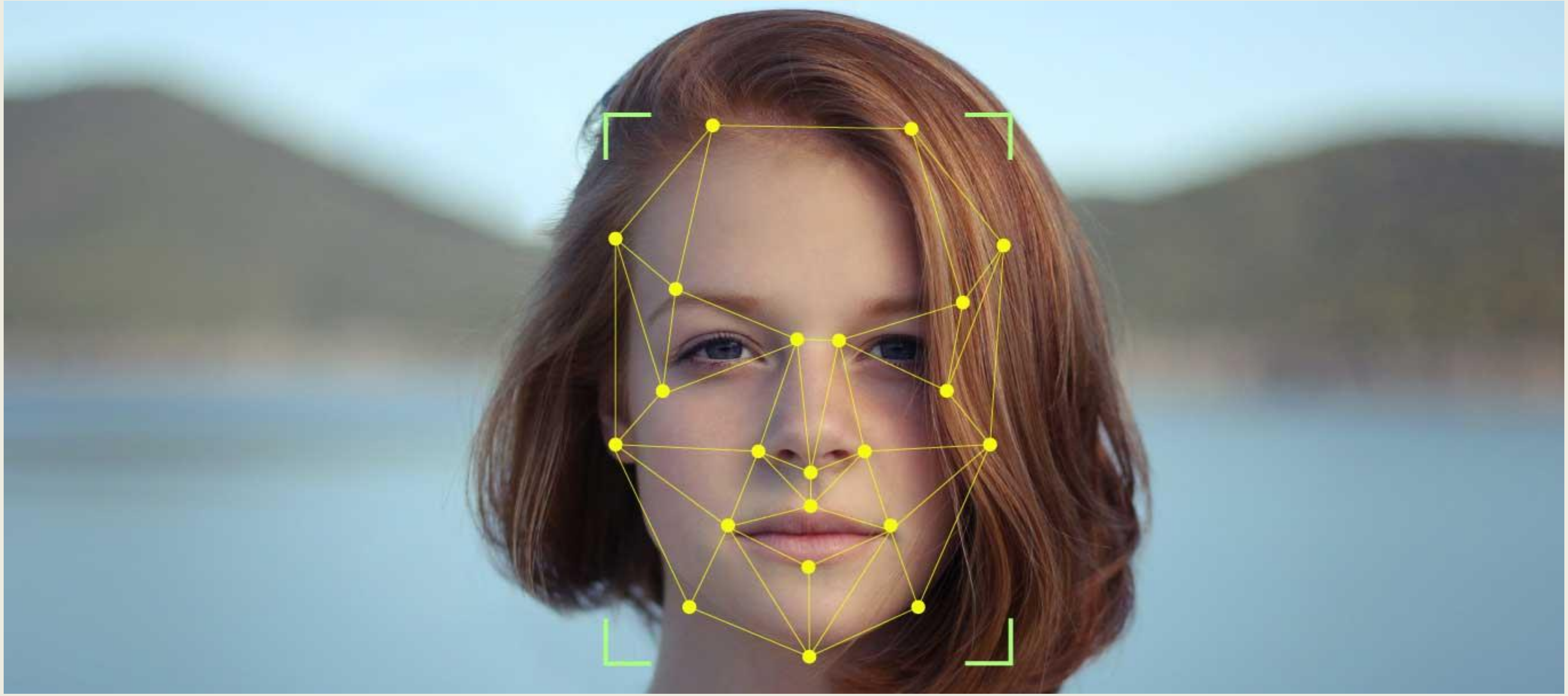
Face Recognition



Based on material from Ruben Tolosana

BiDA Lab
Biometrics & Data Pattern Analytics Lab

UAM
Universidad Autónoma de Madrid

What is it?

Unknown Subject
(image from CCTV,
Passport, etc.)

BiDALab-UAM ©

What is it?

Labelled Face Database (million of people?)

Unknown Subject
(image from CCTV,
Passport, etc.)

Automatic Face
Recognition

What is it?

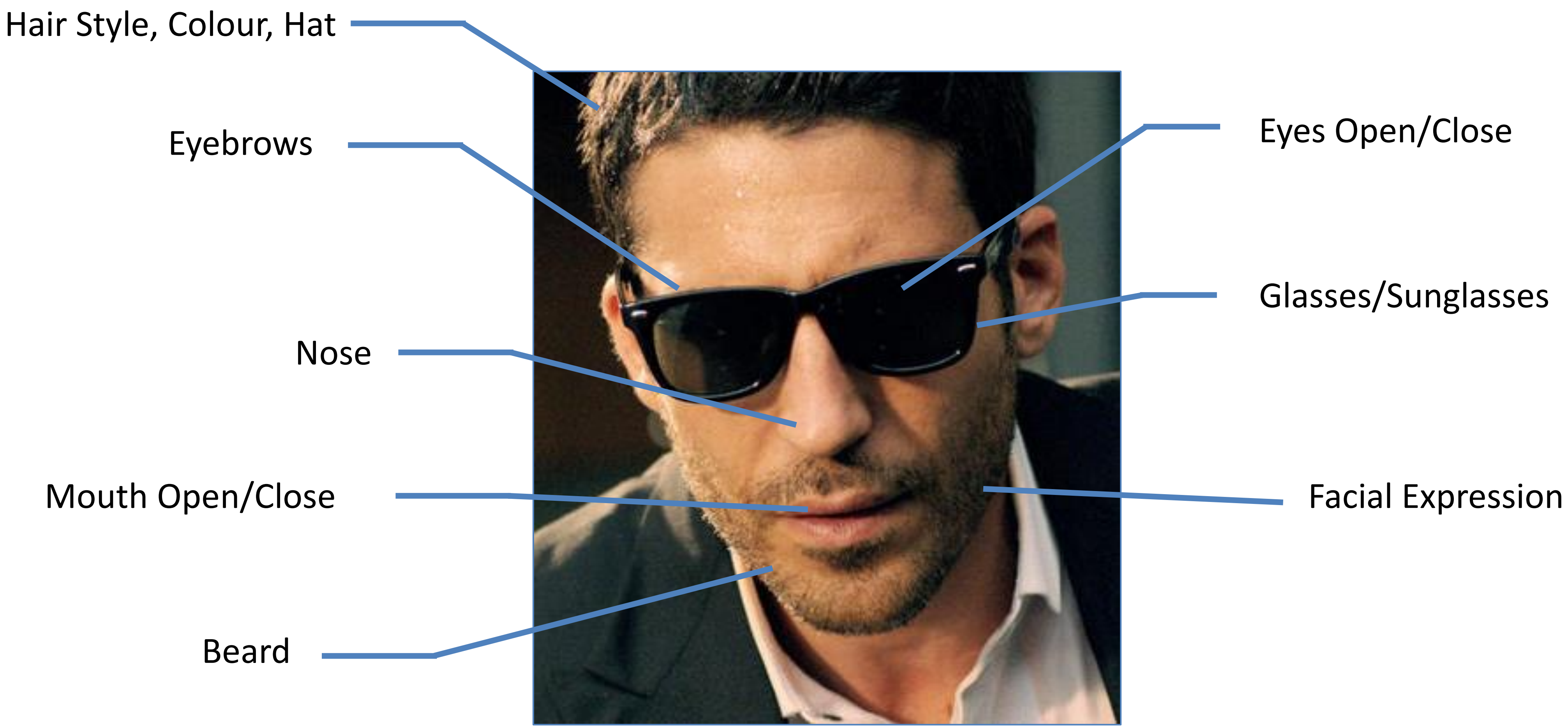
Labelled Face Database (million of people?)

Unknown Subject (image from CCTV, Passport, etc.)

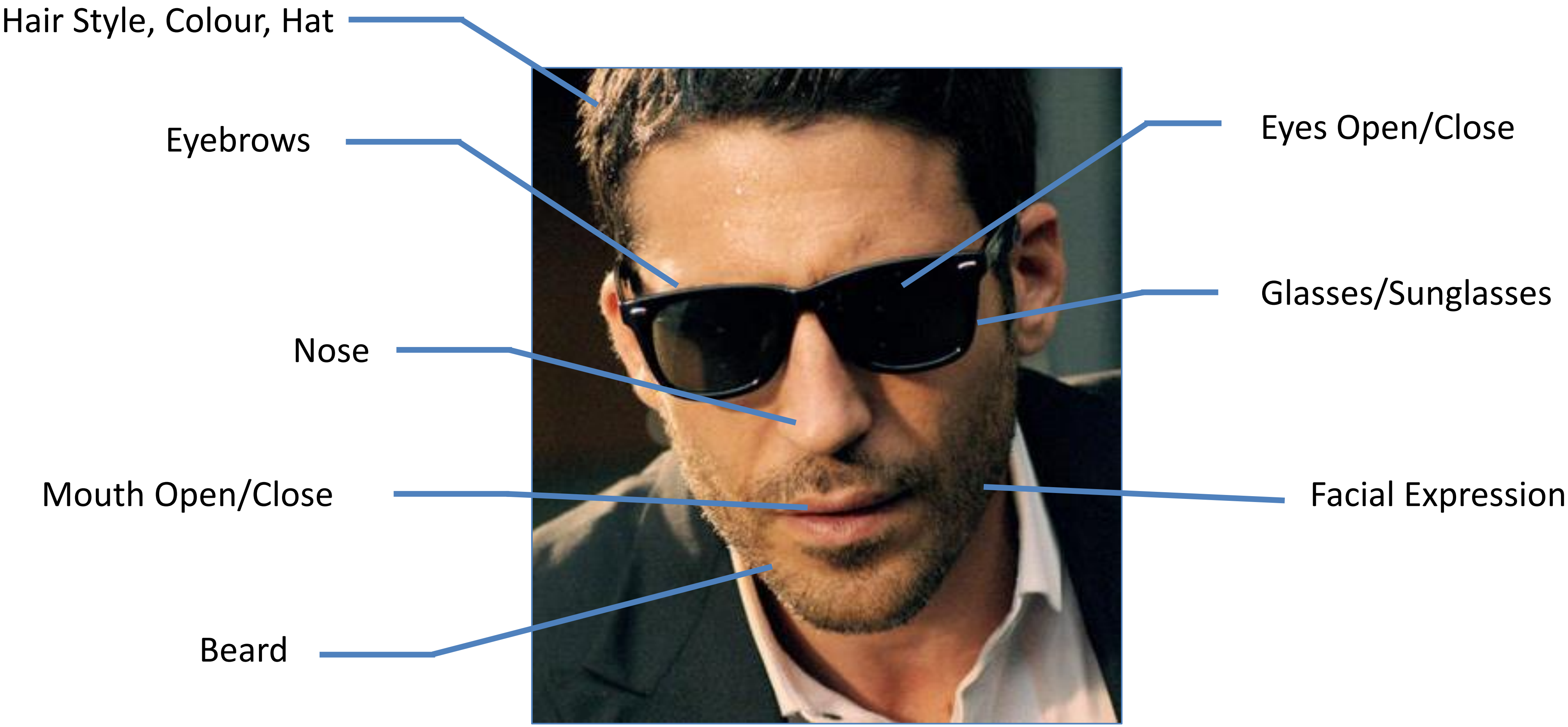
Automatic Face Recognition

Match

Is it challenging?

