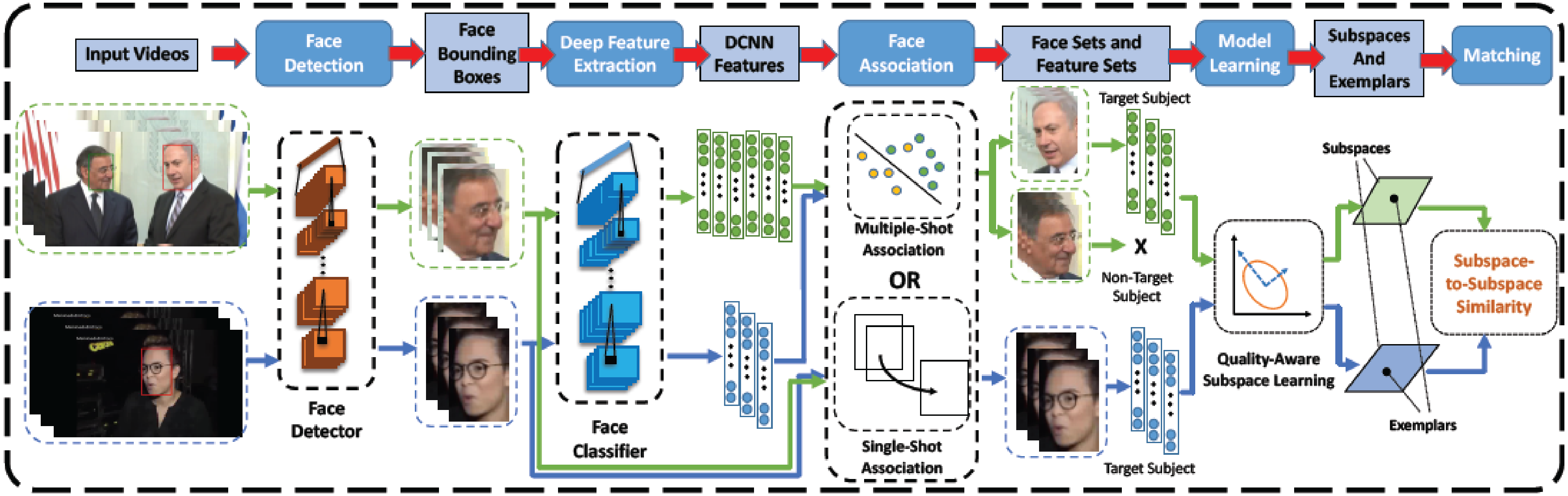
| | TPIR % @ FPIR (G1,G2) | | Retrieval Rate (%) (G1,G2) | | |
|---|---|---|---|---|---|
| | 0.01 | 0.1 | Rank=1 | Rank=5 | Rank=10 |
| Center Loss | 79.1, 75.3 | 86.4, 84.2 | 91.7, 89.8 | 94.6, 93.6 | 95.6, 94.9 |
| $\text{UMD}_A$ | 87.7, 82.4 | 93.5, 91.0 | 95.7, 92.8 | 97.4, 95.4 | 97.9, 96.4 |
| $\text{UMD}_R$ | 88.0, 84.2 | 93.2, 90.6 | 95.9, 93.2 | 97.6, 96.1 | 98.1, **97.0** |
| $\text{UMD}_{(\text{Fused})}$ | **89.6, 85.0** | **93.8, 91.3** | **96.2, 93.6** | **97.7, 96.2** | **98.2**, 96.9 |

# Unconstrained video-based face recognition

- Recognize the identity of the target face in a video
  - Conventional task: frame-by-frame bounding boxes of the target are given in the single-shot video. (e.g. Youtube Faces dataset, PaSC dataset, etc.)
  - End-to-end face identification tasks for JANUS dataset:
    - Video-template creation
    - Open-set face identification

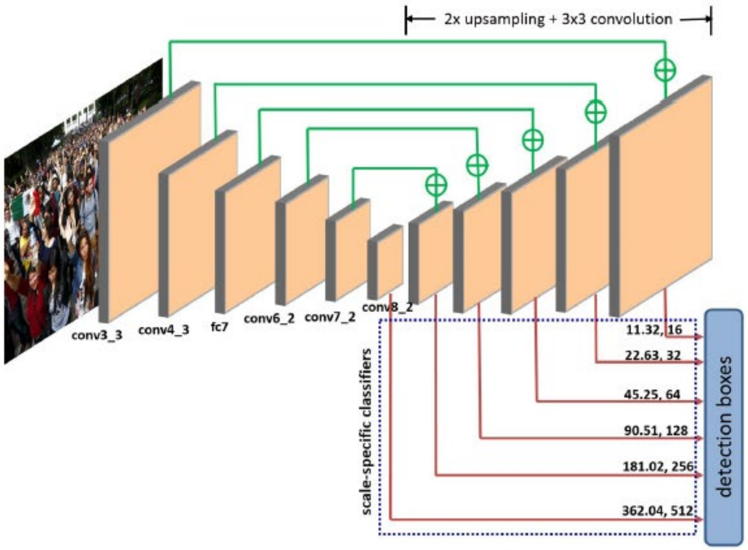
CS6 (single-shot surveillance Videos)

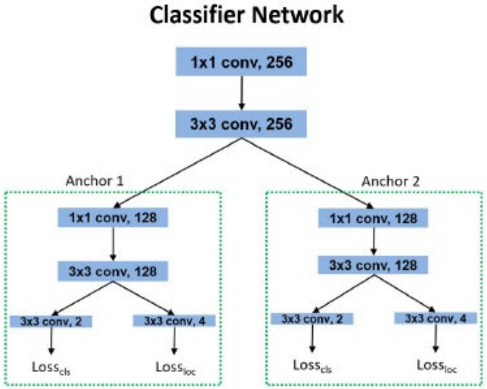
# Video-based face recognition pipeline



- Jingxiao Zheng, Rajeev Ranjan, Ching-Hui Chen, Jun-Cheng Chen,Carlos D. Castillo, and Rama Chellappa. "An Automatic System for Unconstrained Video-based Face Recognition." IEEE T-BIOM, July 2020.

# Deep pyramid single shot face detector (DPSSD)



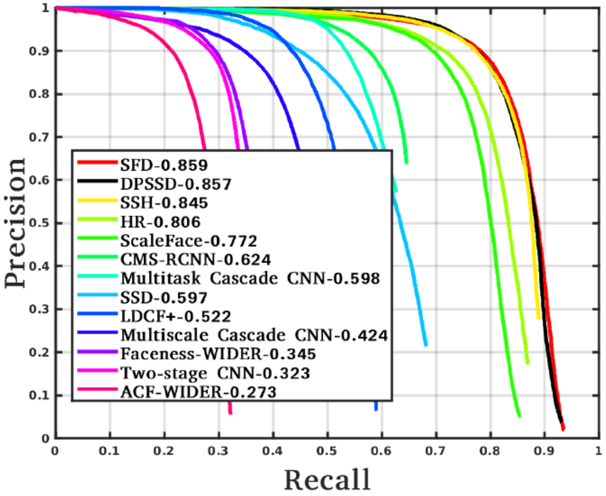

Classifier Network

1x1 conv, 256

3x3 conv, 256

Anchor 1: 1x1 conv, 128 → 3x3 conv, 128 → 3x3 conv, 2 → $Loss_{cls}$; 3x3 conv, 4 → $Loss_{loc}$

Anchor 2: 1x1 conv, 128 → 3x3 conv, 128 → 3x3 conv, 2 → $Loss_{cls}$; 3x3 conv, 4 → $Loss_{loc}$

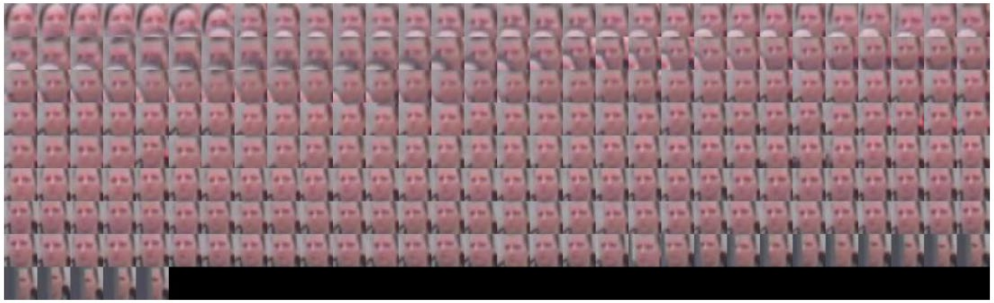


WiderFace Hard

- Ranjan, Rajeev, Ankan Bansal, Jingxiao Zheng, Hongyu Xu, Joshua Gleason, Boyu Lu, Anirudh Nanduri, Jun-Cheng Chen, Carlos D. Castillo, and Rama Chellappa. "A Fast and Accurate System for Face Detection, Identification, and Verification." *arXiv preprint arXiv:1809.07586* (2018).

# Face association for single-shot video

- Simple Online and Real-Time Tracking (SORT)
  - Multi-target data association for detected boxes using Kalman filters
  - Leverage the temporal contiguousness  for the bounding boxes



- Bewley, Alex, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. "Simple online and realtime tracking." In *Image Processing (ICIP), 2016 IEEE International Conference on*, pp. 3464-3468. IEEE, 2016.

# Subspace-based representations

- Given deep features $\mathbf{Y}$, we learn the subspace representation $\mathbf{P}$ by

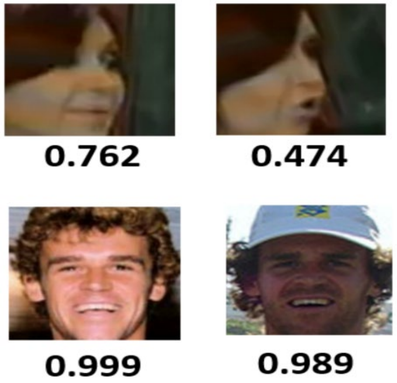
1. Subspace Learning (Sub):

$$\underset{\mathbf{P},\mathbf{X}}{\text{minimize}} \ \|\mathbf{Y} - \mathbf{P}\mathbf{X}\|_F^2 \quad s.t. \ \mathbf{P}^T\mathbf{P} = \mathbf{I}$$

2. Quality-aware Subspace Learning (QSub):

$$\underset{\mathbf{P},\mathbf{X}}{\text{minimize}} \sum_{i=1}^{N} \boxed{\tilde{d}_i} \|\mathbf{y}_i - \mathbf{P}\mathbf{x}_i\|_2^2 \quad s.t. \ \mathbf{P}^T\mathbf{P} = \mathbf{I}$$

Normalized Detection Confidence as face quality indicator



Examples of faces with detection probability.