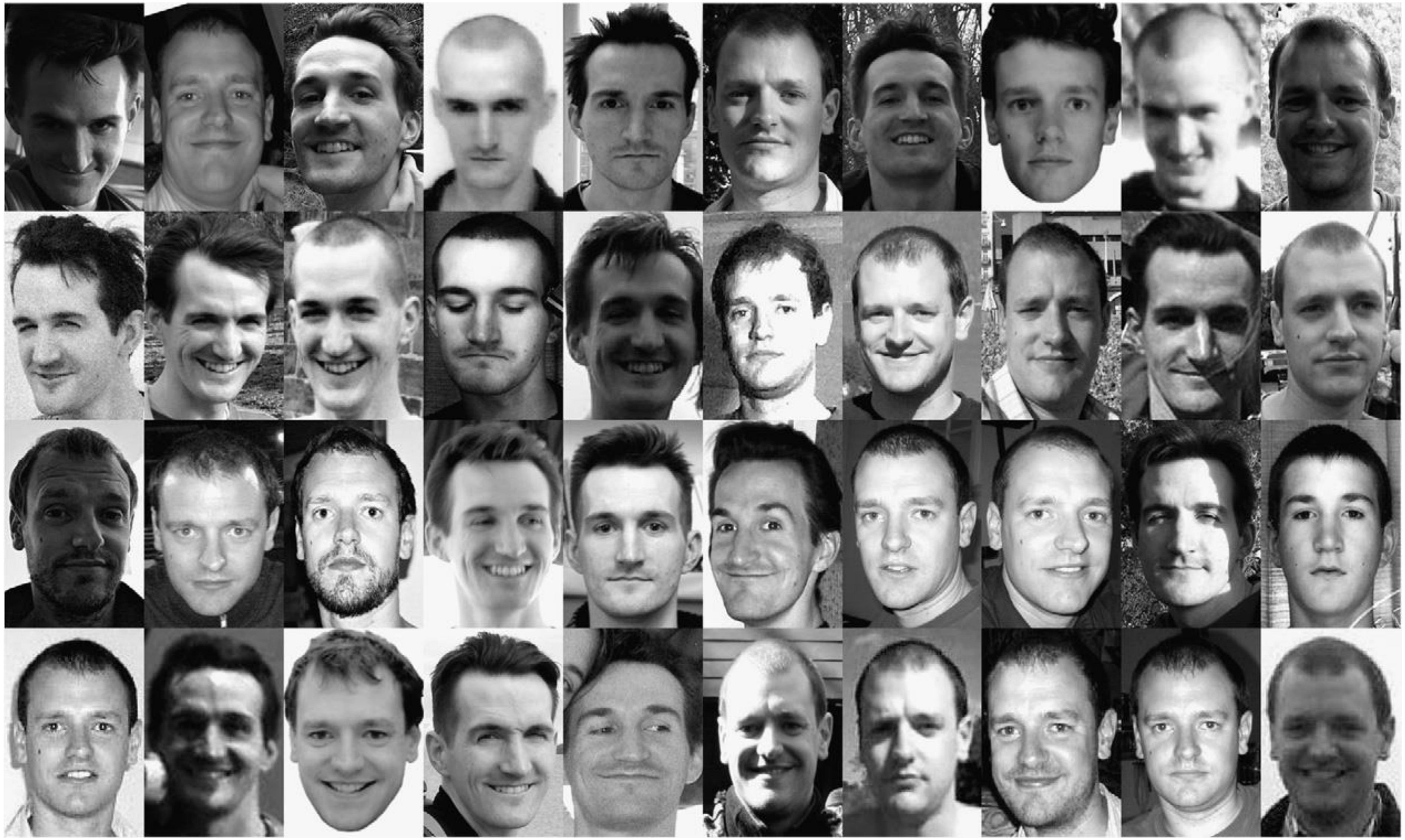


Is it challenging?



Most facial attributes can be changed!!

How many people are there?



- R. Jenkins, D. White, X. Van Montfort, A.M. Burton, “Variability in photos of the same face,” Cognition, 121(3), 313-323, 2011

BiDALab-UAM ©

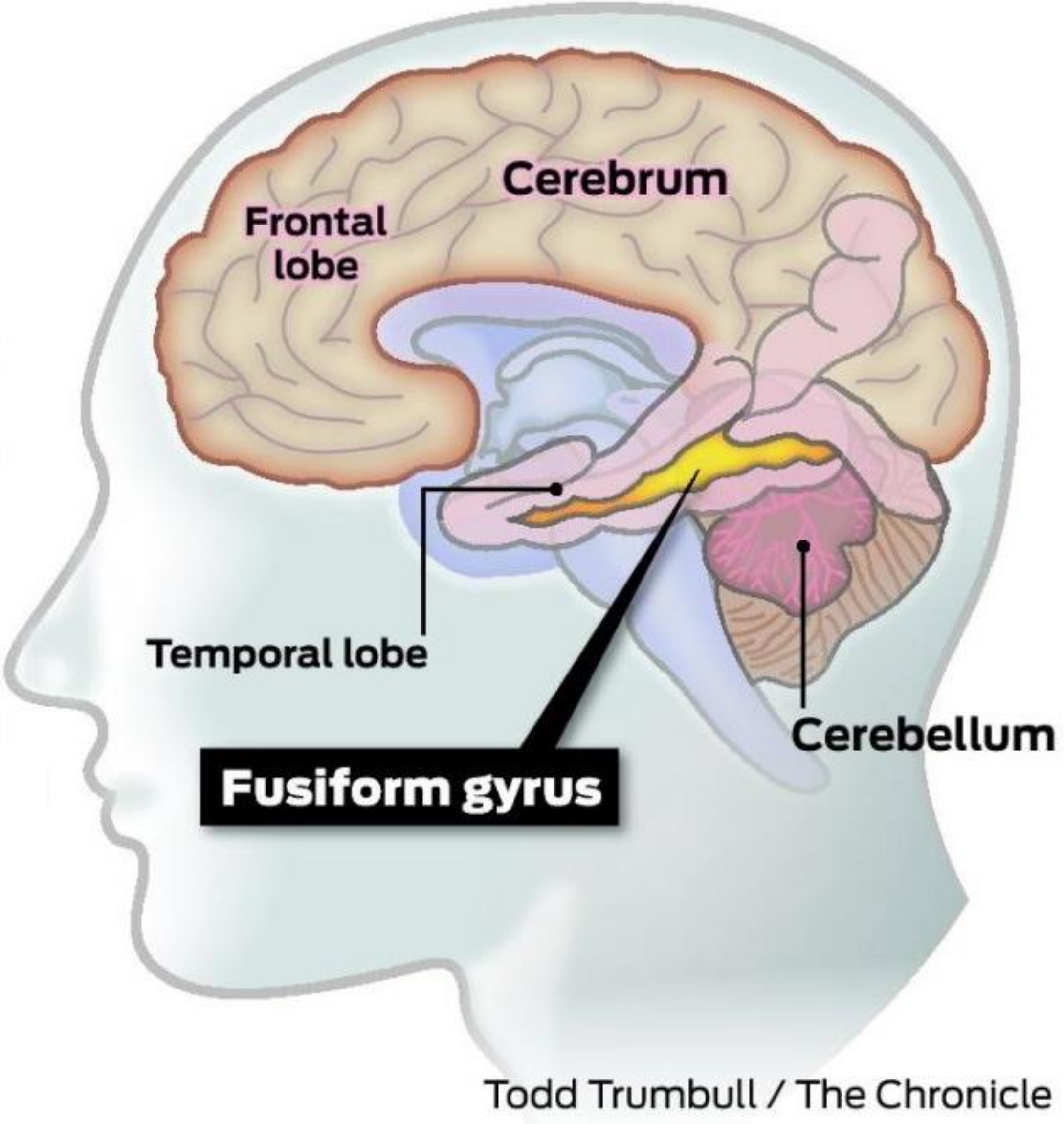
How many people are there?

• R. Jenkins, D. White, X. Van Montfort, A.M. Burton, "Variability in photos of the same face," Cognition, 121(3), 313-323, 2011

BiDALab-UAM ©

Face Recognition: How it works in our brain

- The temporal lobe of the brain is partly responsible for our ability to recognize faces.
- Some neurons in the temporal lobe (Fusiform girus) respond to particular features of faces.
- Some people who suffer damage to the temporal lobe lose their ability to recognize and identify familiar faces.



Face Recognition: How it works in our brain

- The temporal lobe of the brain is partly responsible for our ability to recognize faces.
- Some neurons in the temporal lobe (Fusiform girus) respond to particular features of faces.
- Some people who suffer damage to the temporal lobe lose their ability to recognize and identify familiar faces.

BiDALab-UAM ©

Face Recognition: How it works in our brain

- The temporal lobe of the brain is partly responsible for our ability to recognize faces.
- Some neurons in the temporal lobe (Fusiform girus) respond to particular features of faces.
- Some people who suffer damage to the temporal lobe lose their ability to recognize and identify familiar faces.

Face Recognition: How it works in our brain

- The temporal lobe of the brain is partly responsible for our ability to recognize faces.
- Some neurons in the temporal lobe (Fusiform girus) respond to particular features of faces.
- Some people who suffer damage to the temporal lobe lose their ability to recognize and identify familiar faces.

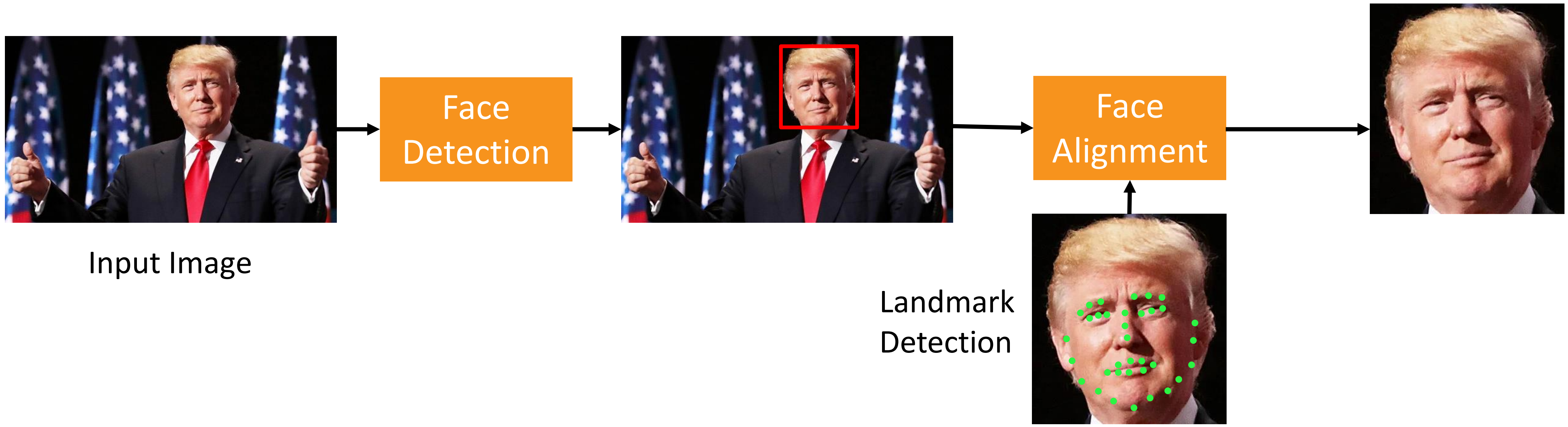
Automatic Face Recognition: Architecture

Input Image

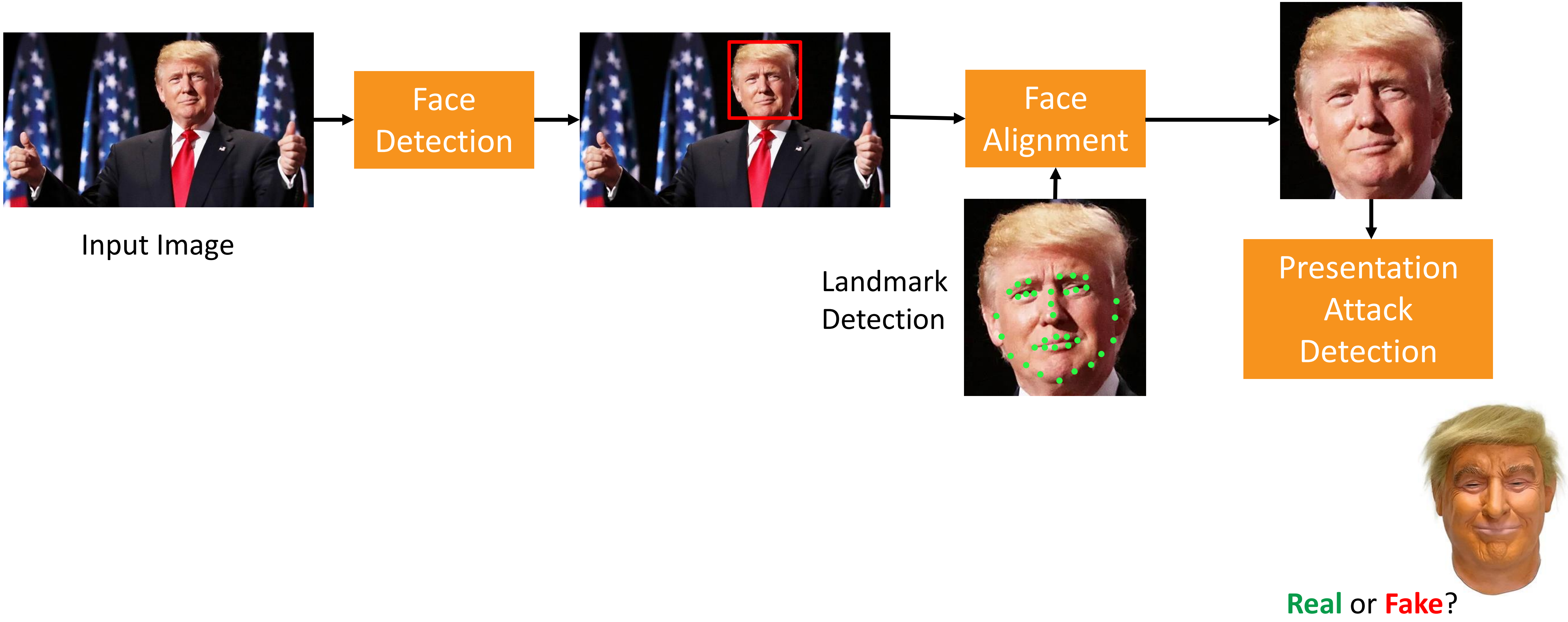
Automatic Face Recognition: Architecture