# Similarity metrics

- The similarity metrics between two sets of deep representations $\mathbf{Y}_1$ and $\mathbf{Y}_2$:

1. Projection Metric (PM)

Principle angles between bases

$$s_{PM}(\mathbf{P}_1, \mathbf{P}_2) = \sqrt{\frac{1}{r}\sum_{k=1}^{r} \boxed{\cos^2\theta_k}} = \sqrt{\frac{1}{r}\|\mathbf{P}_1^T\mathbf{P}_2\|_F^2}$$

1. Variance-aware Projection Metric (VPM)

$$S_{VPM}(\mathbf{P}_1, \mathbf{P}_2) = \sqrt{\frac{1}{r}\sum_{k=1}^{r} \alpha^2(\lambda_{1k})\alpha^2(\lambda_{2k})\cos^2\theta_k} = \sqrt{\frac{1}{r}\|\tilde{\mathbf{P}}_1^T\tilde{\mathbf{P}}_2\|_F^2}$$

$$\text{where } \tilde{\mathbf{P}}_i = \mathbf{P}_i diag\{\alpha(\lambda_{ik})\}$$

Eigenvalues in PCA

# Similarity metrics - 2

3. Cosine similarity (Cos):

$$s_{cos}(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{\mathbf{e}_1^T \mathbf{e}_2}{\|\mathbf{e}_1\|_2 \|\mathbf{e}_2\|_2}$$

4. Quality-aware cosine similarity (QCos):

$$s_{Qcos}(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{\mathbf{e}_{D1}^T \mathbf{e}_{D2}}{\|\mathbf{e}_{D1}\|_2 \|\mathbf{e}_{D2}\|_2}$$

5. Combining the quality-aware subspace learning, quality-aware average pooling and variance-aware projection metric, the overall similarity is

$$s(\mathbf{Y}_1, \mathbf{Y}_2) = s_{Qcos}(\mathbf{Y}_1, \mathbf{Y}_2) + \lambda s_{VPM}(\mathbf{P}_{D1}, \mathbf{P}_{D2})$$

# Deep networks for face representation

| | Deep networks | |
|---|---|---|
| | Rajeev-G1 | Ankan-G1 |
| Training Set | MS1M-Curated + UMDFace Still/Videos | MS1M-Curated + UMDFace Still/Videos |
| Base Architecture | ResNet-101 | Inception-ResNet |
| Loss Function | L2-Softmax | L2-softmax |
| Embedding | TPE (UMDFace stills) | TPE (UMDFace stills) |
| Alignment + Box Size | All-in-One Face<br>224x224 | All-in-One Face<br>299x299 |

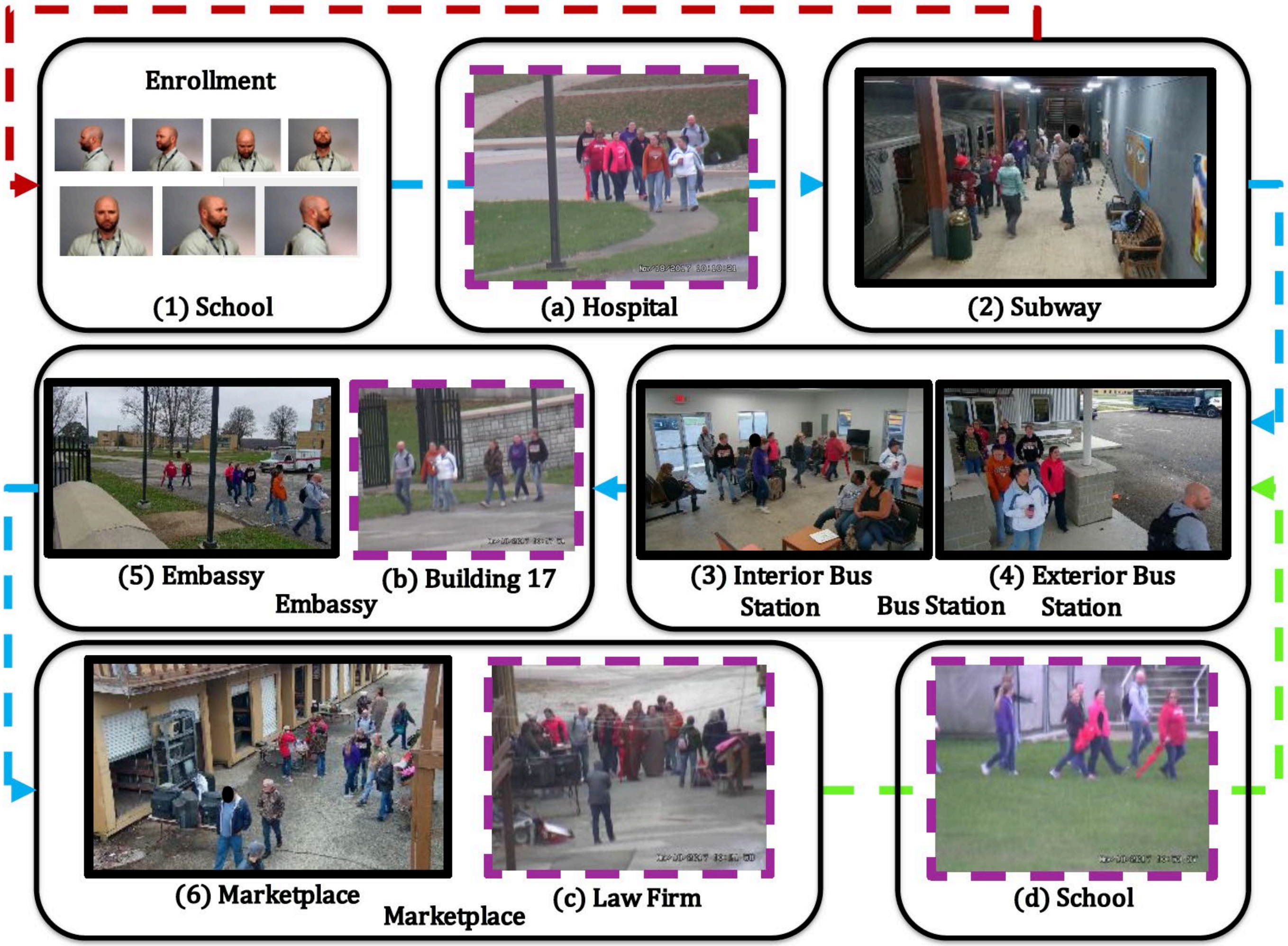
- Ranjan, Rajeev, Ankan Bansal, Hongyu Xu, Swami Sankaranarayanan, Jun-Cheng Chen, Carlos D. Castillo, and Rama Chellappa. "Crystal Loss and Quality Pooling for Unconstrained Face Verification and Recognition." *arXiv preprint arXiv:1804.01159* (2018).
- Ranjan, Rajeev, Ankan Bansal, Jingxiao Zheng, Hongyu Xu, Joshua Gleason, Boyu Lu, Anirudh Nanduri, Jun-Cheng Chen, Carlos D. Castillo, and Rama Chellappa. "A Fast and Accurate System for Face Detection, Identification, and Verification." *arXiv preprint arXiv:1809.07586* (2018).

# System details

- Face Detection
  - Multi-task SSD (Chen *et al.* 2018) for high quality faces,
  - DPSSD (Ranjan *et al.* 2019) for tiny faces.
- Facial Landmark Estimation
  - All-in-One Face (Ranjan *et al.* 2017)
- Face Association
  - SORT tracking for single-shot videos,
  - TFA (Chen *et al.* 2017) association for multi-shot videos.
- ResNet-101 and Inception-ResNet-v2, both trained on the union of MSCeleb-1M, UMDFaces, and UMDFaces Video datasets with the crystal loss.
- Features are further reduced to 128-dimensional by a Triplet Probabilistic Embedding (TPE).

# IARPA JANUS surveillance video benchmark (IJB-S)

- An unconstrained video-based face recognition dataset.
- Galleries: high-resolution still images. Probes: low quality, remotely captured surveillance videos.
- 202 subjects from 1421 images and 398 single-shot surveillance videos.
- We focus on surveillance-to-single , surveillance-to-booking and surveillance-to-surveillance identification protocols.


Enrollment
(1) School
(a) Hospital
(2) Subway
(5) Embassy
Embassy
(b) Building 17
(3) Interior Bus Station
Bus Station
(4) Exterior Bus Station
(6) Marketplace
Marketplace
(c) Law Firm
(d) School

# Identification results on IJB-S

## Surveillance-to-Single

| Methods | Top-K Average Accuracy **with Filtering** | | | | | | EERR metric **without Filtering** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R=1 | R=2 | R=5 | R=10 | R=20 | R=50 | R=1 | R=2 | R=5 | R=10 | R=20 | R=50 |
| Arc-Cos (Deng *et al.*) | 52.03% | 56.83% | 63.16% | 69.05% | 76.13% | 88.95% | 24.45% | 26.54% | 29.35% | 32.33% | 36.38% | 44.81% |
| Arc-QCos+QSub-PM | 60.92% | 65.06% | 70.45% | 75.19% | 80.69% | 90.29% | 28.73% | 30.44% | 32.98% | 35.40% | 38.70% | 45.46% |
| Cos | 64.86% | 70.87% | 77.09% | 81.53% | 86.11% | 93.24% | 29.62% | 32.34% | 35.60% | 38.36% | 41.53% | 46.78% |
| QCos | 65.42% | 71.34% | 77.37% | 81.78% | 86.25% | 93.29% | 29.94% | 32.60% | 35.85% | 38.52% | 41.70% | 46.78% |
| Cos+Sub-PM | 69.52% | 75.15% | 80.41% | 84.14% | 87.83% | 94.27% | 32.22% | 34.70% | 37.66% | 39.91% | 42.65% | 47.54% |
| QCos+Sub-PM | 69.65% | 75.26% | 80.43% | 84.22% | 87.81% | 94.25% | 32.27% | 34.73% | 37.66% | 39.91% | 42.67% | 47.54% |
| QCos+QSub-PM | **69.82%** | **75.38%** | **80.54%** | **84.36%** | **87.91%** | **94.34%** | **32.43%** | **34.89%** | **37.74%** | **40.01%** | **42.77%** | **47.60%** |
| QCos+QSub-VPM | 69.43% | 75.24% | 80.34% | 84.14% | 87.86% | 94.28% | 32.19% | 34.75% | 37.68% | 39.88% | 42.56% | 47.50% |

## Surveillance-to-Booking

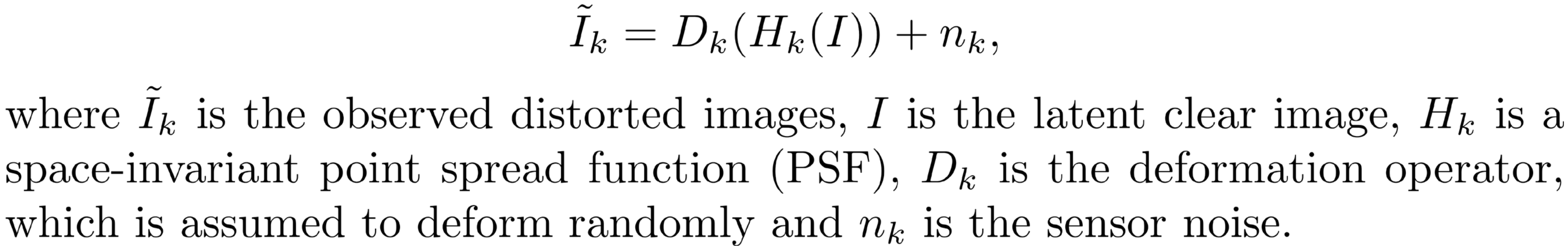
| Methods | Top-K Average Accuracy **with Filtering** | | | | | | EERR metric **without Filtering** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R=1 | R=2 | R=5 | R=10 | R=20 | R=50 | R=1 | R=2 | R=5 | R=10 | R=20 | R=50 |
| Arc-Cos (Deng *et al.*) | 54.59% | 59.12% | 65.43% | 71.05% | 77.84% | 89.16% | 25.38% | 27.58% | 30.59% | 33.42% | 37.60% | 45.05% |
| Arc-QCos+QSub-VPM | 60.86% | 65.36% | 71.30% | 76.15% | 81.63% | 90.70% | 28.66% | 30.64% | 33.43% | 36.11% | 39.57% | 45.70% |
| Cos | 66.48% | 71.98% | 77.80% | 82.25% | 86.56% | 93.41% | 30.38% | 32.91% | 36.15% | 38.77% | 41.86% | 46.79% |
| QCos | 66.94% | 72.41% | 78.04% | 82.37% | 86.63% | 93.43% | 30.66% | 33.17% | 36.28% | 38.84% | 41.88% | 46.84% |
| Cos+Sub-PM | 69.39% | 74.55% | 80.06% | 83.91% | 87.87% | **94.34%** | 32.02% | 34.42% | 37.59% | 39.97% | 42.64% | **47.58%** |
| QCos+Sub-PM | 69.57% | 74.78% | 80.06% | 83.89% | 87.94% | 94.33% | 32.16% | 34.61% | 37.62% | 39.99% | 42.71% | 47.57% |
| QCos+QSub-PM | 69.67% | 74.85% | 80.25% | 84.10% | 88.04% | 94.22% | 32.28% | 34.77% | 37.76% | 40.11% | **42.76%** | 47.57% |
| QCos+QSub-VPM | **69.86%** | **75.07%** | **80.36%** | **84.32%** | **88.07%** | 94.33% | **32.44%** | **34.93%** | **37.80%** | **40.14%** | 42.72% | **47.58%** |