Input Image

Automatic Face Recognition: Architecture

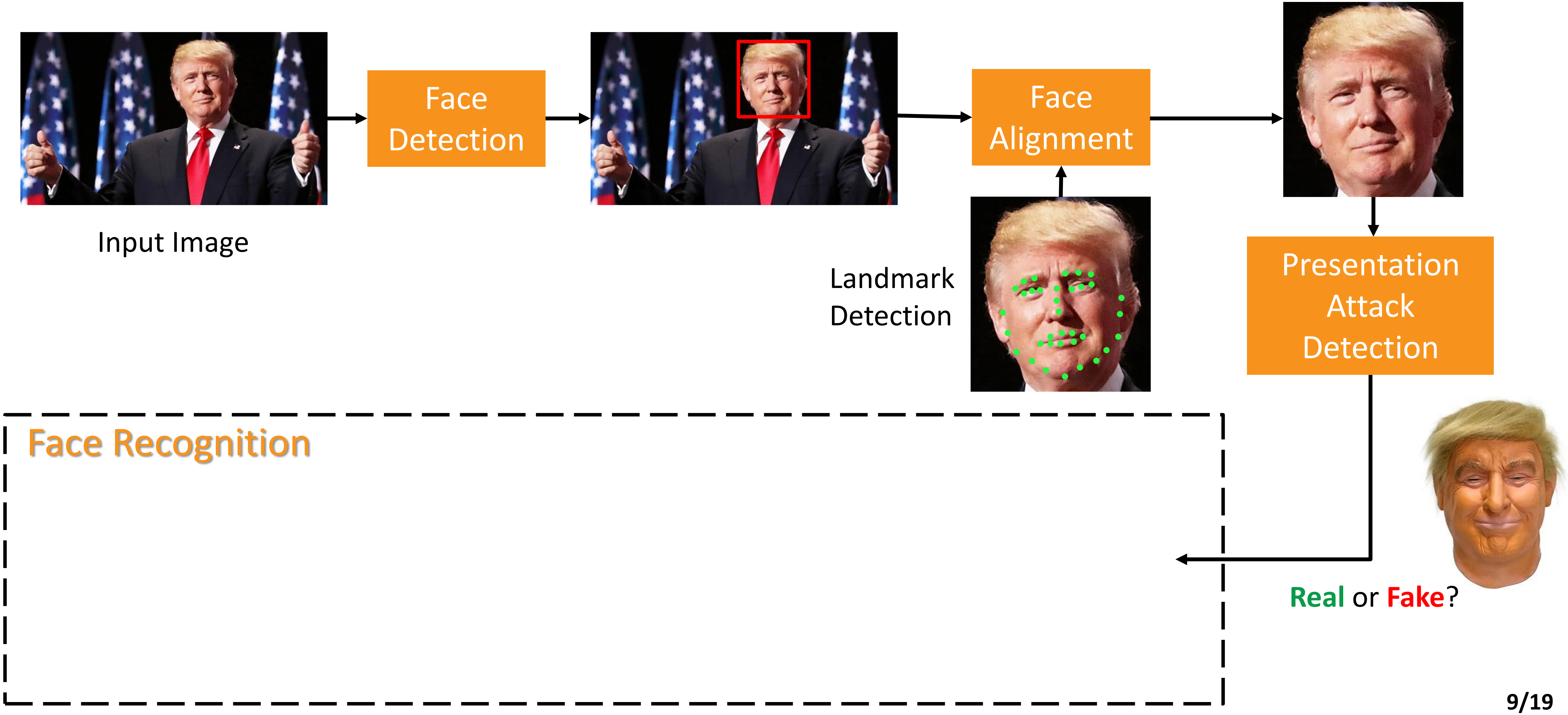


Automatic Face Recognition: Architecture

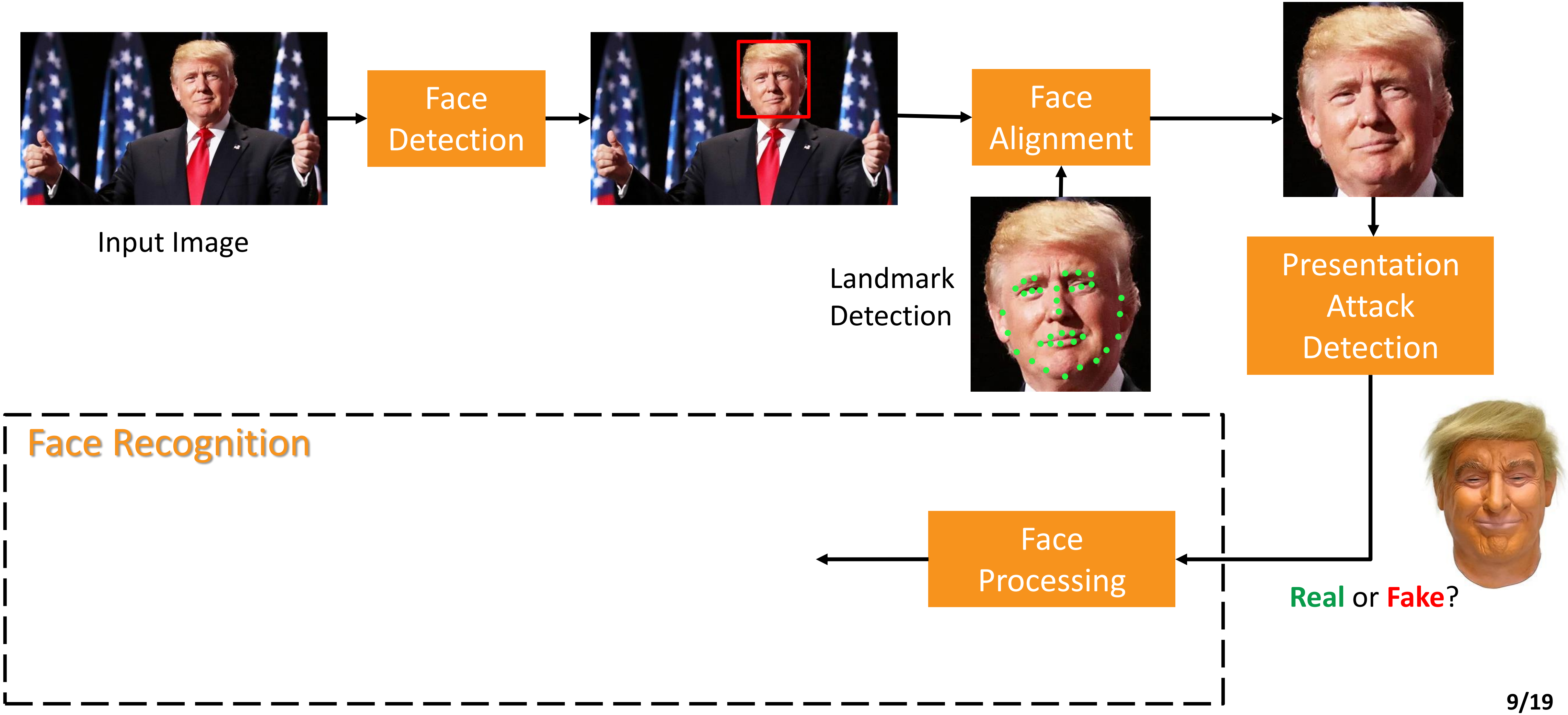


BiDALab-UAM ©

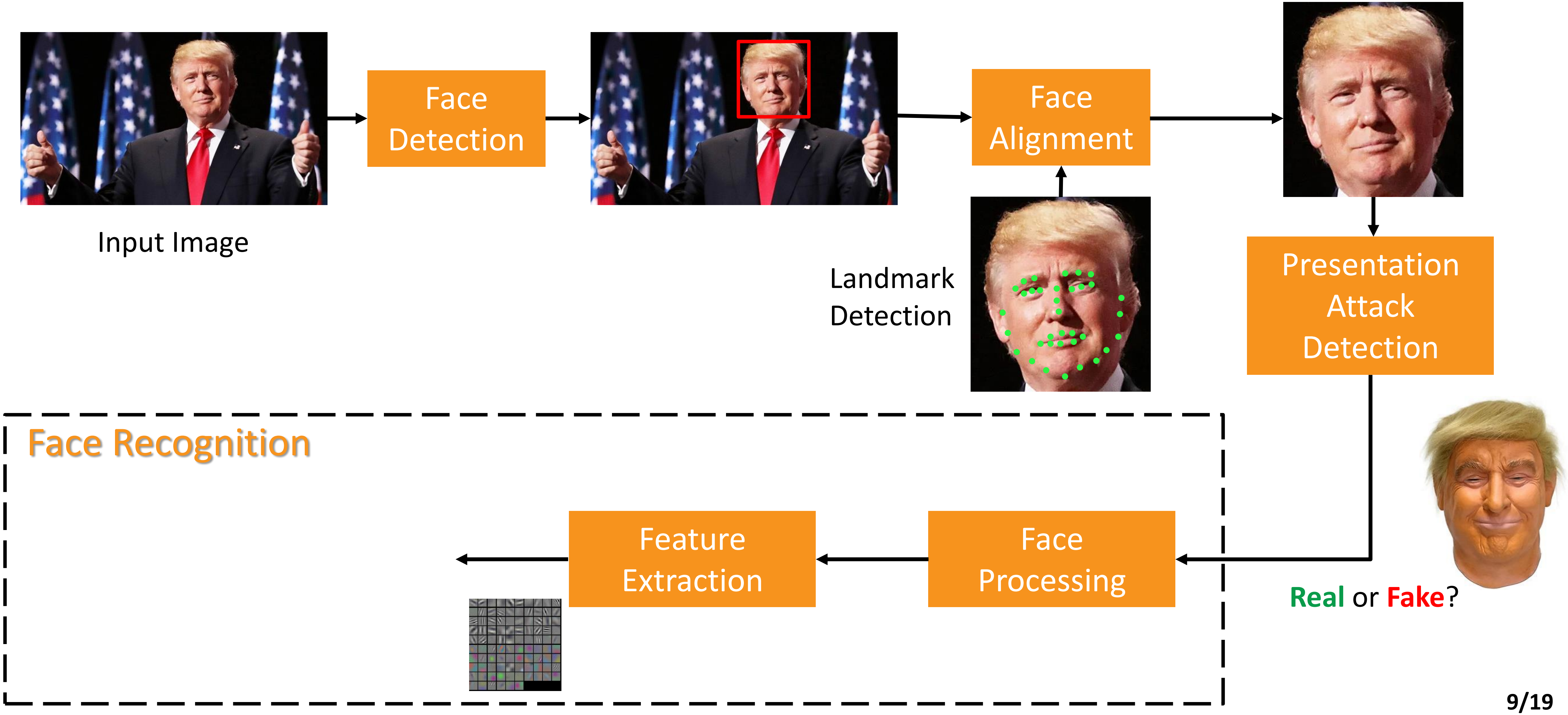
Automatic Face Recognition: Architecture



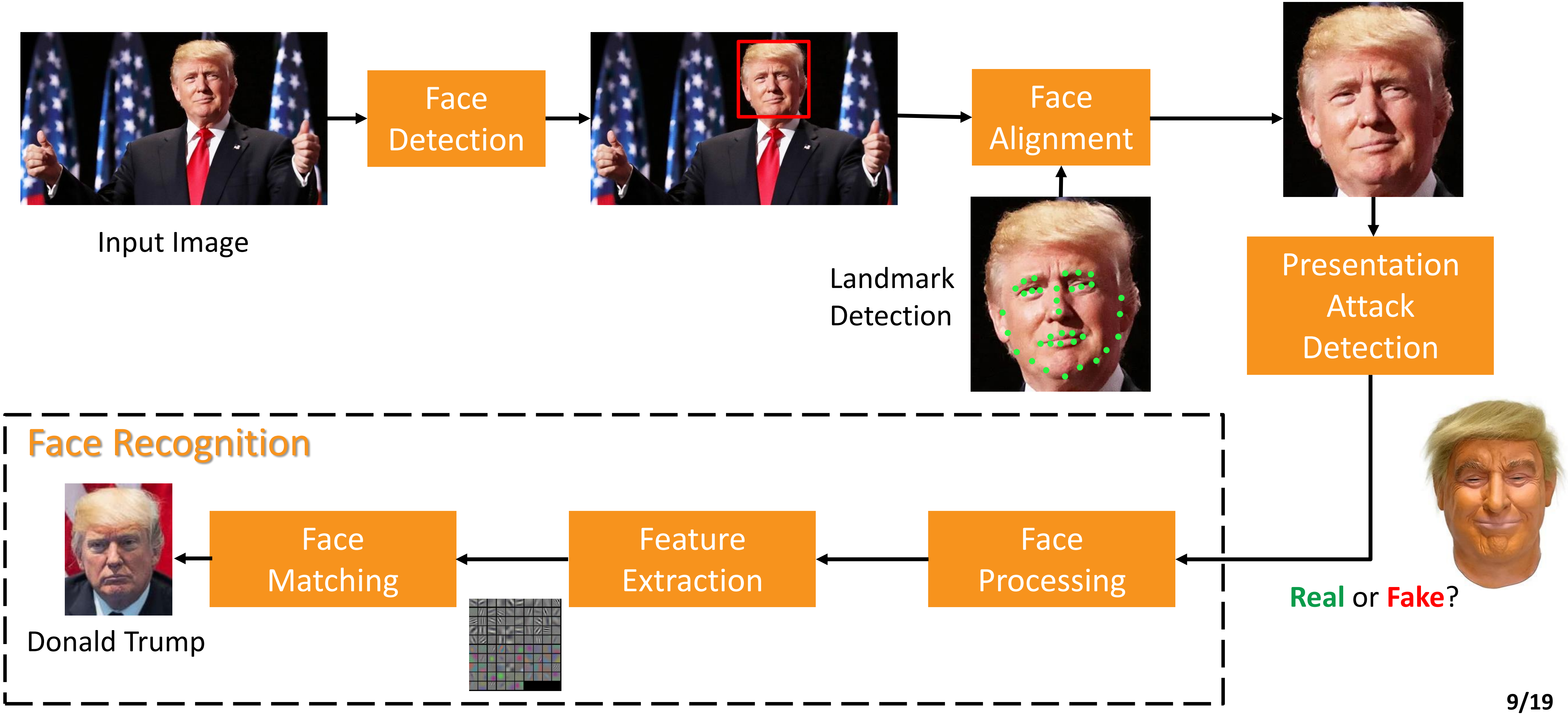
Automatic Face Recognition: Architecture



Automatic Face Recognition: Architecture



Automatic Face Recognition: Architecture



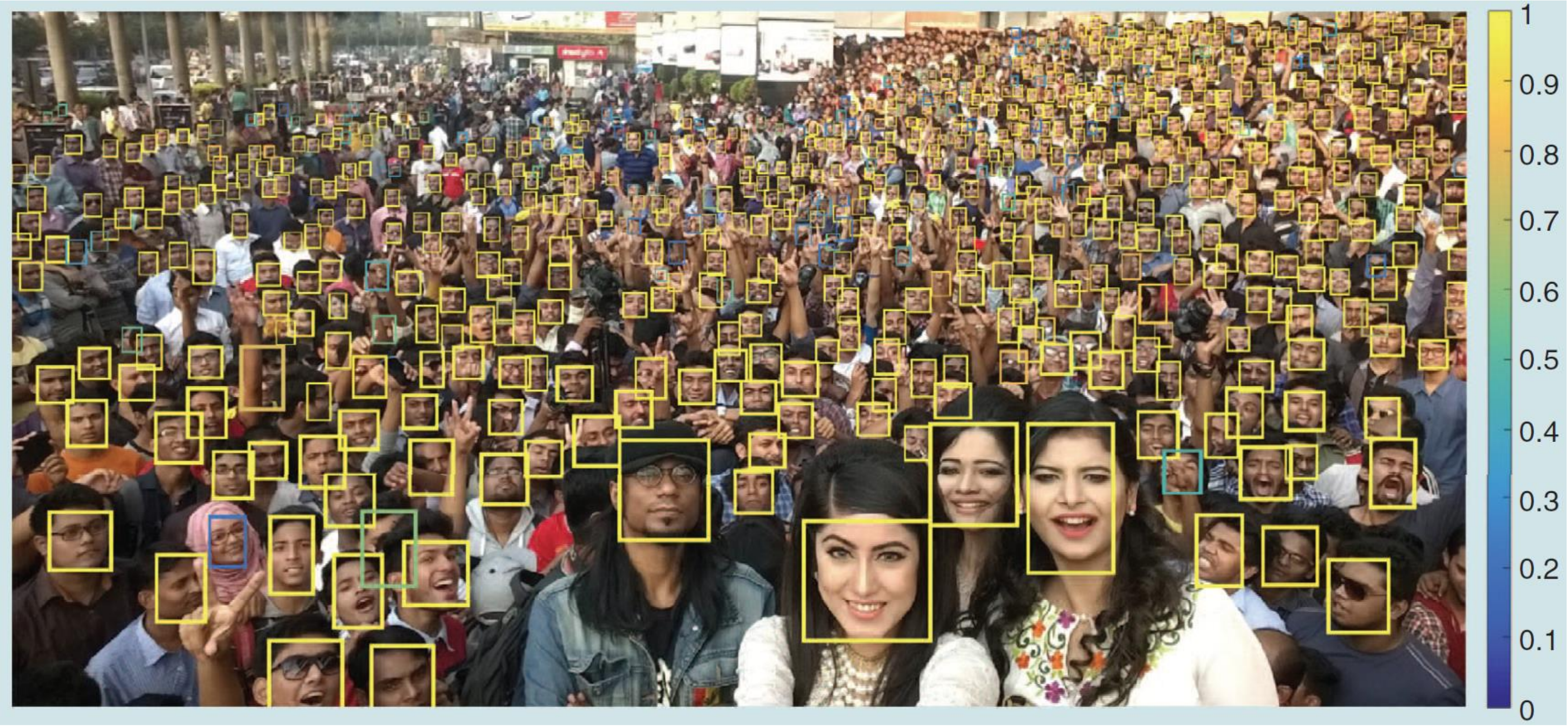
Automatic Face Recognition: Architecture

Input Image

Face Detection

Given an input image, a face detector needs to detect the face/s in the image and return bounding-box coordinates for each of them.

Challenges: different resolution, size, illumination, expression, skin color, occlusions, cosmetics…



Detector confidence is given by the color bar on the right

Face Detection

Popular databases:

- Wider Face Database.
- Face Detection Data Set and Benchmark (FDDB).
- MALF Database.

- S. Yang, P. Luo, C.-C. Loy, and X. Tang, “Wider face: A face detection benchmark,” in Proc. IEEE Conf. Computer Vision Pattern Recognition, 2016, pp. 5525–5533.
- V. Jain and E. Learned-Miller, “FDDB: A benchmark for face detection in unconstrained settings,” Tech. Rep. vol. 88, Univ. Massachusetts, Amherst, 2010.
- B. Yang, J. Yan, Z. Lei, and, S. Z. Li, “Fine-grained evaluation on face detection in the wild,” in, Proc. 11th IEEE Int. Conf. Workshops Automatic Face and Gesture Recognition, vol. 1, 2015, pp. 1–7.

BiDALab-UAM ©

Face Detection

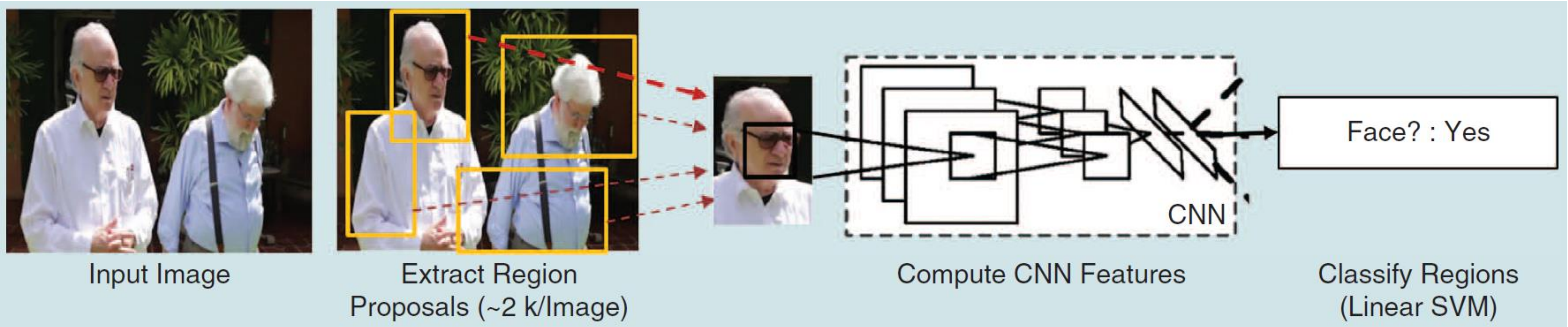
Region-Based Approach: it generates a pool of generic object-proposals (approximately 2,000 per image) and a DCNN is used to classify whether or not a given proposal contains a face.

Cons:

- Difficult faces are hard to capture in any object proposal.
- High computation time.

Popular Approaches:

- Faster R-CNN (with a 3D mean face model)...



- H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," in, Proc. 12th IEEE Int. Conf. Workshops Automatic Face and Gesture Recognition, 2017
- Y. Li, B. Sun, T. Wu, and Y. Wang, "Face detection with end-to-end integration of a convnet and a 3D model," in Proc. European Conf. Computer Vision, 2016

BiDALab-UAM ©

Face Detection

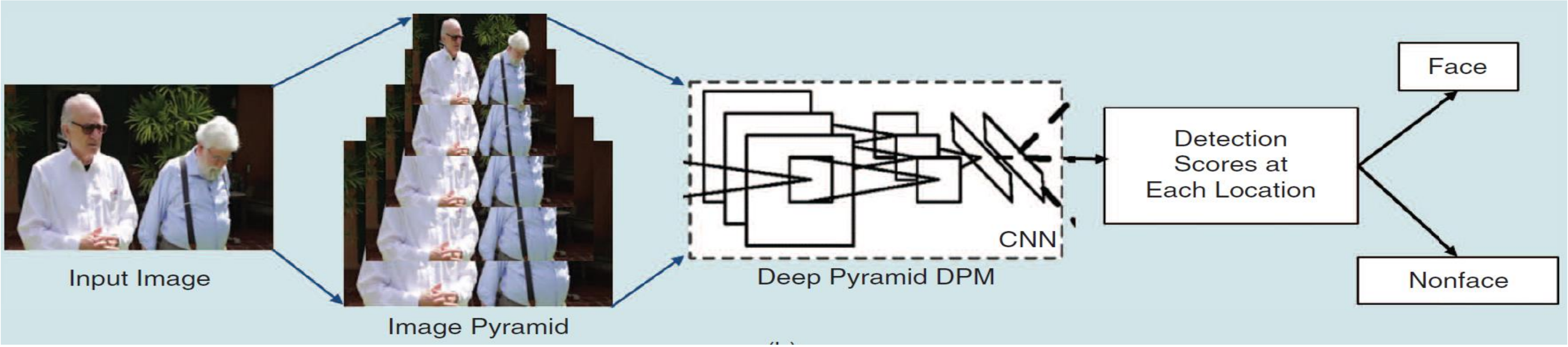
Sliding-Window Approach: it computes a face detection score and bounding-box coordinates at every location in a feature map at a given scale.
Detection at different scales are typically carried out by creating an image pyramid at multiple scales

Pros:

- Faster than region-based methods.

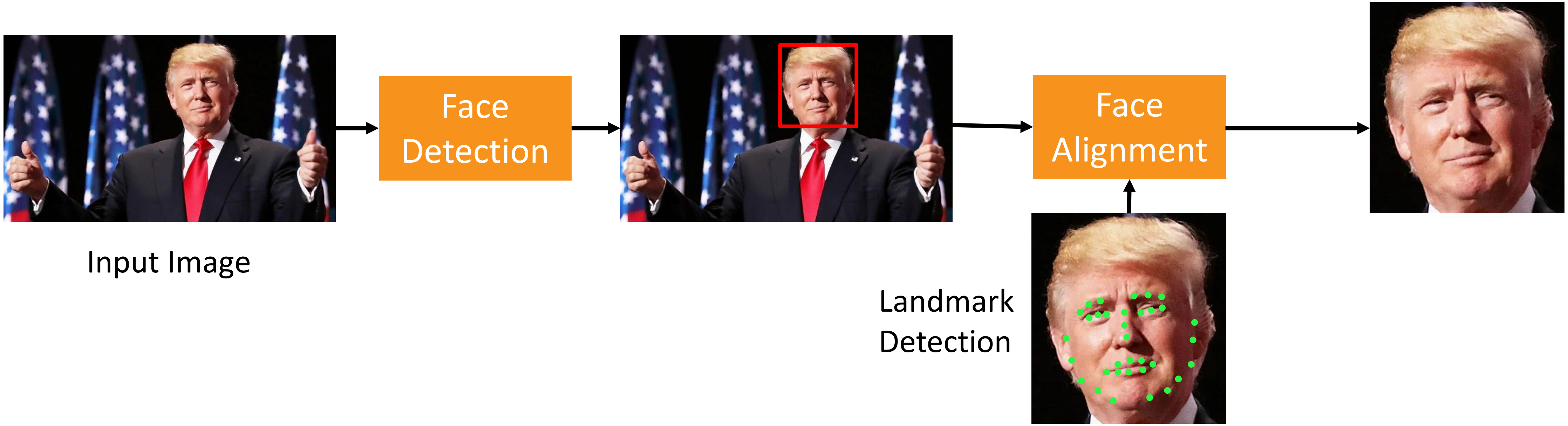
Popular Approaches:

- DDFD, Faceness, Single-Shot Detector…



- W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in Proc. European Conf. Computer Vision, 2016
- S. S. Farfade, M. J. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in Proc. ACM Int. Conf. Multimedia Retrievals, 2015
- S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," in Proc. IEEE Int. Conf. Computer Vision, 2015

Automatic Face Recognition: Architecture

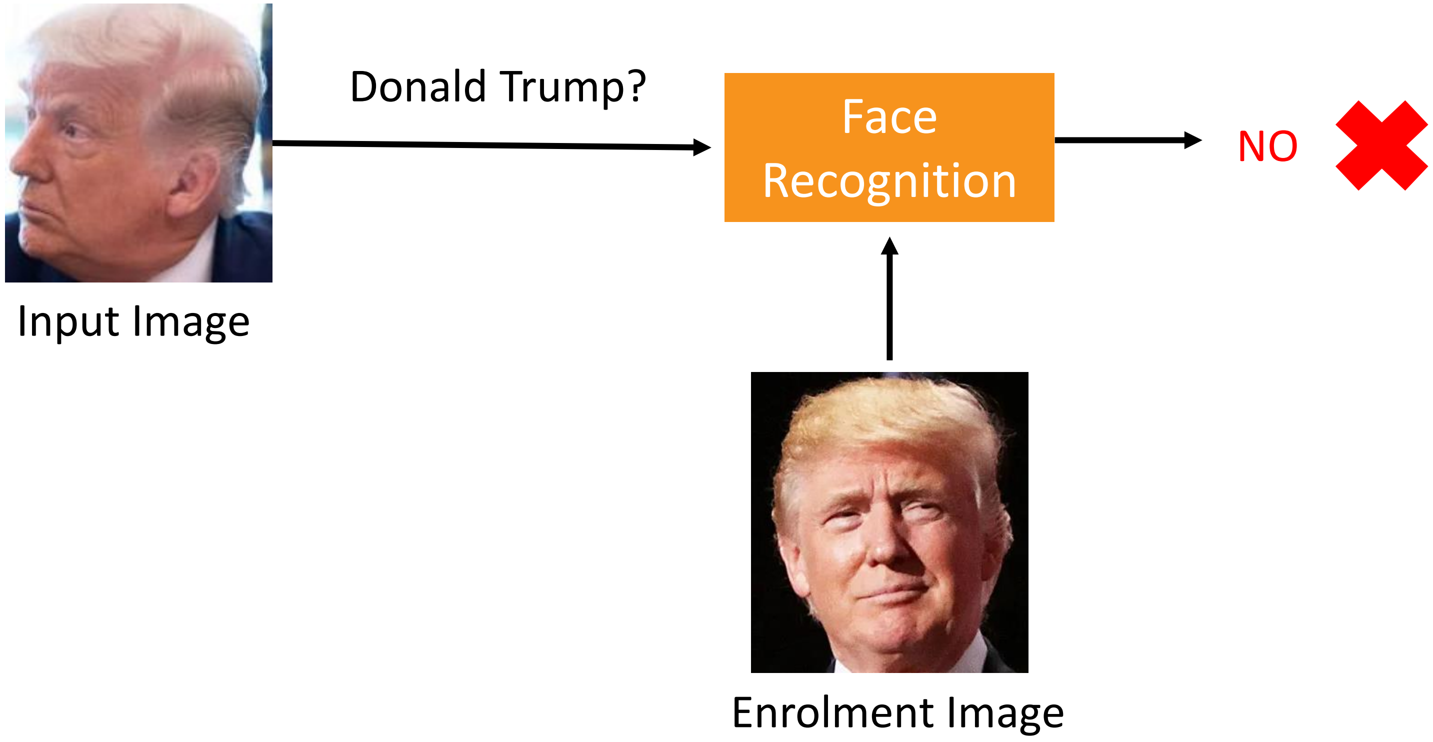


Face Alignment

Facial keypoints such as eye centers, nose tip, mouth corners, etc. can be used to align the face into canonical coordinates. Goal: improve the robustness of face recognition systems against pose variations.

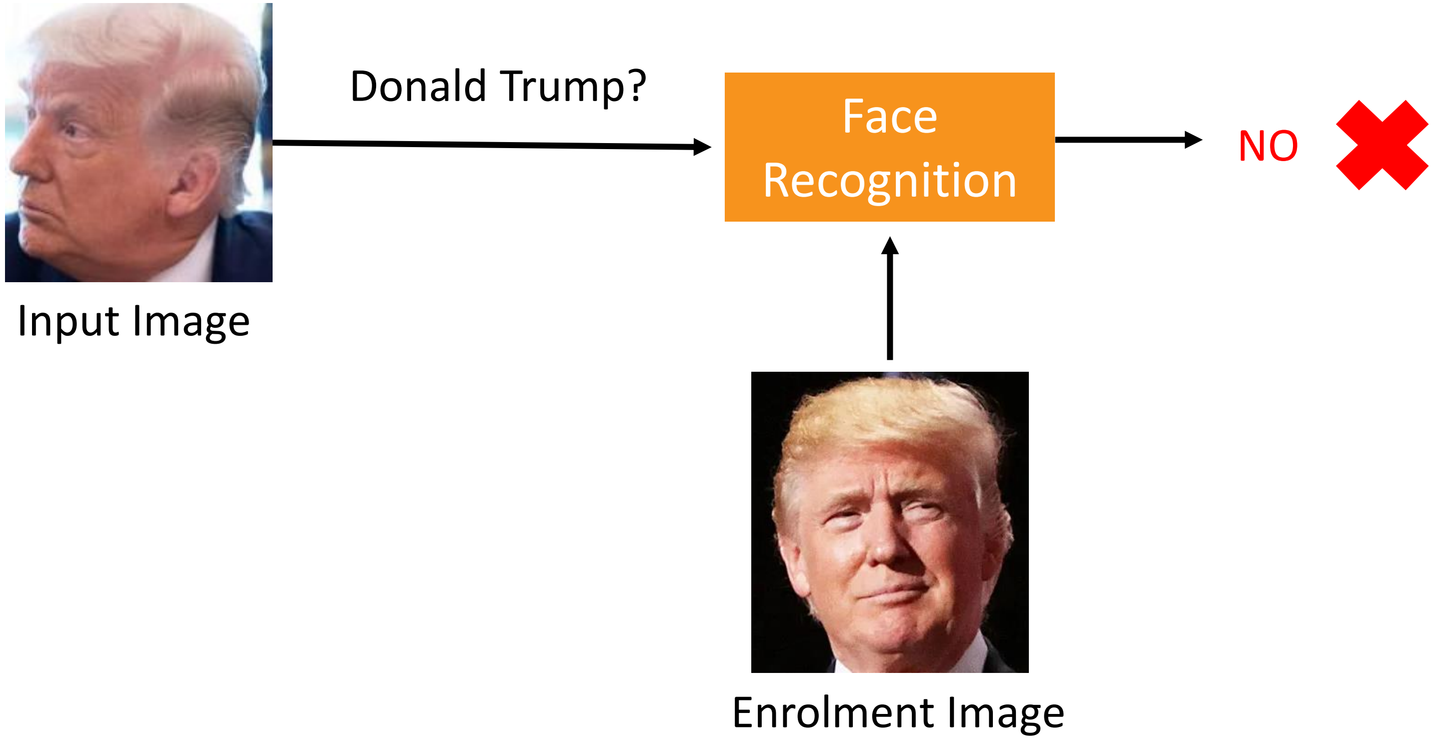
Face Alignment

Facial keypoints such as eye centers, nose tip, mouth corners, etc. can be used to align the face into canonical coordinates. Goal: improve the robustness of face recognition systems against pose variations.