# Identification results on IJB-S

## Surveillance-to-Surveillance

| Methods | Top-K Average Accuracy **with Filtering** | | | | | | EERR metric **without Filtering** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R=1 | R=2 | R=5 | R=10 | R=20 | R=50 | R=1 | R=2 | R=5 | R=10 | R=20 | R=50 |
| Arc-Cos (Deng *et al.*) | 8.68% | 12.58% | 18.79% | 26.66% | 39.22% | 68.19% | 4.98% | 7.17% | 10.86% | 15.42% | 22.34% | 37.68% |
| Arc-QCos+QSub-PM | 8.64% | 12.57% | 18.84% | 26.86% | 39.78% | **68.21%** | **5.26%** | **7.44%** | **11.31%** | 15.90% | **22.68%** | **37.83%** |
| Cos | 8.54% | 11.99% | 19.60% | 28.00% | 37.71% | 59.44% | 4.42% | 6.15% | 10.84% | 15.73% | 21.14% | 33.21% |
| QCos | 8.62% | 12.11% | 19.62% | 28.14% | 37.78% | 59.21% | 4.46% | 6.20% | 10.80% | 15.81% | 21.06% | 33.17% |
| Cos+Sub-PM | 8.19% | 11.79% | 19.56% | 28.62% | 39.77% | 63.15% | 4.26% | 6.25% | 10.79% | 16.18% | 22.48% | 34.82% |
| QCos+Sub-PM | 8.24% | 11.82% | 19.68% | 28.68% | 39.68% | 62.96% | 4.27% | 6.25% | 10.92% | 16.18% | 22.39% | 34.69% |
| QCos+QSub-PM | 8.33% | 11.88% | 19.82% | 28.65% | 39.78% | 62.79% | 4.33% | 6.21% | 10.96% | 16.19% | 22.48% | 34.69% |
| QCos+QSub-VPM | 8.66% | 12.27% | **19.91%** | **29.03%** | **40.20%** | 63.20% | 4.30% | 6.30% | 10.99% | **16.23%** | 22.50% | 34.76% |

# Face recognition from atmospheric turbulence degraded images

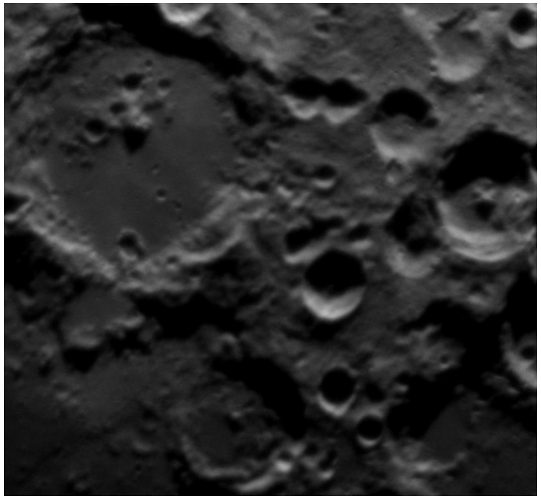


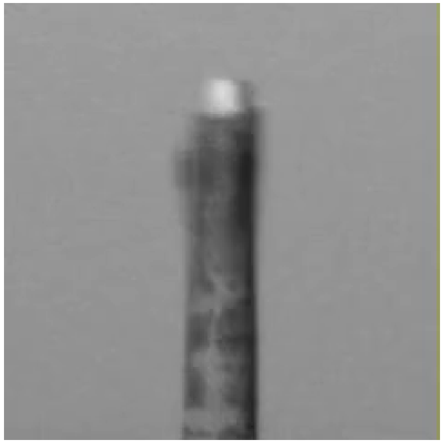
$$\tilde{I}_k = D_k(H_k(I)) + n_k,$$

where $\tilde{I}_k$ is the observed distorted images, $I$ is the latent clear image, $H_k$ is a space-invariant point spread function (PSF), $D_k$ is the deformation operator, which is assumed to deform randomly and $n_k$ is the sensor noise.

- Turbulence degraded images: effects of the turbulent flow of air and changes in temperature, density of air particles, humidity and carbon dioxide level, the captured image is blurry and deformed due to variations in the refractive index.
- Decreases the visual quality and the performance of different computer vision tasks such as object detection and face recognition.

# Typical examples





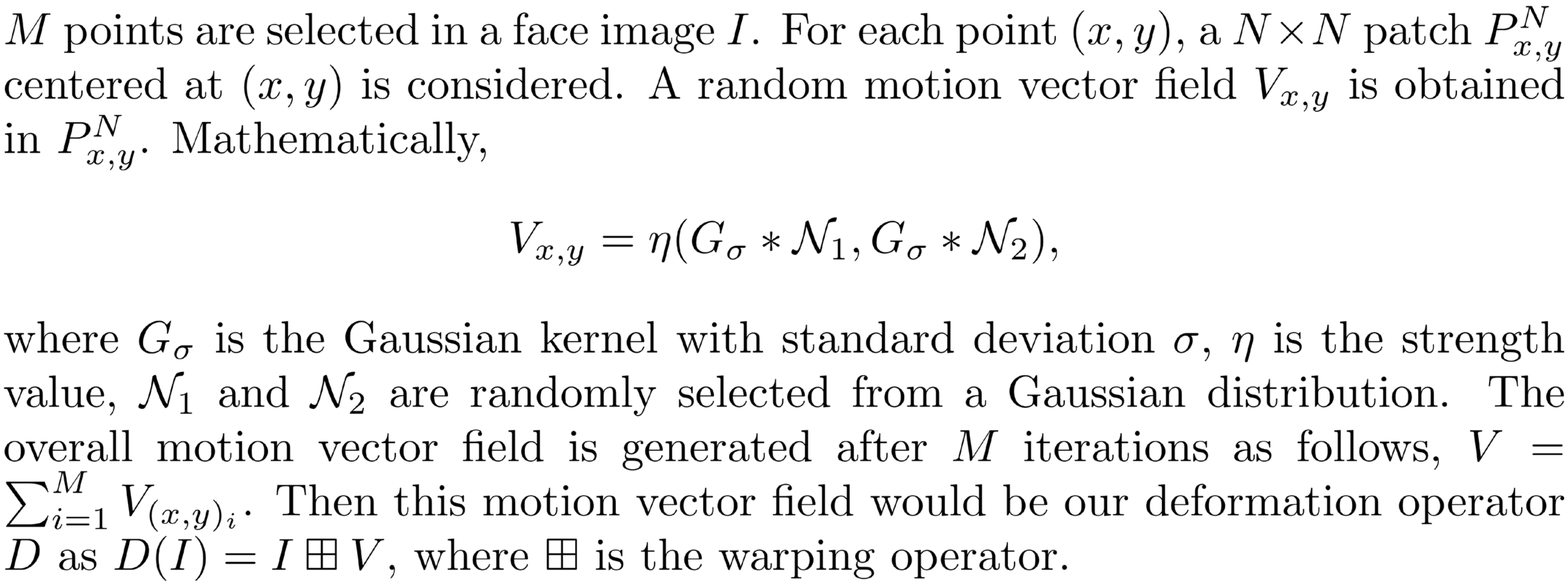
# Problem formulation

$$\tilde{I} = D(H(I)) + n,$$

- A more challenging and practical setting: one frame is available to reconstruct the latent clean image
- Build a restoration function *G* to restore the distorted face image, i.e. $G(\tilde{I}) = I$
- The **Wasserstein GAN with gradient penalty** is employed.
- Denote blurry image and deformed image as $I_b$ and $I_d$ respectively. Build a deblur function $G_d$ and a deformation correction $G_b$ to remove undesired blur and deformation, i.e. $G_d(\tilde{I}) = I_d$, $G_b(\tilde{I}) = I_b$ and $G(\tilde{I}) = G_d(G_b(\tilde{I})) = I$.
- Then the **turbulence is decomposed into blur and deformation**.
- The "mixing" of deformation and blur in realistic turbulence face images is very fast and we could not be sure whether deformation precedes blur or blur precedes deformation.
- **Commutative constraint** is enforced, i.e. $D(H(I)) = H(D(I)) = \tilde{I}$.

# Data augmentation

- 10000 aligned face images are picked from UMDFaces
- Use Gaussian blurring kernel with different variance as H
- Construct D as follows:

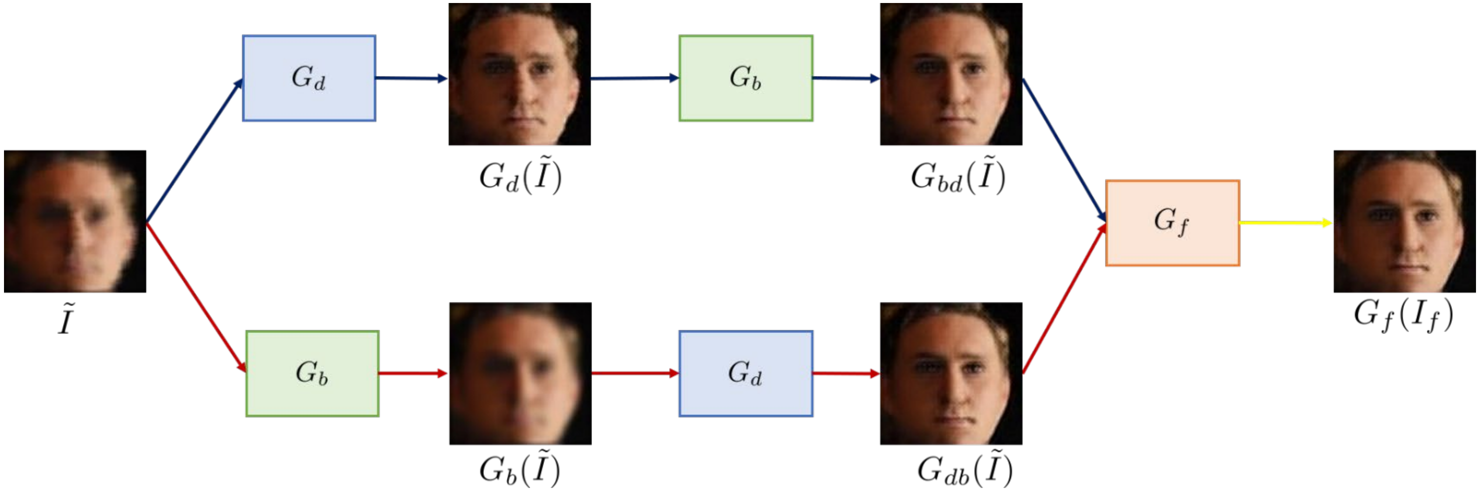


$M$ points are selected in a face image $I$. For each point $(x, y)$, a $N \times N$ patch $P^N_{x,y}$ centered at $(x, y)$ is considered. A random motion vector field $V_{x,y}$ is obtained in $P^N_{x,y}$. Mathematically,

$$V_{x,y} = \eta(G_\sigma * \mathcal{N}_1, G_\sigma * \mathcal{N}_2),$$

where $G_\sigma$ is the Gaussian kernel with standard deviation $\sigma$, $\eta$ is the strength value, $\mathcal{N}_1$ and $\mathcal{N}_2$ are randomly selected from a Gaussian distribution. The overall motion vector field is generated after $M$ iterations as follows, $V = \sum_{i=1}^{M} V_{(x,y)_i}$. Then this motion vector field would be our deformation operator $D$ as $D(I) = I \boxplus V$, where $\boxplus$ is the warping operator.