Face Alignment

Facial keypoints such as eye centers, nose tip, mouth corners, etc. can be used to align the face into canonical coordinates. Goal: improve the robustness of face recognition systems against pose variations.



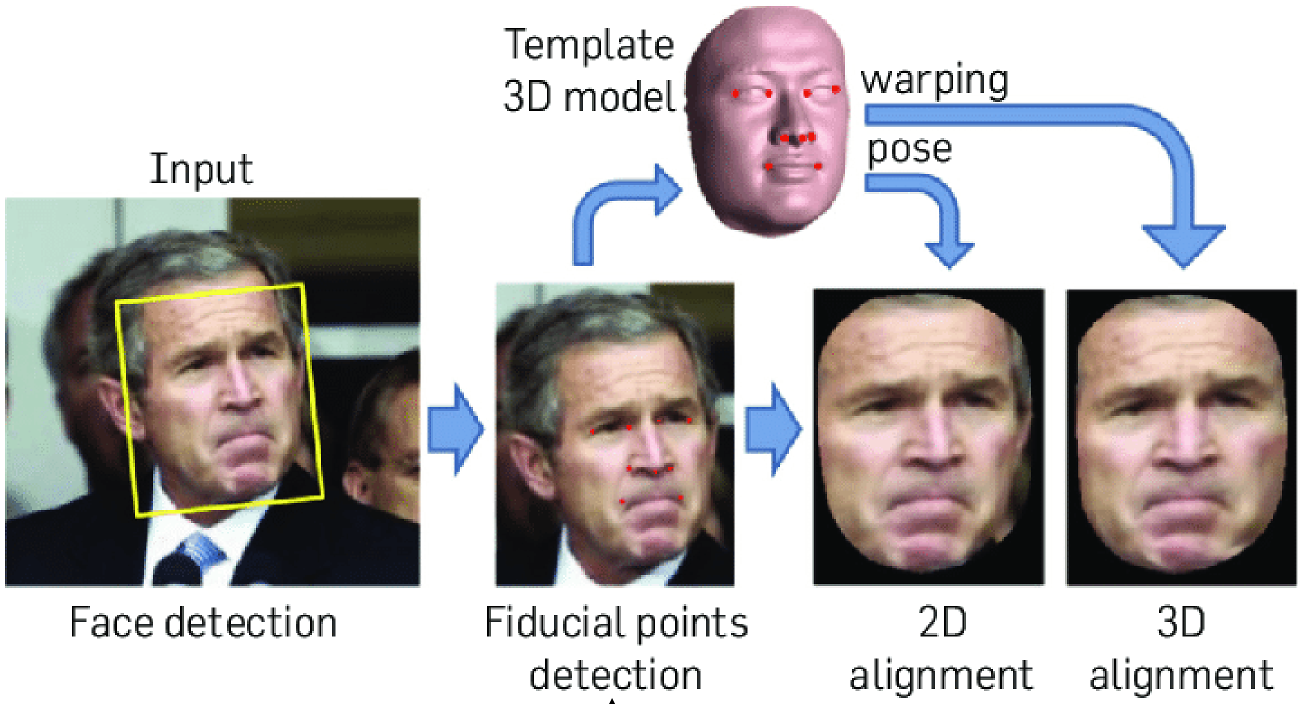
Important and essential intermediary step for many face tasks:

- Face recognition, face tracking, facial expression recognition, head pose estimation, etc.

BiDALab-UAM ©

Face Alignment

Facial keypoints such as eye centers, nose tip, mouth corners, etc. can be used to align the face into canonical coordinates. Goal: improve the robustness of face recognition systems against pose variations.



Landmark Detection

Face Alignment

Facial keypoints such as eye centers, nose tip, mouth corners, etc. can be used to align the face into canonical coordinates. Goal: improve the robustness of face recognition systems against pose variations.

- Luan Tran, Xi Yin, Xiaoming Liu; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1415-1424

Face Alignment

Landmark Detection: detection of facial keypoints such as eye centers, nose tip, mouth corners, etc.

Is it challenging?

(a) Pose (b) Occlusion (c) Expression (d) Illumination

• X. Jin and X. Tan, "Face alignment in-the-wild: A Survey," Computer Vision and Image Understanding, 2017

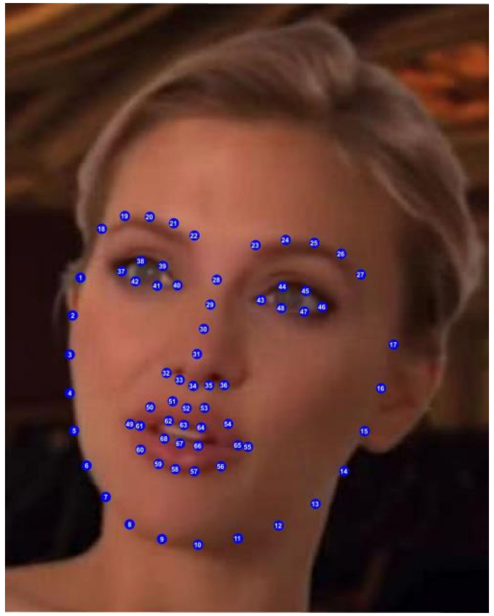
Face Alignment

Landmark Detection: detection of facial keypoints such as eye centers, nose tip, mouth corners, etc.

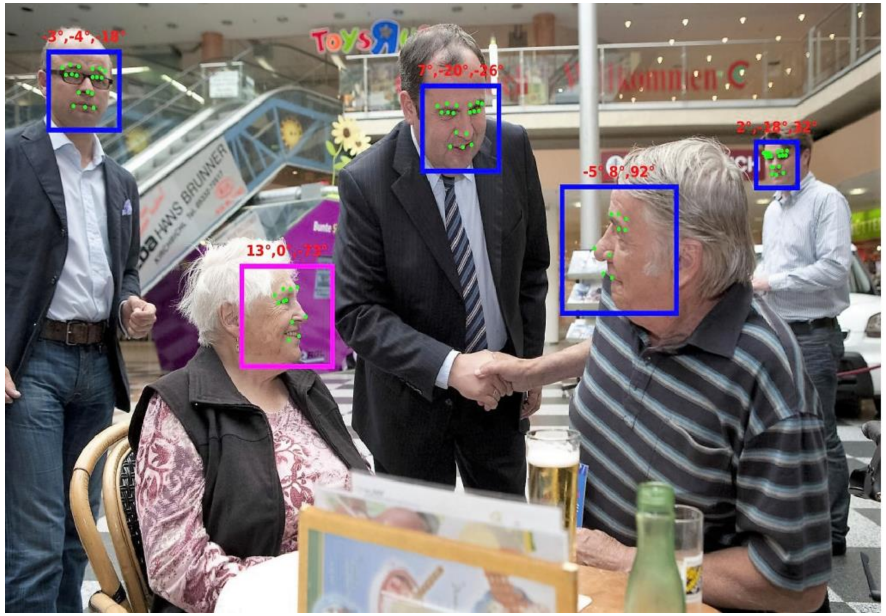
Popular approaches (68 facial keypoints):

- OpenFace2: based on the CE-CLM deep learning implementation
- HyperFace: deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition

OpenFace2



HyperFace
Male
Female
Pose (roll, pitch, yaw)



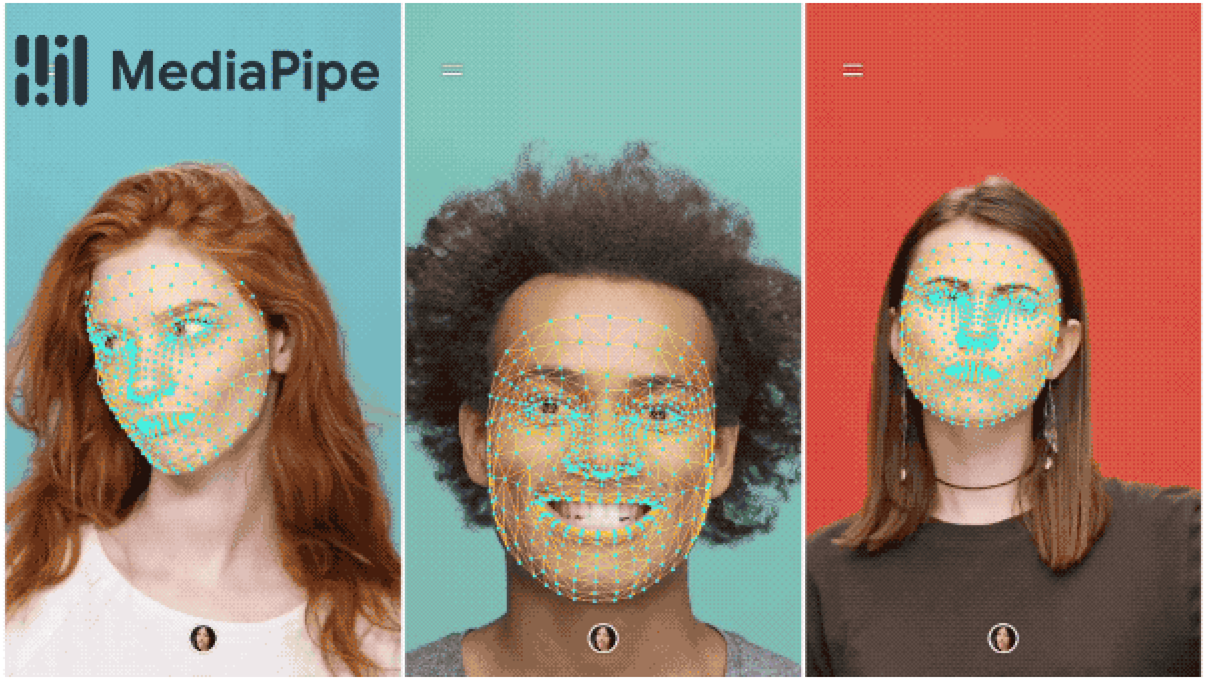
- T. Baltrusaitis, A. Zadeh, Y. Lim, and L. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," in Proc. International Conference on Automatic Face & Gesture Recognition, 2018.
- Rajeev Ranjan, Vishal M. Patel, and Rama Chellappa. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41:121–135, 2019.