# Proposed method



- Using Wasserstein GAN with gradient penalty
- Split the turbulence degradation due to blur and deformation in the training stage

  Introduce deblur function $G_d$ and deformation correction function $G_b$.
- Enforce novel constraint: commutative constraint:

  Denote $G_{bd} = G_b \circ G_d$ and $G_{db} = G_d \circ G_b$. Mathematically, $G(\tilde{I}) = G_f(I_f)$, where $G_f$ is a image fusion function and $I_f$ is the pixelwise average of the restored image pair $(G_{bd}(\tilde{I}), G_{db}(\tilde{I}))$.

# Loss functions

- Content Loss: $\mathcal{L}_{con} = \|G_b(\tilde{I}) - I_b\|_2^2 + \|G_d(\tilde{I}) - I_d\|_2^2$
- Commutative Constraint: $\mathcal{L}_{cc} = \|G_{db}(\tilde{I}) - I)\|_2^2 + \|G_{bd}(\tilde{I}) - I)\|_2^2$
- Fusion Loss: $\mathcal{L}_f = \|G_f(I_f) - I\|_2^2$
- Adversarial Loss:

$$\mathcal{L}^{\mathcal{I}_i}_{Dis} = \mathbb{E}_{\tilde{I}\sim\tilde{\mathcal{I}}}[D_i(G_i(\tilde{I}))] - \mathbb{E}_{I_i\sim\mathcal{I}_i}[D_i(I_i))] + \lambda_{WGAN}\cdot\mathbb{E}_{\hat{I}_i\sim\widehat{\mathcal{I}}_i}[(\|\nabla_{\hat{I}_i} D_i(\hat{I})\|_2 - 1)^2],$$

$$\mathcal{L}^{\mathcal{I}_j}_{Dis} = \mathbb{E}_{I_j\sim\mathcal{I}_j}[D_f(G_j(I_j))] - \mathbb{E}_{I\sim\mathcal{I}}[D_f(I))] + \lambda_{WGAN}\cdot\mathbb{E}_{\hat{I}_j\sim\widehat{\mathcal{I}_j}}[(\|\nabla_{\hat{I}} D_f(\hat{I})\|_2 - 1)^2],$$

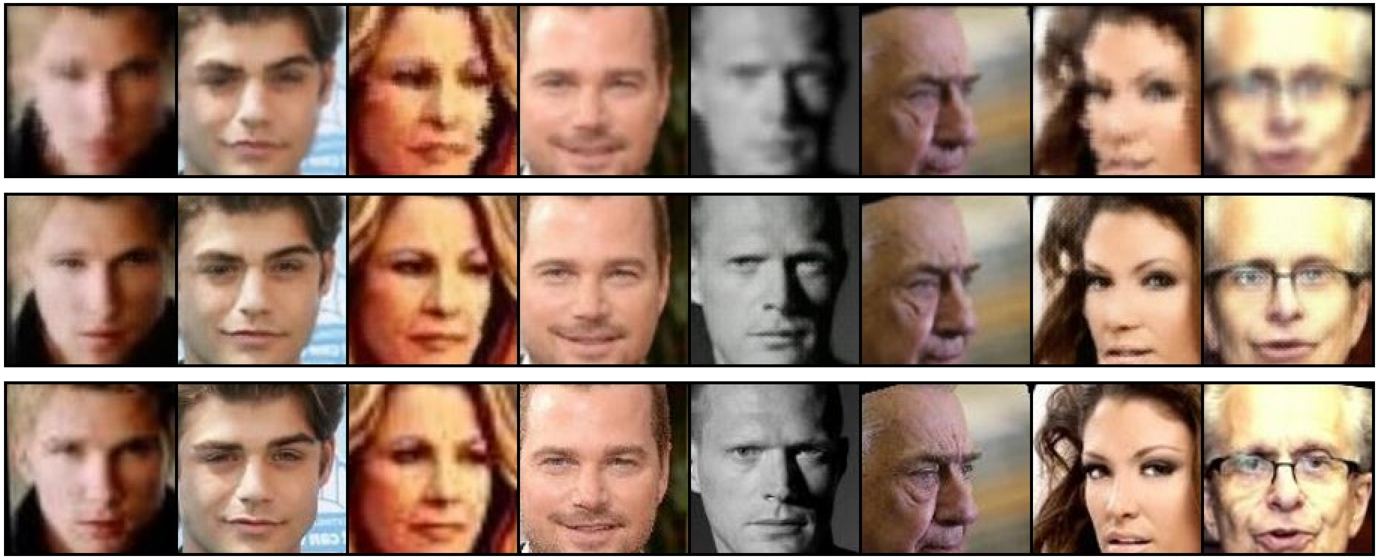
$\mathcal{L}^{\mathcal{I}_k}_{Gen} = -\mathbb{E}_{I_k\sim\mathcal{I}_k}[D_k(G_k(I_k))]$, where $\widehat{\mathcal{I}}_i$ is the distribution obtained by randomly interpolating between real images $I_i$ and restored images $G_i(\tilde{I})$, $i \in \{b, d\}$, $j \in \{bd, db, f\}$ and $k \in \{b, d, bd, db, f\}$. For convenience of notation, $I_{bd} = I_{db} = \tilde{I}$, $\mathcal{I}_{bd} = \mathcal{I}_{db} = \tilde{\mathcal{I}}$ and $D_{bd} = D_{db} = D_f$.

- Perceptual Loss:

$\mathcal{L}^{\mathcal{I}_i}_p = \|\phi_l(G_i(\tilde{I})) - \phi_l(I_i)\|_2^2, \quad i \in \{b, d\}, \quad \mathcal{L}^{\mathcal{I}_j}_p = \|\phi_l(G_j(I_j)) - \phi_l(I)\|_2^2, \quad j \in \{bd, db, f\}$, where $\phi_l(\cdot)$ is the features of the $l^{\text{th}}$ layer of a pretrained CNN.

- Full Loss Function: $\mathcal{L} = \mathcal{L}_{adv} + \lambda_{con}\mathcal{L}_{con} + \lambda_{cc}\mathcal{L}_{cc} + \lambda_f\mathcal{L}_f + \lambda_p\mathcal{L}_p$

# Results



- The first row is the synthetic atmospheric turbulence degraded images. The second row is the corresponding restored images. The third row is the latent groun dtruth images

# Ablation study

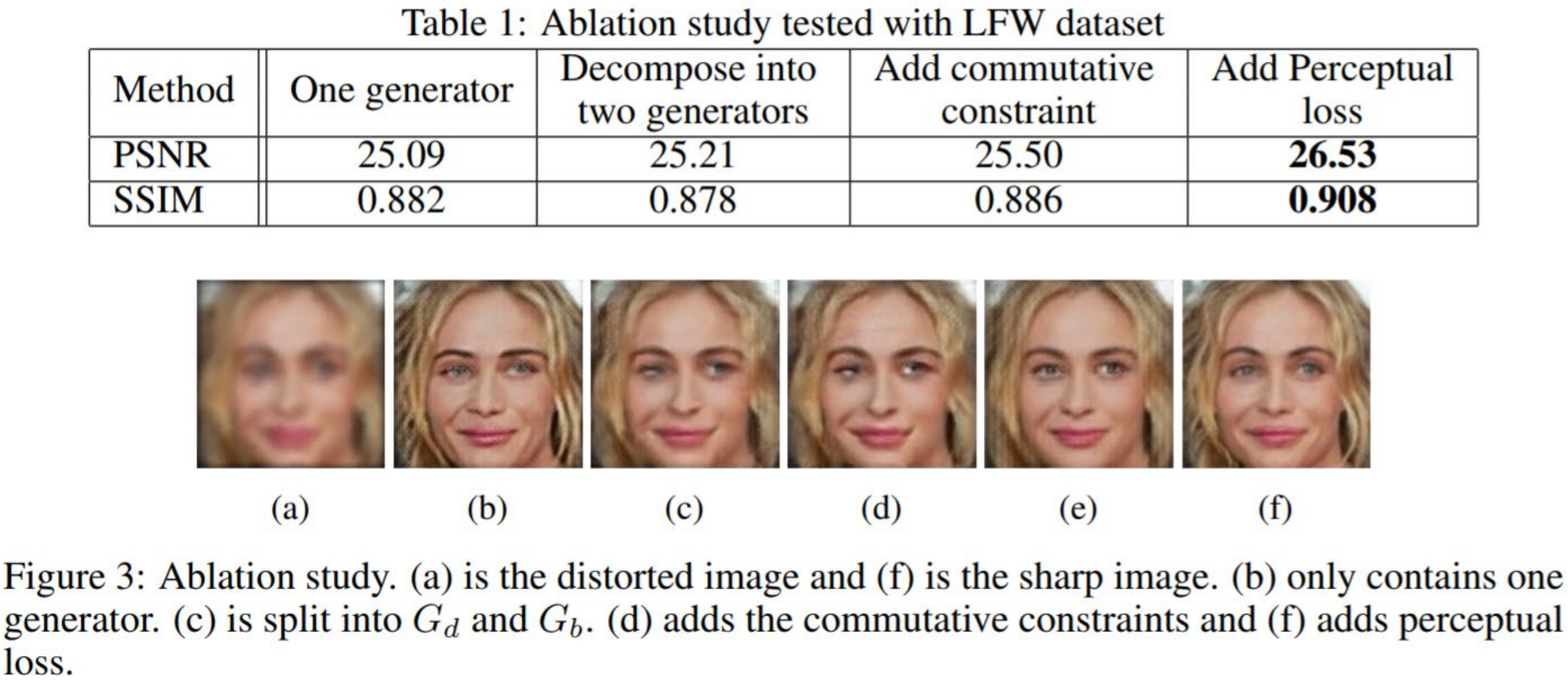


Table 1: Ablation study tested with LFW dataset

| Method | One generator | Decompose into two generators | Add commutative constraint | Add Perceptual loss |
|---|---|---|---|---|
| PSNR | 25.09 | 25.21 | 25.50 | **26.53** |
| SSIM | 0.882 | 0.878 | 0.886 | **0.908** |

(a) (b) (c) (d) (e) (f)

Figure 3: Ablation study. (a) is the distorted image and (f) is the sharp image. (b) only contains one generator. (c) is split into $G_d$ and $G_b$. (d) adds the commutative constraints and (f) adds perceptual loss.

# Qualitative and quantitative evaluation

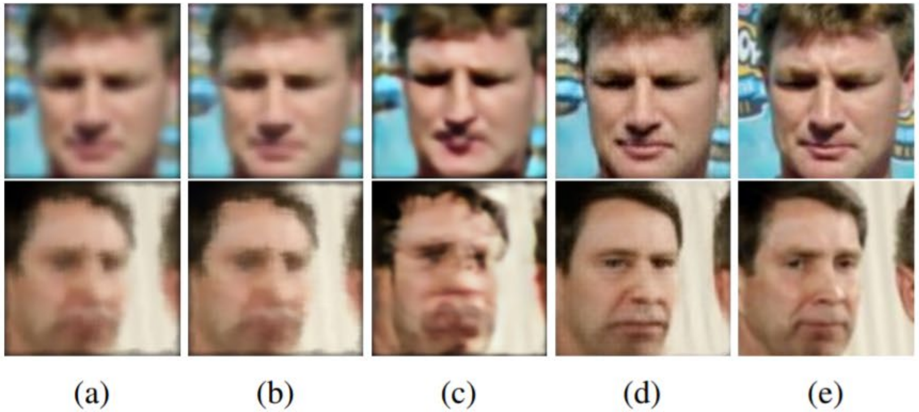


Figure 4: Visual performance comparison with state-of-the-art methods. (a) is the distorted image. (b) Kupyn et al. [17]. (c) Shen et al. [32]. (d) Ours. (e) Groundtruth.

Table 2: Quantitative performance comparison with state-of-the-art methods on LFW dataset

| Distorted | | Kupyn et al. [17] | | Shen et al. [32] | | Ours | |
|---|---|---|---|---|---|---|---|
| PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| 24.17 | 0.878 | 23.89 | 0.867 | 19.60 | 0.768 | **26.53** | **0.908** |

Table 3: Face verification results on the LFW dataset.

| Method | Sharp | Distorted | Kupyn et al. [17] | Shen et al. [32] | Ours |
|---|---|---|---|---|---|
| Accuracy | 0.998 | 0.726 | 0.783 | 0.647 | **0.799** |

# Performance of the disentangled representation

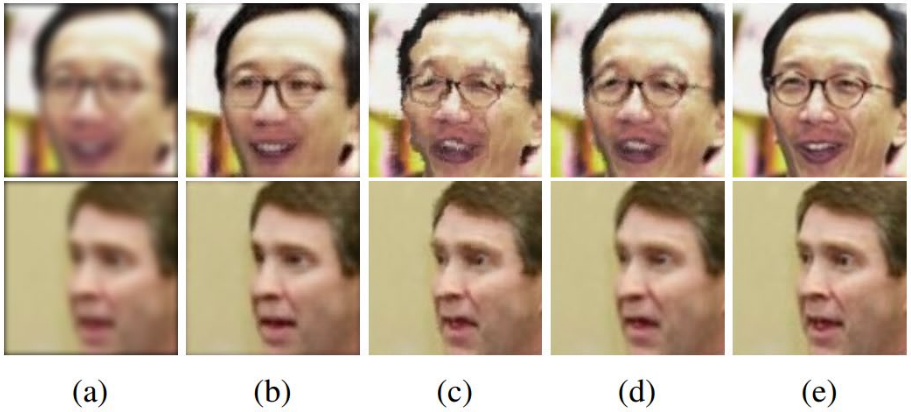
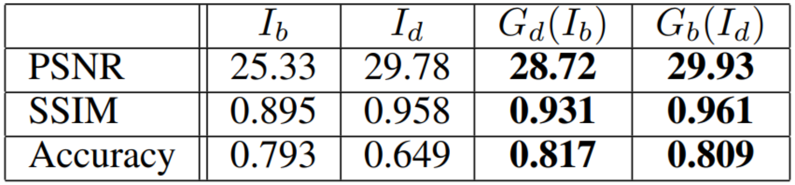


Figure 5: Visual performance comparison of the deblur function $G_d$ and deformation correction $G_b$ with the LFW dataset. (a) Blurry image. (b) Restored image of (a) by $G_d$. (c) Deformed image. (d) Restored image of (c) by $G_b$. (e) Groundtruth.

Table 4: PSNR, SSIM and face verification results for LFW dataset with $G_b$ and $G_d$.



| | $I_b$ | $I_d$ | $G_d(I_b)$ | $G_b(I_d)$ |
|---|---|---|---|---|
| PSNR | 25.33 | 29.78 | **28.72** | **29.93** |
| SSIM | 0.895 | 0.958 | **0.931** | **0.961** |
| Accuracy | 0.793 | 0.649 | **0.817** | **0.809** |

# Bias in gender classification

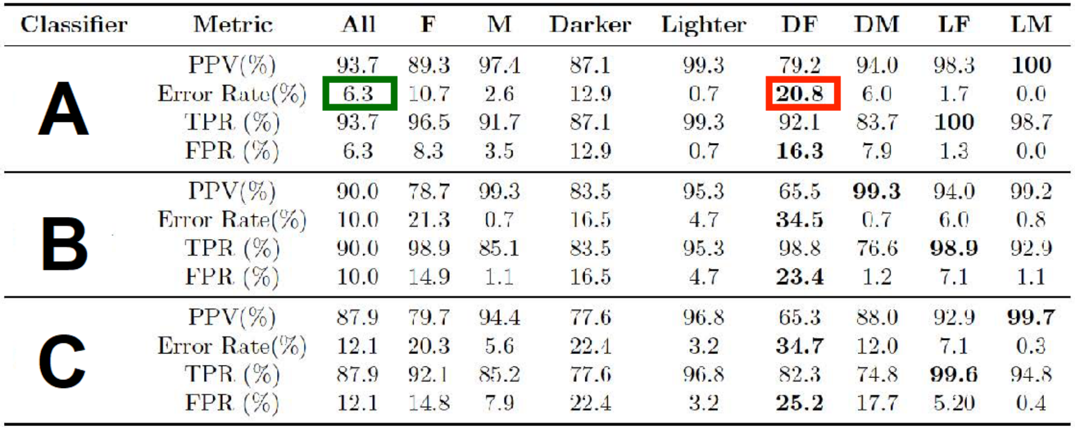
The PPB dataset

AFRICAN SCANDINAVIAN

PPB: Pilot Parliaments Benchmark

6.3%

20.8%

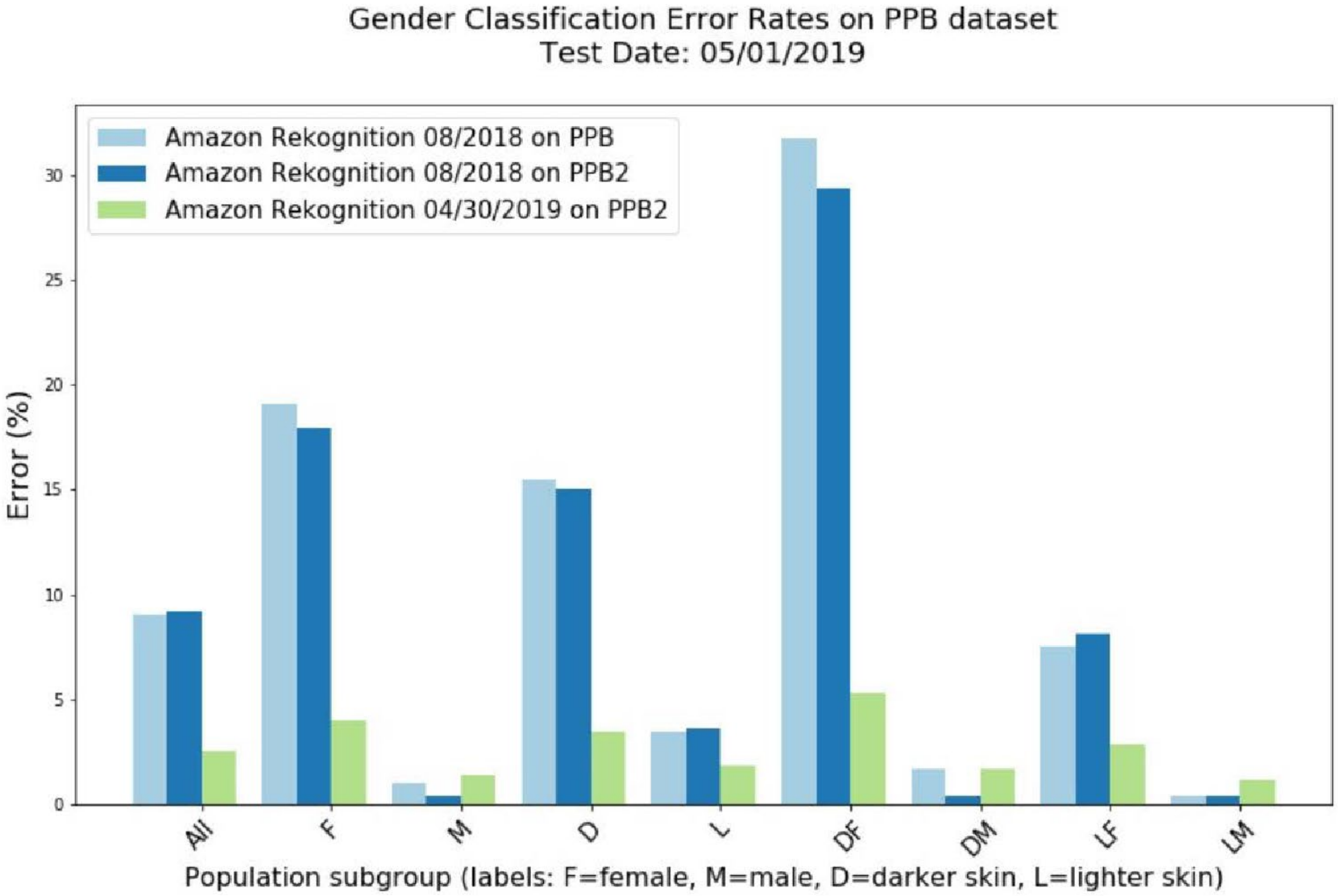


| Classifier | Metric | All | F | M | Darker | Lighter | DF | DM | LF | LM |
|---|---|---|---|---|---|---|---|---|---|---|
| A | PPV(%) | 93.7 | 89.3 | 97.4 | 87.1 | 99.3 | 79.2 | 94.0 | 98.3 | **100** |
| | Error Rate(%) | 6.3 | 10.7 | 2.6 | 12.9 | 0.7 | **20.8** | 6.0 | 1.7 | 0.0 |
| | TPR (%) | 93.7 | 96.5 | 91.7 | 87.1 | 99.3 | 92.1 | 83.7 | **100** | 98.7 |
| | FPR (%) | 6.3 | 8.3 | 3.5 | 12.9 | 0.7 | **16.3** | 7.9 | 1.3 | 0.0 |
| B | PPV(%) | 90.0 | 78.7 | 99.3 | 83.5 | 95.3 | 65.5 | **99.3** | 94.0 | 99.2 |
| | Error Rate(%) | 10.0 | 21.3 | 0.7 | 16.5 | 4.7 | **34.5** | 0.7 | 6.0 | 0.8 |
| | TPR (%) | 90.0 | 98.9 | 85.1 | 83.5 | 95.3 | 98.8 | 76.6 | **98.9** | 92.9 |
| | FPR (%) | 10.0 | 14.9 | 1.1 | 16.5 | 4.7 | **23.4** | 1.2 | 7.1 | 1.1 |
| C | PPV(%) | 87.9 | 79.7 | 94.4 | 77.6 | 96.8 | 65.3 | 88.0 | 92.9 | **99.7** |
| | Error Rate(%) | 12.1 | 20.3 | 5.6 | 22.4 | 3.2 | **34.7** | 12.0 | 7.1 | 0.3 |
| | TPR (%) | 87.9 | 92.1 | 85.2 | 77.6 | 96.8 | 82.3 | 74.8 | **99.6** | 94.8 |
| | FPR (%) | 12.1 | 14.8 | 7.9 | 22.4 | 3.2 | **25.2** | 17.7 | 5.20 | 0.4 |

GenderShades.Org

[Buolamwini & Gebru 2018]

[Buolamwini 2018]


Gender Classification Error Rates on PPB dataset
Test Date: 05/01/2019
Amazon Rekognition 08/2018 on PPB
Amazon Rekognition 08/2018 on PPB2
Amazon Rekognition 04/30/2019 on PPB2
Error (%)
0
5
10
15
20
25
30
All
F
M
D
L
DF
DM
LF
LM
Population subgroup (labels: F=female, M=male, D=darker skin, L=lighter skin)

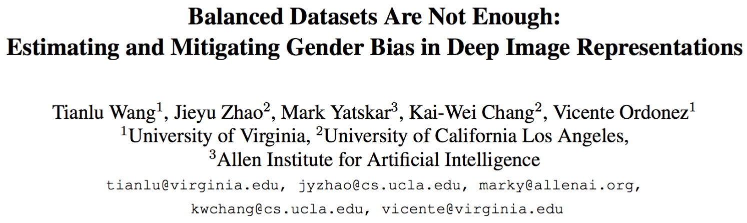
# Why are face recognition systems biased?