Face Alignment

Landmark Detection: detection of facial keypoints such as eye centers, nose tip, mouth corners, etc.

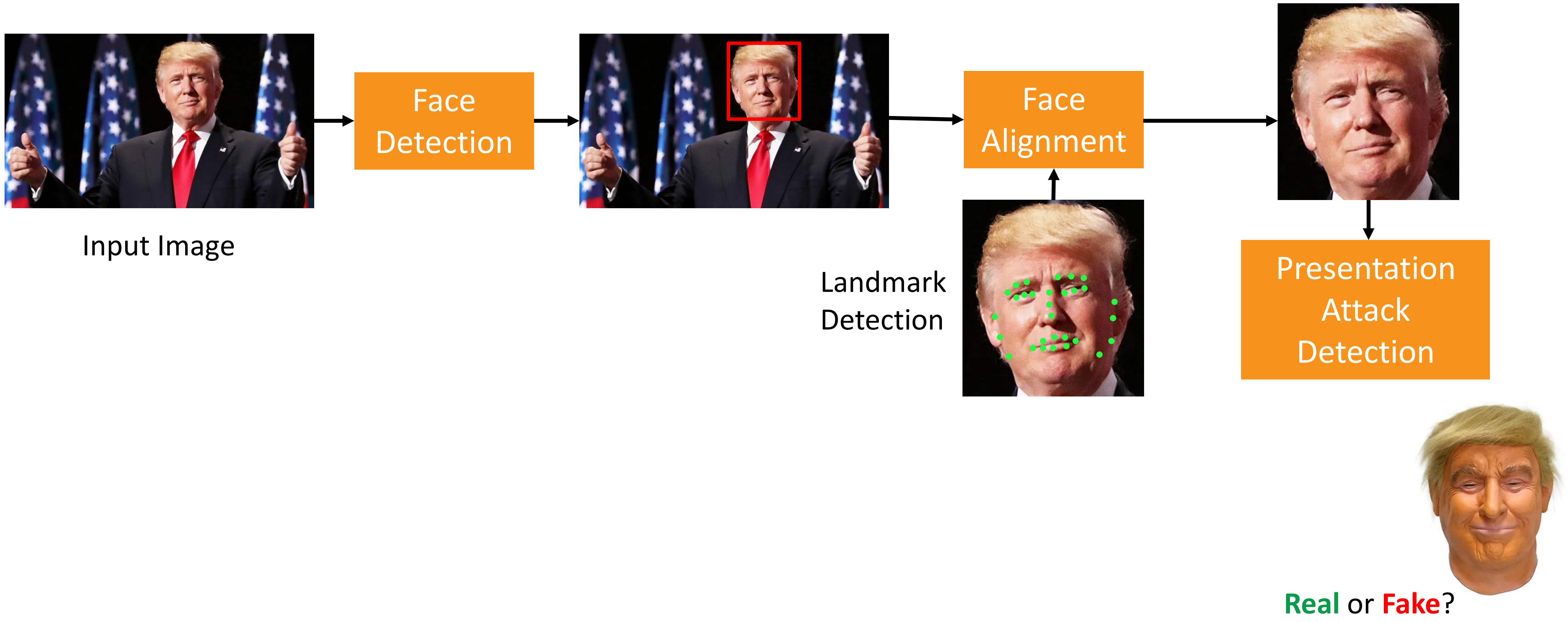
Popular approaches (468 facial keypoints):

- MediaPipe: based on a custom residual neural network architecture



- Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris Mc-Clanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. Mediapipe: A framework for building perception pipelines.arXivpreprint arXiv:1906.08172, 2019. Link: https://google.github.io/mediapipe/

Automatic Face Recognition: Architecture



Presentation Attack Detection

Detection of facial attacks from real faces.

Which one is real?

Presentation Attack Detection

Detection of facial attacks from real faces.

Which one is real?

Real

Fake: printed photo

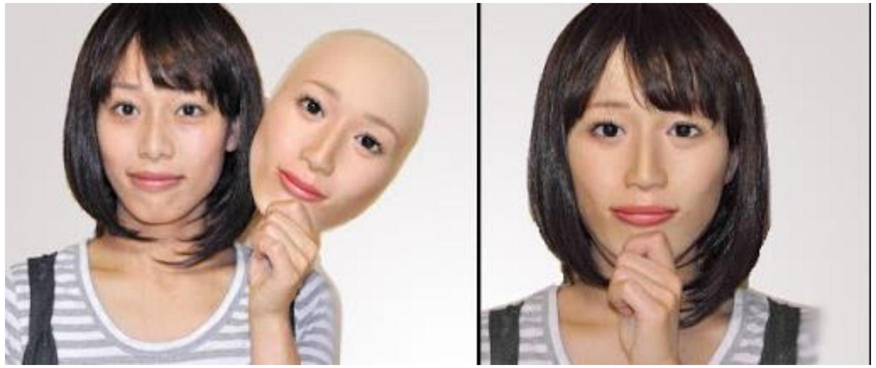
Fake: LCD photo

Fake: HD printed photo

BiDALab-UAM ©

Presentation Attack Detection

Detection of facial attacks from real faces.



Presentation Attack Detection

Police arrest passenger who boarded plane in Hong Kong as an old man in flat cap and arrived in Canada a young Asian refugee

By DAILY MAIL REPORTER

'Unbelievable concealment': The plane passenger boarded with a silicone face and neck mask that gave him the appearance of an old man. He agreed to put the disguise on to be photographed after being picked up on arrival in Canada

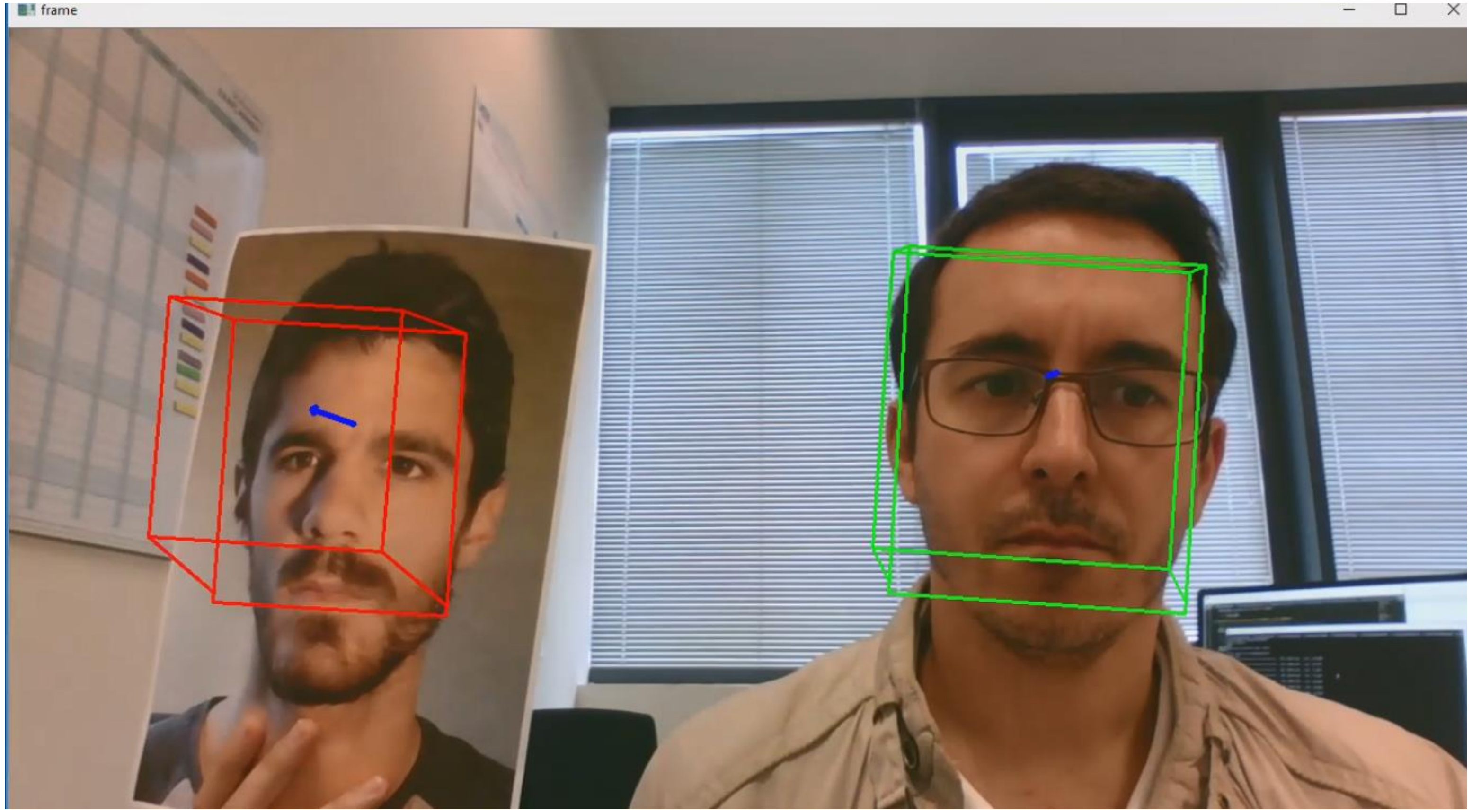
BiDALab-UAM ©

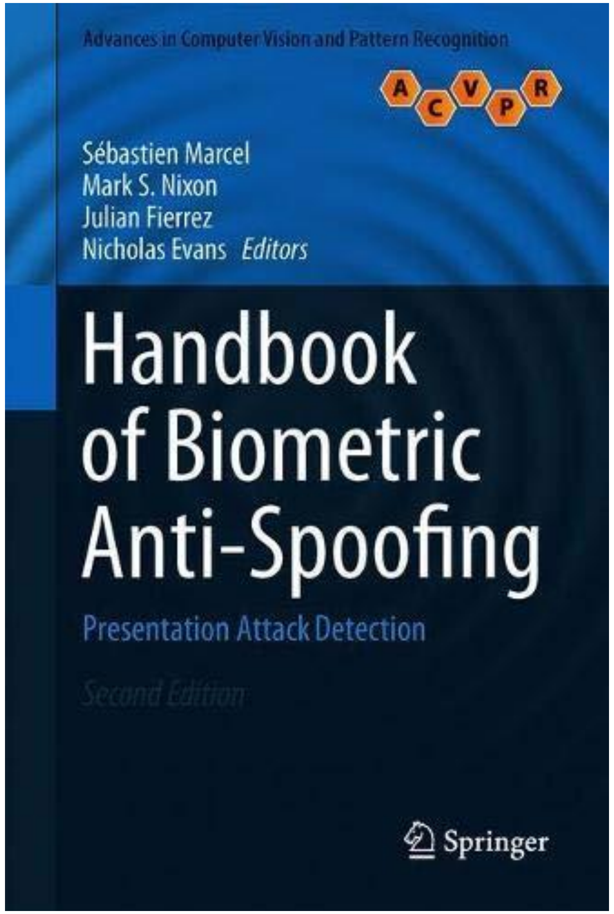
Presentation Attack Detection

BiDA Lab

Biometrics & Data Pattern Analytics Lab

Detetection of facial attacks from real faces.





- S. Marcel, M.S. Nixon, J. Fierrez, N. Evans, " Handbook of Biometric Anti-Spoofing (2nd Edition)," Springer, 2019.
- J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," IEEE Access, 2, 1530-1552, 2014
- J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," IEEE Transactions on Image Processing, 23(2), 710-724, 2013.