- Imbalanced training datasets [Albiero et. al, IJCB 2020]
- Use of cosmetics (gender-bias) [Albiero et. al, WACV-W 2020]
- Gendered Hairstyles (gender-bias) [Albiero et. al, BMVC 2020]
- Implicit encoding of gender and skin tone in face recognition features [Hill et. al, Nature MI 2019; Dhar et. al, FG 2020]

# Balancing does not work

- Several papers have experimentally verified that training a network on a balanced dataset does not mitigate bias

Wang et. al, ICCV 2019

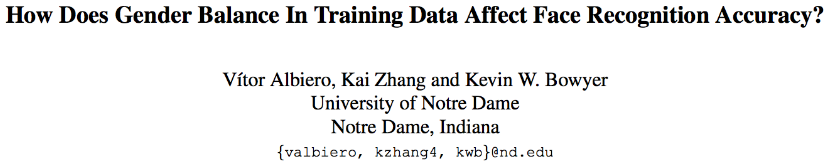


**Balanced Datasets Are Not Enough:**
**Estimating and Mitigating Gender Bias in Deep Image Representations**

Tianlu Wang[1], Jieyu Zhao[2], Mark Yatskar[3], Kai-Wei Chang[2], Vicente Ordonez[1]
[1]University of Virginia, [2]University of California Los Angeles,
[3]Allen Institute for Artificial Intelligence
tianlu@virginia.edu, jyzhao@cs.ucla.edu, marky@allenai.org,
kwchang@cs.ucla.edu, vicente@virginia.edu

"…we show that balanced datasets do not lead to unbiased predictions…"

Albiero et. al, IJCB 2020



**How Does Gender Balance In Training Data Affect Face Recognition Accuracy?**

Vítor Albiero, Kai Zhang and Kevin W. Bowyer
University of Notre Dame
Notre Dame, Indiana
{valbiero, kzhang4, kwb}@nd.edu

"… there is little if any empirical support for the premise that training with a gender-balanced training set will result in gender-balanced accuracy on a test set."

# Why doesn’t balancing work?

- We cannot completely ‘balance’ a dataset.
- Even if the training dataset has equal number of male and female identities, we cannot control the appearance variation in both genders.
- Appearance variation can be affected by yaw (pose), image quality, lighting etc.
- Building a dataset where the lighting, pose, expression etc. is exactly same for males and females in not feasible.

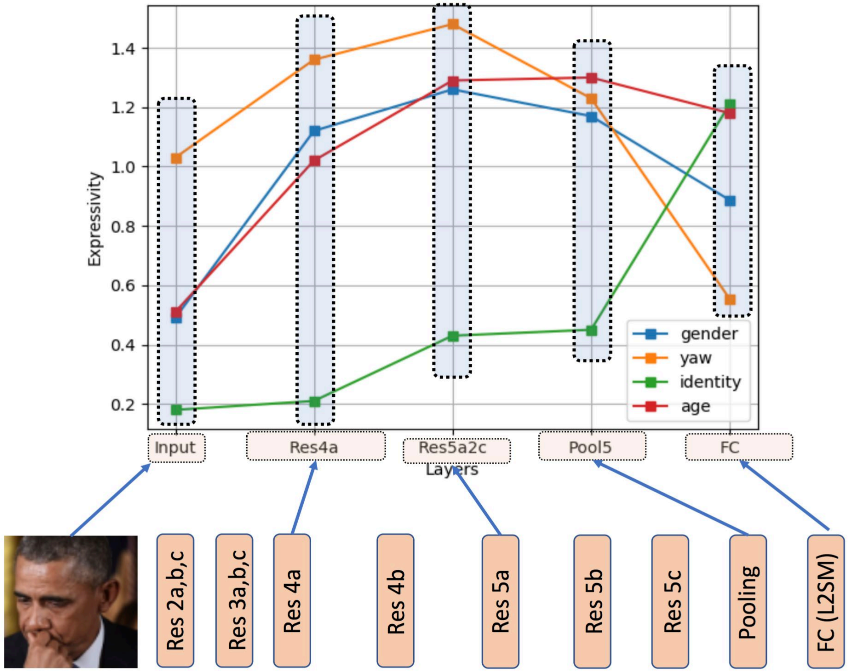
# How are attributes expressed in face DCNNs? (Dhar et. al, FG 2020)

- Face recognition networks are trained to classify identities.
- However, if we train an MLP network to classify gender using features extracted from a trained network, we obtain a very high accuracy.
- High predictability → Implicit encoding of sensitive attributes

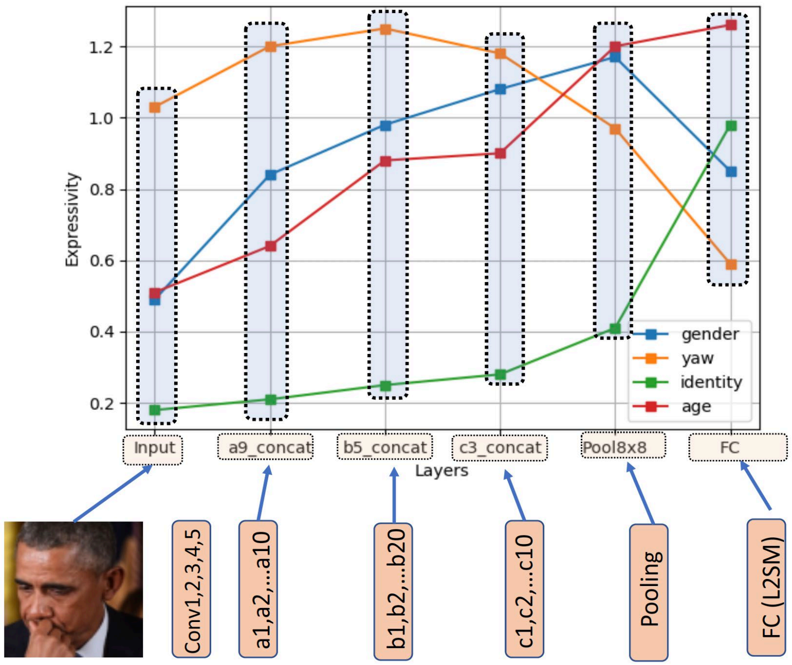
# Expressivity of attributes

- *Expressivity of an entity =* the *ease* with which that entity can be predicted using a given set of features.
- We compute expressivity of facial attributes (yaw, age, gender, identity* ) in a given set of face descriptors.
- To compute expressivity, we approximate the mutual information (MI) between features and attributes, by using an existing approach called Mutual Information Neural Estimation (MINE) [Belghazi et. al, ICML 2018].

# Expressivity of yaw, gender and age



Crystalface
(Resnet-101)



Inception
Resnet-v2

# Key takeaways

- Face recognition features implicitly encode attributes like yaw, gender and age.
- During the training process, the expressivity of identity increases while that of yaw, gender and age decreases, thus showing that *un-learning is a part of learning*. Expressivity of yaw, especially, decreases very rapidly.
- Rate of un-learning: **Age < Gender < Yaw** (opposite to the order of attribute-wise relevance)

# Adversarial removal of sensitive attributes

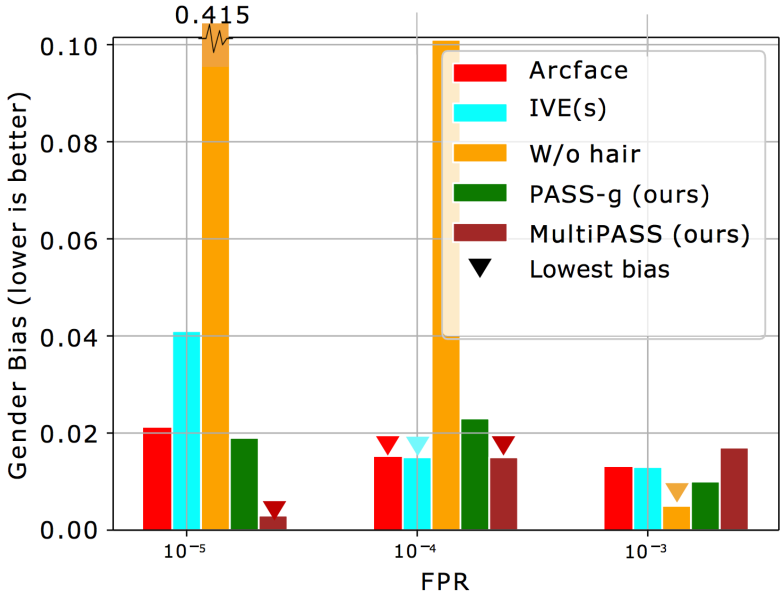
- Implicit encoding of attributes may result in networks demonstrating bias in face recognition.
- Potential solution: Train networks to classify identities, while adversarially removing sensitive attributes
- P. Dhar, A. Roy, J. Gleason, C.D. Castillo and R. Chellappa, ‘’PASS: Protected Attribute Suppression System for Mitigating Bias in Face Recognition”, ICCV 2021.

# Bias-performance Tradeoff

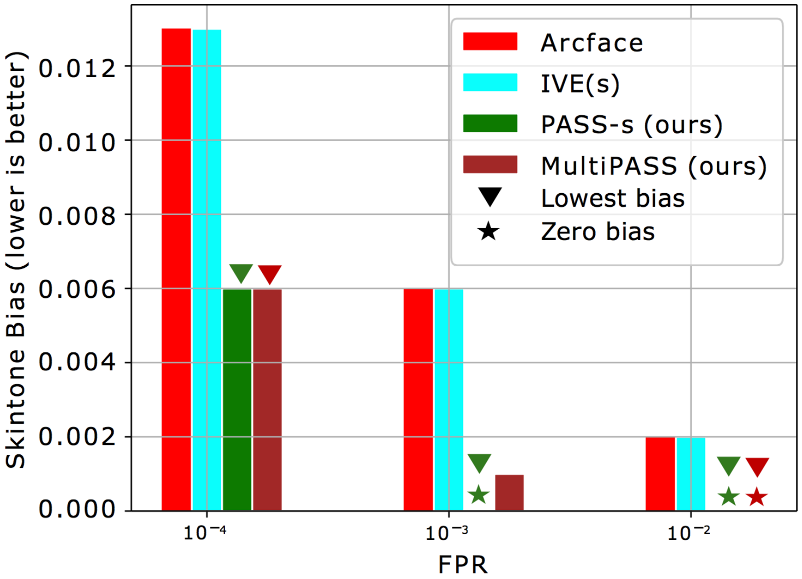
- Most adversarial de-biasing systems demonstrate a drop in face verification performance.
- An ideal face recognition system should demonstrate high bias reduction and low drop in performance.
- To measure this tradeoff between reduction in bias and drop in verification performance, we propose a new metric called Bias Performance Coefficient:

$$\mathrm{BPC}^{(F)} = \underbrace{\frac{\mathrm{Bias}^{(F)} - \mathrm{Bias}_{deb}^{(F)}}{\mathrm{Bias}^{(F)}}}_{\text{\% drop in bias}} - \underbrace{\frac{\mathrm{TPR}^{(F)} - \mathrm{TPR}_{deb}^{(F)}}{\mathrm{TPR}^{(F)}}}_{\text{\% drop in TPR}}$$

# Results (Arcface)

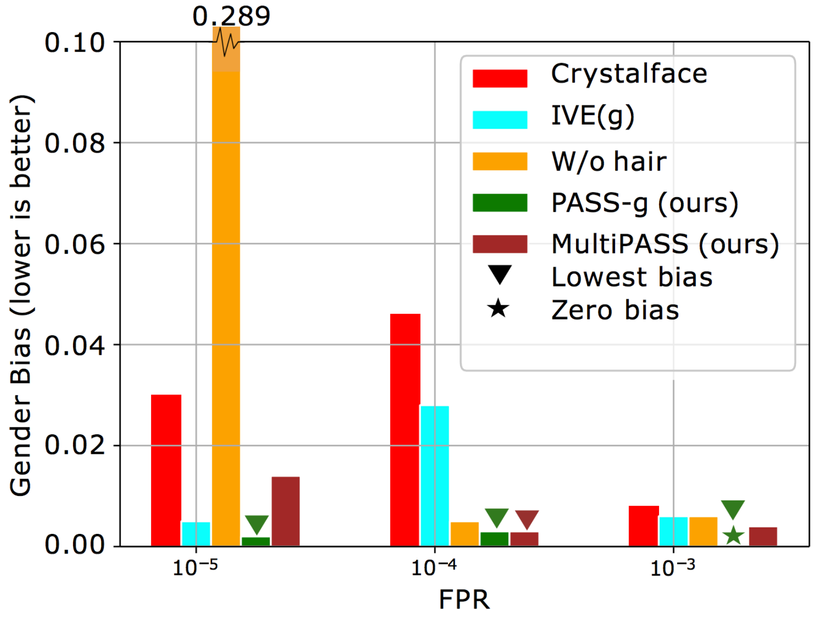


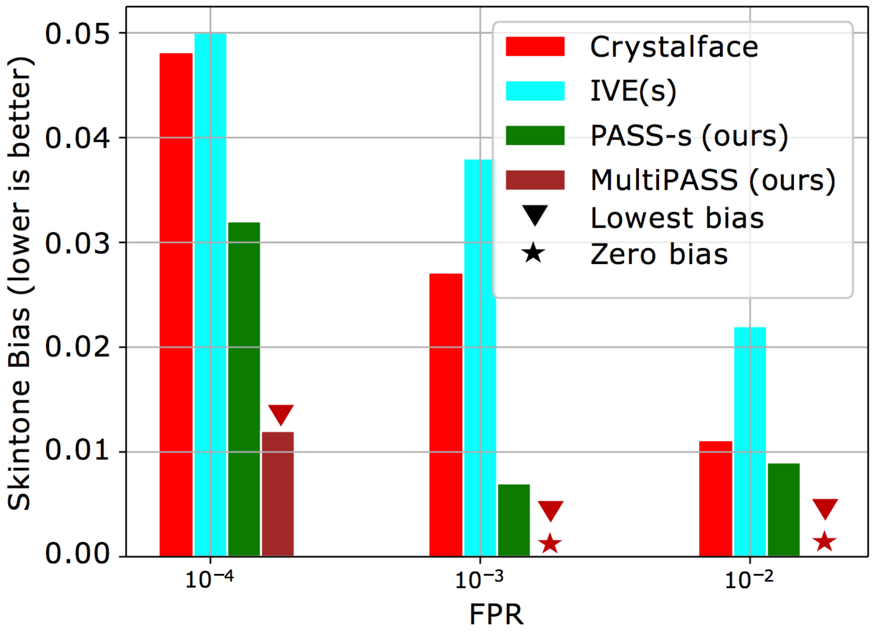
Genderbias



Skintone bias

# Results (Crystalface)





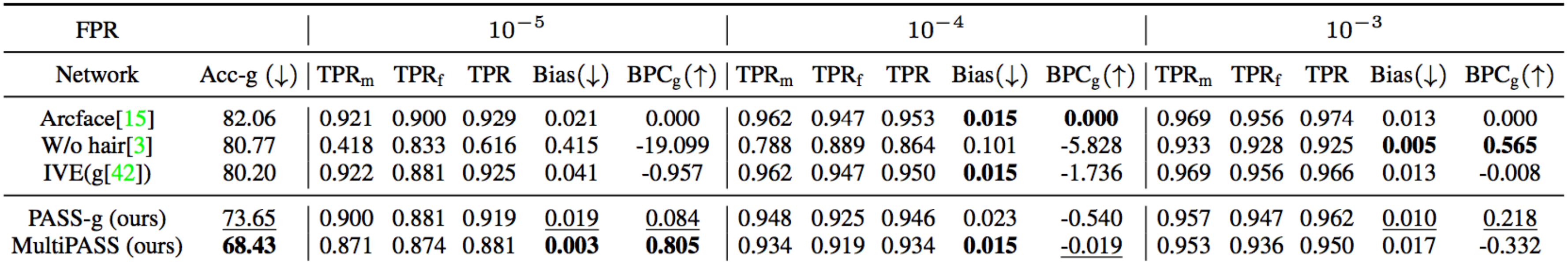
Genderbias

Skintone bias

Crystalface: R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J.C.Chen, C. D. Castillo, and R. Chellappa, “A Fast and Accurate System for Face Detection, Identification, and Verification”, IEEE T-BIOM, vol. 1, pp. 82-96, April 2019.

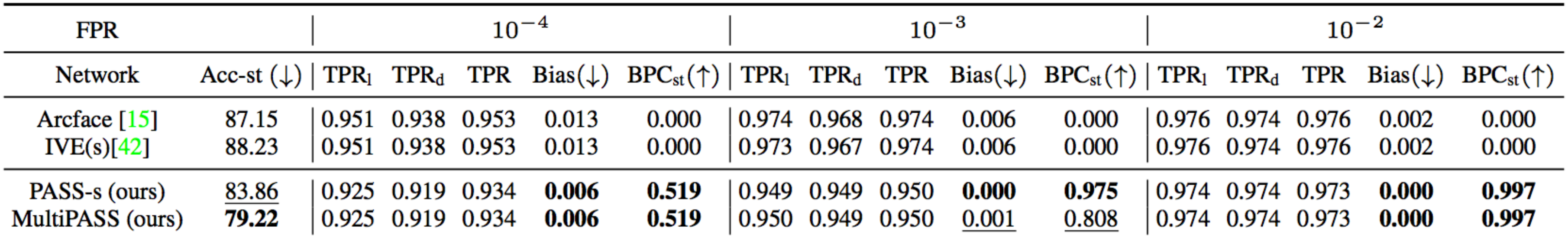
# PASS/MultiPASS Systems Achieve High BPCs

## Arcface – Gender bias analysis



| FPR | | $10^{-5}$ | | | | | $10^{-4}$ | | | | | $10^{-3}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Network | Acc-g (↓) | $\text{TPR}_\text{m}$ | $\text{TPR}_\text{f}$ | TPR | Bias(↓) | $\text{BPC}_\text{g}$(↑) | $\text{TPR}_\text{m}$ | $\text{TPR}_\text{f}$ | TPR | Bias(↓) | $\text{BPC}_\text{g}$(↑) | $\text{TPR}_\text{m}$ | $\text{TPR}_\text{f}$ | TPR | Bias(↓) | $\text{BPC}_\text{g}$(↑) |
| Arcface[15] | 82.06 | 0.921 | 0.900 | 0.929 | 0.021 | 0.000 | 0.962 | 0.947 | 0.953 | **0.015** | **0.000** | 0.969 | 0.956 | 0.974 | 0.013 | 0.000 |
| W/o hair[3] | 80.77 | 0.418 | 0.833 | 0.616 | 0.415 | -19.099 | 0.788 | 0.889 | 0.864 | 0.101 | -5.828 | 0.933 | 0.928 | 0.925 | **0.005** | **0.565** |
| IVE(g[42]) | 80.20 | 0.922 | 0.881 | 0.925 | 0.041 | -0.957 | 0.962 | 0.947 | 0.950 | **0.015** | -1.736 | 0.969 | 0.956 | 0.966 | 0.013 | -0.008 |
| PASS-g (ours) | 73.65 | 0.900 | 0.881 | 0.919 | 0.019 | 0.084 | 0.948 | 0.925 | 0.946 | 0.023 | -0.540 | 0.957 | 0.947 | 0.962 | 0.010 | 0.218 |
| MultiPASS (ours) | **68.43** | 0.871 | 0.874 | 0.881 | **0.003** | **0.805** | 0.934 | 0.919 | 0.934 | **0.015** | -0.019 | 0.953 | 0.936 | 0.950 | 0.017 | -0.332 |

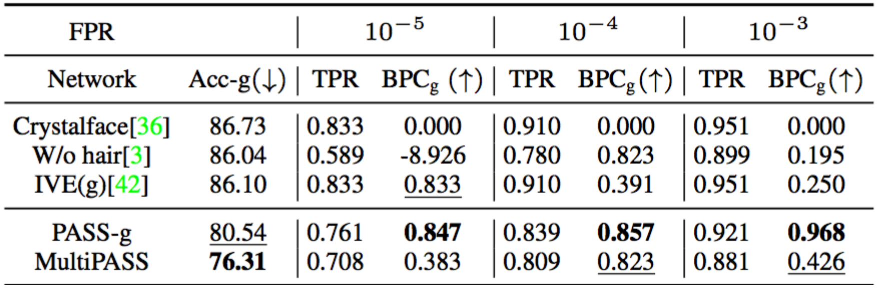
## Arcface - Skin tone bias analysis



| FPR | | $10^{-4}$ | | | | | $10^{-3}$ | | | | | $10^{-2}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Network | Acc-st (↓) | $\text{TPR}_\text{l}$ | $\text{TPR}_\text{d}$ | TPR | Bias(↓) | $\text{BPC}_\text{st}$(↑) | $\text{TPR}_\text{l}$ | $\text{TPR}_\text{d}$ | TPR | Bias(↓) | $\text{BPC}_\text{st}$(↑) | $\text{TPR}_\text{l}$ | $\text{TPR}_\text{d}$ | TPR | Bias(↓) | $\text{BPC}_\text{st}$(↑) |
| Arcface [15] | 87.15 | 0.951 | 0.938 | 0.953 | 0.013 | 0.000 | 0.974 | 0.968 | 0.974 | 0.006 | 0.000 | 0.976 | 0.974 | 0.976 | 0.002 | 0.000 |
| IVE(s)[42] | 88.23 | 0.951 | 0.938 | 0.953 | 0.013 | 0.000 | 0.973 | 0.967 | 0.974 | 0.006 | 0.000 | 0.976 | 0.974 | 0.976 | 0.002 | 0.000 |
| PASS-s (ours) | 83.86 | 0.925 | 0.919 | 0.934 | **0.006** | **0.519** | 0.949 | 0.949 | 0.950 | **0.000** | **0.975** | 0.974 | 0.974 | 0.973 | **0.000** | **0.997** |
| MultiPASS (ours) | **79.22** | 0.925 | 0.919 | 0.934 | **0.006** | **0.519** | 0.950 | 0.949 | 0.950 | 0.001 | 0.808 | 0.974 | 0.974 | 0.973 | **0.000** | **0.997** |

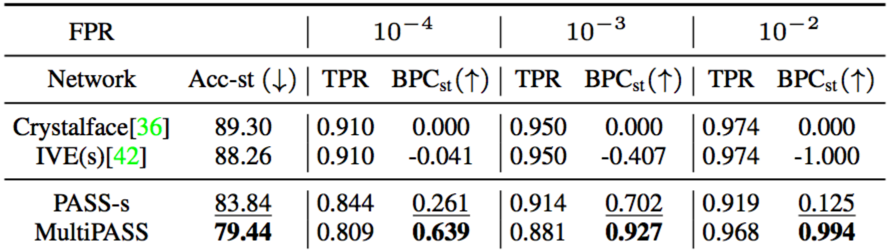
# PASS/MultiPASS Systems Achieve High BPCs

## Crystalface – Gender bias analysis



| FPR | | $10^{-5}$ | | $10^{-4}$ | | $10^{-3}$ | |
|---|---|---|---|---|---|---|---|
| Network | Acc-g(↓) | TPR | $\text{BPC}_g$ (↑) | TPR | $\text{BPC}_g$(↑) | TPR | $\text{BPC}_g$(↑) |
| Crystalface[36] | 86.73 | 0.833 | 0.000 | 0.910 | 0.000 | 0.951 | 0.000 |
| W/o hair[3] | 86.04 | 0.589 | -8.926 | 0.780 | 0.823 | 0.899 | 0.195 |
| IVE(g)[42] | 86.10 | 0.833 | 0.833 | 0.910 | 0.391 | 0.951 | 0.250 |
| PASS-g | 80.54 | 0.761 | **0.847** | 0.839 | **0.857** | 0.921 | **0.968** |
| MultiPASS | **76.31** | 0.708 | 0.383 | 0.809 | 0.823 | 0.881 | 0.426 |

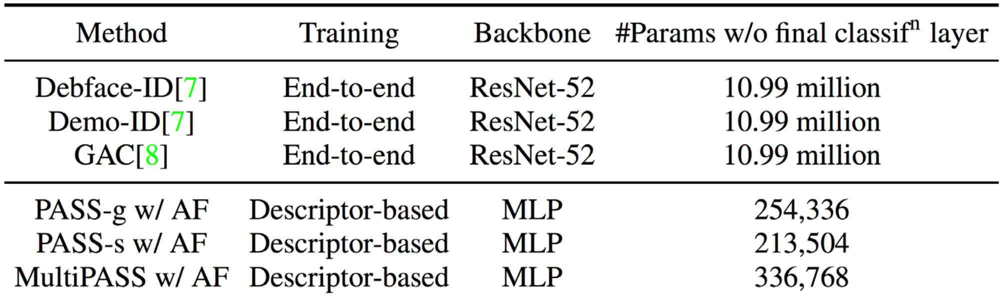
## Crystalface – Skin tone bias analysis



| FPR | | $10^{-4}$ | | $10^{-3}$ | | $10^{-2}$ | |
|---|---|---|---|---|---|---|---|
| Network | Acc-st (↓) | TPR | $\text{BPC}_{st}$(↑) | TPR | $\text{BPC}_{st}$(↑) | TPR | $\text{BPC}_{st}$(↑) |
| Crystalface[36] | 89.30 | 0.910 | 0.000 | 0.950 | 0.000 | 0.974 | 0.000 |
| IVE(s)[42] | 88.26 | 0.910 | -0.041 | 0.950 | -0.407 | 0.974 | -1.000 |
| PASS-s | 83.84 | 0.844 | 0.261 | 0.914 | 0.702 | 0.919 | 0.125 |
| MultiPASS | **79.44** | 0.809 | **0.639** | 0.881 | **0.927** | 0.968 | **0.994** |

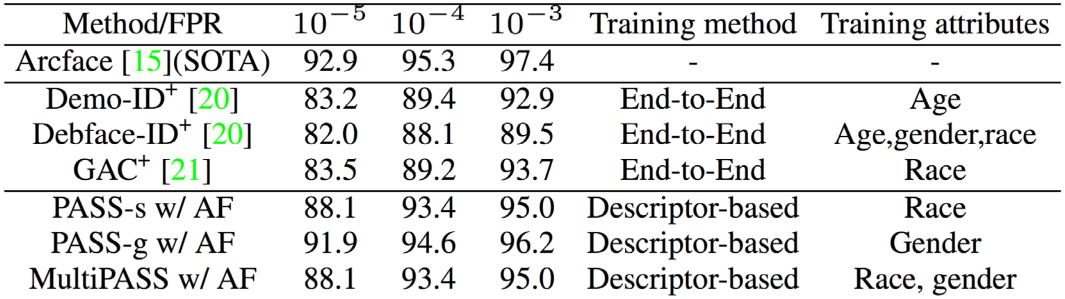
# End-to-end Systems v/s PASS

- PASS-based systems require fewer parameters than their end-to-end counterparts.



| Method | Training | Backbone | #Params w/o final classif$^{n}$ layer |
|---|---|---|---|
| Debface-ID[7] | End-to-end | ResNet-52 | 10.99 million |
| Demo-ID[7] | End-to-end | ResNet-52 | 10.99 million |
| GAC[8] | End-to-end | ResNet-52 | 10.99 million |
| PASS-g w/ AF | Descriptor-based | MLP | 254,336 |
| PASS-s w/ AF | Descriptor-based | MLP | 213,504 |
| MultiPASS w/ AF | Descriptor-based | MLP | 336,768 |

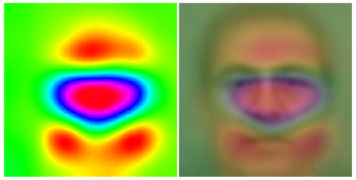
- With PASS, we have the freedom to start with SOTA face descriptors. So, the verification performance obtained by PASS is closer to SOTA.



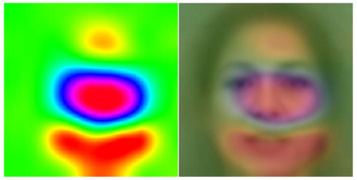
| Method/FPR | $10^{-5}$ | $10^{-4}$ | $10^{-3}$ | Training method | Training attributes |
|---|---|---|---|---|---|
| Arcface [15](SOTA) | 92.9 | 95.3 | 97.4 | - | - |
| Demo-ID$^{+}$ [20] | 83.2 | 89.4 | 92.9 | End-to-End | Age |
| Debface-ID$^{+}$ [20] | 82.0 | 88.1 | 89.5 | End-to-End | Age,gender,race |
| GAC$^{+}$ [21] | 83.5 | 89.2 | 93.7 | End-to-End | Race |
| PASS-s w/ AF | 88.1 | 93.4 | 95.0 | Descriptor-based | Race |
| PASS-g w/ AF | 91.9 | 94.6 | 96.2 | Descriptor-based | Gender |
| MultiPASS w/ AF | 88.1 | 93.4 | 95.0 | Descriptor-based | Race, gender |

# FR Networks Attend to Different Spatial Regions, Depending on Demographic Groups

Dissimilar attention regions for male and female

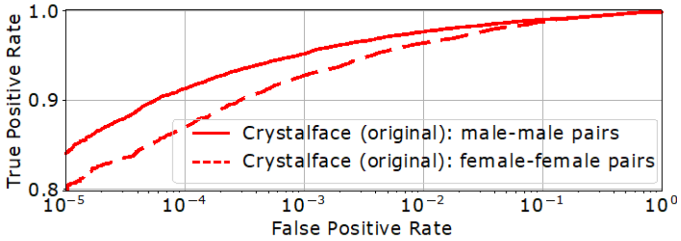


Average attention map :
**Male**

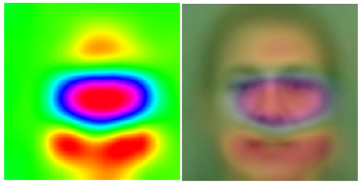


Average attention map :
**Female**

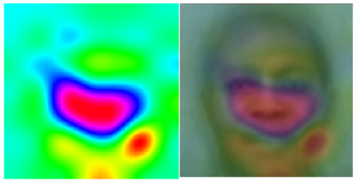
Gender bias in face verification



Dissimilar attention regions for dark and light skintone

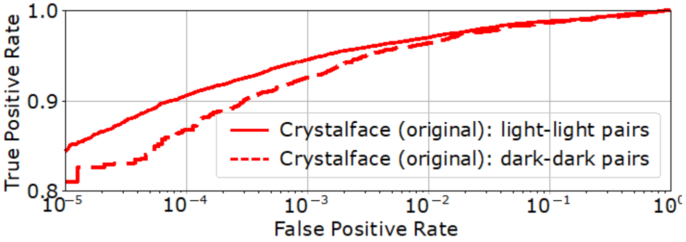


Average attention map :
**Dark skintone**



Average attention map :
**Light skintone**

Skintone bias in face verification



# Summary -1

- Learning from degraded data is hard.
- Statistics literature has considered the “errors in variable” formulation
  - Insufficient to deal with blur, deformations and turbulence
- Without incorporating physics-based models of deformation, performance will not improve.
- In the absence of physics, demand on annotated data will be huge.

# Summary -2

- Face recognition systems demonstrate gender and skintone bias.
- Simply balancing the training dataset does not help.
- Face recognition networks implicitly encode sensitive attributes like gender, age etc., without being trained to do so.
- Adversarially removing sensitive attributes is an interesting line of research, that can potentially reduce bias.
- End-to-end adversarial de-biasing systems reduce bias but achieve much lower verification performance, compared to SOTA.
- PASS can reduce gender and skintone bias while achieving SOTA verification performance
- For latest evaluation results see
  - https://pages.nist.gov/frvt/reports/11/frvt_11_report.pdf
  - https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf

# Publications - 1

- C.P.Lau, C. D. Castillo, and R. Chellappa, “ATFaceGAN: Single Face Semantic Aware Image Restoration and Recognition from Atmospheric Turbulence”, IEEE Trans. on Biometrics, Behaviors and Identity Science, vol. 3, pp. 240-251, April 2021.
- B. Lu, J.C. Chen and R. Chellappa, “UID-GAN: Unsupervised Image Deblurring via Disentangled Representations”, IEEE Transactions on Biometrics, Behavior and Identity Science, vol. 2, pp. 26-39, Jan. 2020.
- Zheng, R. Ranjan, C. H. Chen, J. C. Chen, C. D. Castillo, and R. Chellappa “An Automatic System for Unconstrained Video-based Face Recognition”, IEEE Transactions on Biometrics, Behavior and Identity Science, vol. 2, pp. 194 – 209, July 2020.
- R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J.C.Chen, C. D. Castillo, and R. Chellappa, “A Fast and Accurate System for Face Detection, Identification, and Verification”, IEEE Trans. on Biometrics, Behavior and Identity Science, vol. 1, pp. 82-96, April 2019.
- M. Singh, R. Singh, M. Vatsa, N. Ratha and R. Chellappa, “Recognizing Disguised Faces in the Wild”, IEEE Trans. on Biometrics, Behavior and Identity Science, vol. 1, pp. 97-108, April 2019.

# Publications-2

- R. Ranjan, et al., "Deep Learning for Understanding Faces", IEEE Signal Processing Magazine, vol. 35, pp. 66-83, Jan. 2018.
- R. Ranjan, V.M. Patel and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition", IEEE Trans. Patt. Anal. and Mach. Intelligence, vol. 41, pp. 121-135, Jan. 2018.
- J.C. Chen, R. Ranjan, S. Sankaranarayanan, A. Kumar, C.H. Chen, V.M. Patel, C. Castillo and R. Chellappa, "Unconstrained Still/Video-Based Face Verification with Deep Convolutional Neural Networks", International Jl. of Computer Vision, vol. 126, pp. 272-291, 2018.
- P. J, Phillips, et al., "Face Recognition at its Best: Forensic Examiners, Super-recognizers, and Algorithms", Proc. National Academy of Sciences, vol. 115, May 2018.
- A.J. O'Toole, C.D. Castillo, C.J. Parde, M.Q. Hill, and R. Chellappa, "Face Space Representations in Deep Convolutional Neural Networks", *Trends in Cognitive Sciences,* vol. 22, Sept. 2018, Pages 794-809

# Publications-3

- Y.C. Chen, V.M Patel, P.J. Phillips and R. Chellappa, “Dictionary-based Face and Person Recognition from Unconstrained Video”, IEEE Access: Special Issue on 4D’s of Machine Learning for Biometrics: Deep Learning, Dictionary Learning, Domain Adaptation, and Distance Metric Learning, vol. 3, pp. 1783 - 1798, Oct. 2015.
- M. Du, A. Sankaranarayanan and R. Chellappa,” Robust Face Recognition from Multi-View Videos”, IEEE Trans. on Image Processing, vol. 23, pp. 1105-1107, March 2014.
- V. M. Patel, Y. C. Chen, R. Chellappa, and P. J. Phillips, “Dictionaries for Image and Video-based Face Recognition” Invited Paper , 30th Anniversary Issue, Jl. Opt. Society of America, vol. 31, pp. 1090-1103, May 2014.
- H. T. Ho and R. Chellappa, “Pose-Invariant Face Recognition Using Markov Random Fields”, IEEE Transactions on Image Processing, vol. 22, pp. 1573-1584, April 2013.
- W. Zou, P.C. Yuen, and R. Chellappa, “A Low-Resolution Face Tracker Robust to Illumination Variations”, IEEE Trans. on Image Processing, vol. 22, pp. 1726-1739, May 2013.
- P. Vageeswaran, K. Mitra and R. Chellappa, “Blur and Illumination Robust Face Recognition via Set-Theoretic Characterization”, IEEE Trans. on Image Processing, vol. 22, pp. 1362-1372, April 2013.