Presentation Attack Detection

Approaches:

- Texture-based techniques: perform an image analysis of the facial texture to discover unnatural features that may be related to presentation attacks.
  - Limitations: It require high-quality images. It does not work properly with bad illumination.

- I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In Proc. Biometrics Special Interest Group (BIOSIG)
- A. Agarwal, R. Singh, and V. M. Face anti-spoofing using Haralick features. In Proc. IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2016
- J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. IEEE Transactions on Image Processing, 23(2):710–724, 2014

Presentation Attack Detection

Approaches:

- Texture-based techniques: perform an image analysis of the facial texture to discover unnatural features that may be related to presentation attacks.
  - Limitations: It require high-quality images. It does not work properly with bad illumination.

- Depth-based techniques: perform an image analysis of the depth information (using special sensors such as Microsoft Kinect or Intel Real-Sense, or even RGB sensors).
  - Limitations: Inefficient when dealing with 3D presentation attacks.

- A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. In Proc. IEEE International Conference on Computer Vision (ICCV)
- Y. Liu, A. Jourabloo and X. Liu, “Learning deep models for face anti-spoofing: binary or auxiliary supervision,” in: Proceeding of IEEE Computer Vision and Pattern Recognition, 2018

BiDALab-UAM ©

Presentation Attack Detection

Approaches:

- Texture-based techniques: perform an image analysis of the facial texture to discover unnatural features that may be related to presentation attacks.
  - Limitations: It require high-quality images. It does not work properly with bad illumination.

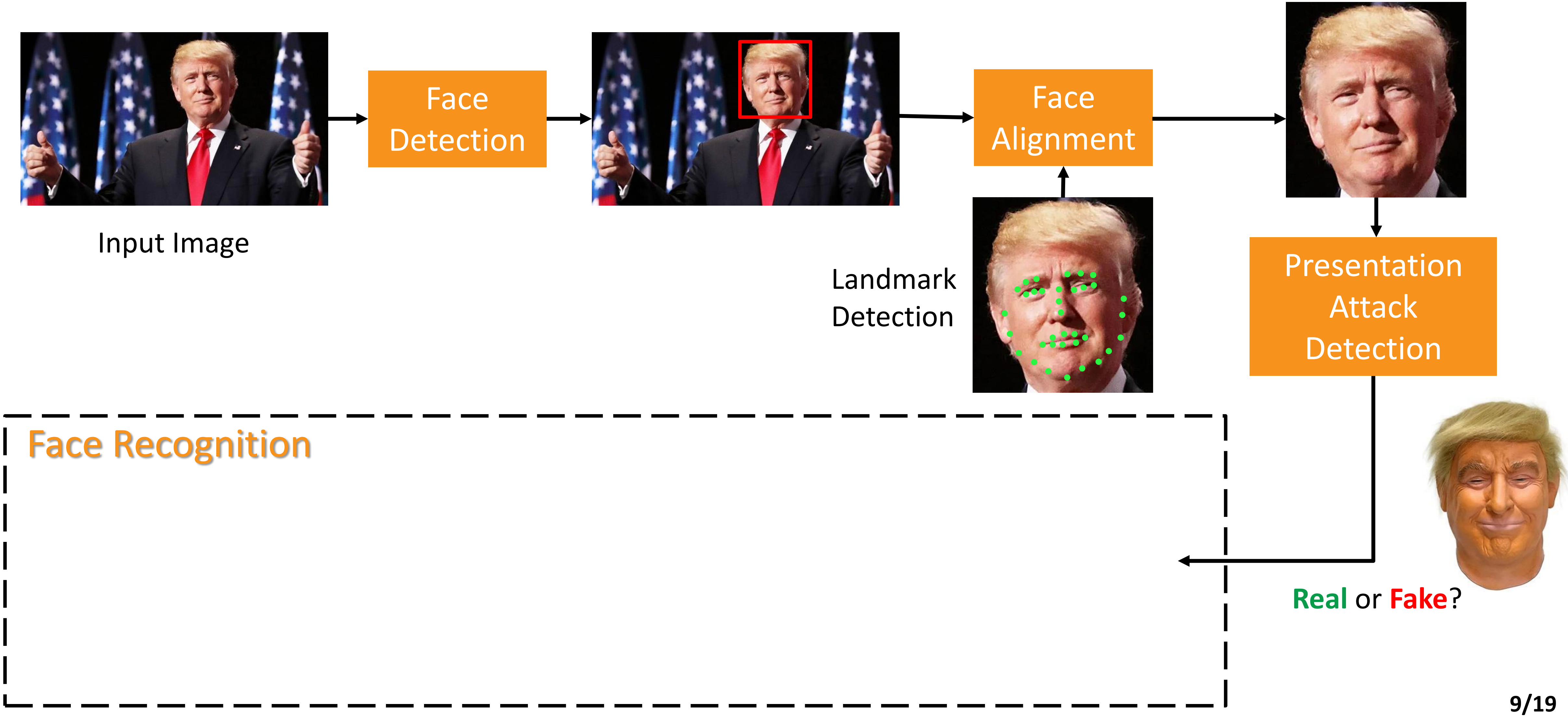
- Depth-based techniques: perform an image analysis of the depth information (using special sensors such as Microsoft Kinect or Intel Real-Sense, or even RGB sensors).
  - Limitations: Inefficient when dealing with 3D presentation attacks.

- Physiological-based techniques: perform a video analysis of the physiological information of the person, e.g., heart-rate information, eye blink information, etc.
  - Limitations: It requires more time. It depends on illumination, resolution, face detection and tracking…

- J. Hernandez-Ortega, J. Fierrez, A. Morales, and P. Tome, "Time analysis of pulse-based face anti-spoofing in visible and NIR," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2018.
- S. Jia, G. Guo, and Z. Xu, "A survey on 3D mask presentation attack detection and countermeasures," Pattern Recognition, 98, 107032, 2020.
- A. Al-Rashid, "A Three Steps Eye-Liveness Validation System. In Proc. IEEE International Conference on Cyber Security for Emerging Technologies, 2019.

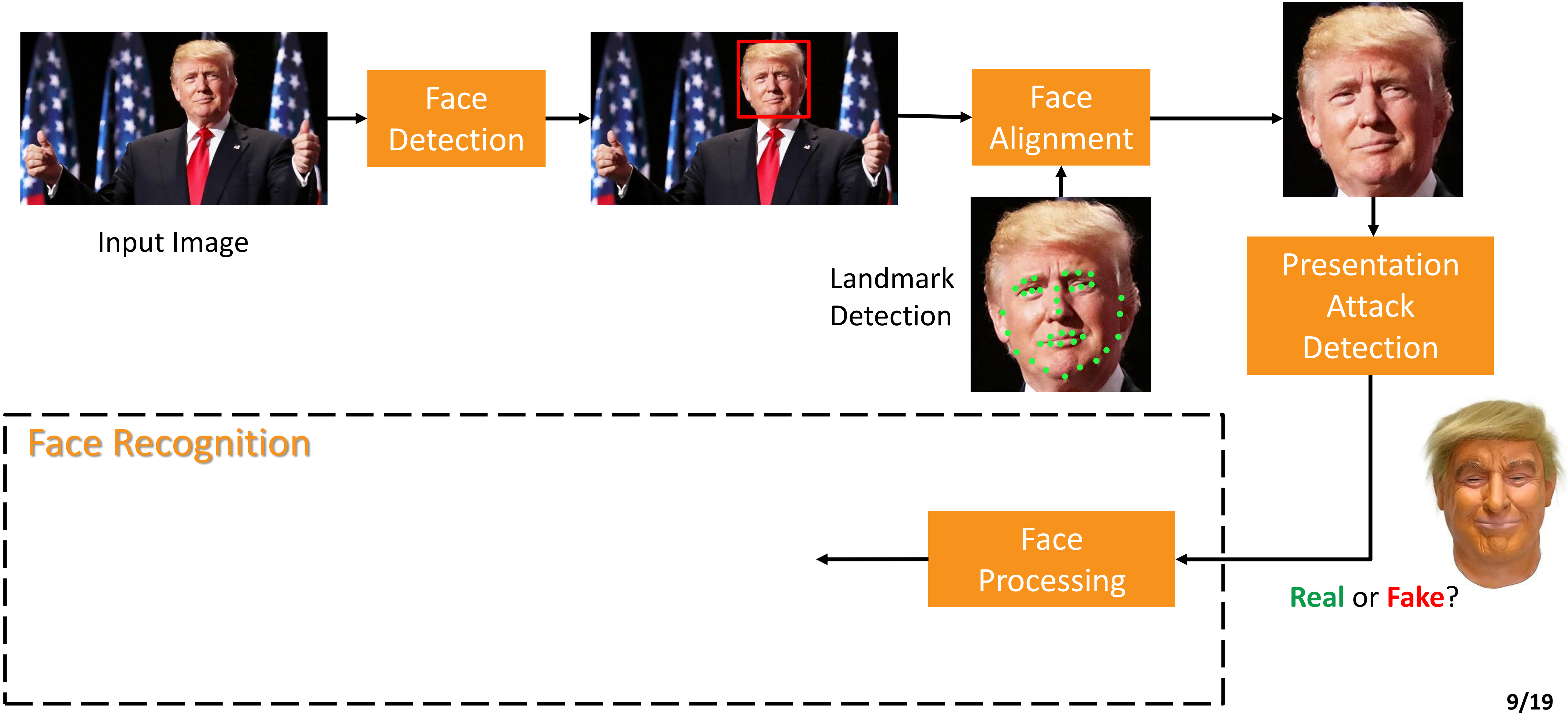
BiDALab-UAM ©

Automatic Face Recognition: Architecture



BiDALab-UAM ©

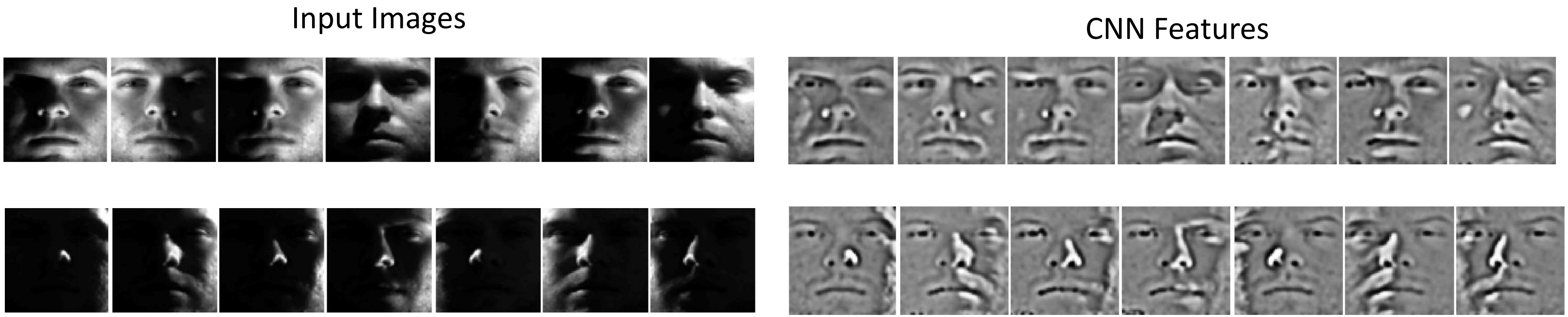
Automatic Face Recognition: Architecture



BiDALab-UAM ©

Face Processing

Improve the robustness of face recognition systems against pose, illumination, expression, and occlusion variations as they affect the performance of the systems (even those based on deep learning technology).



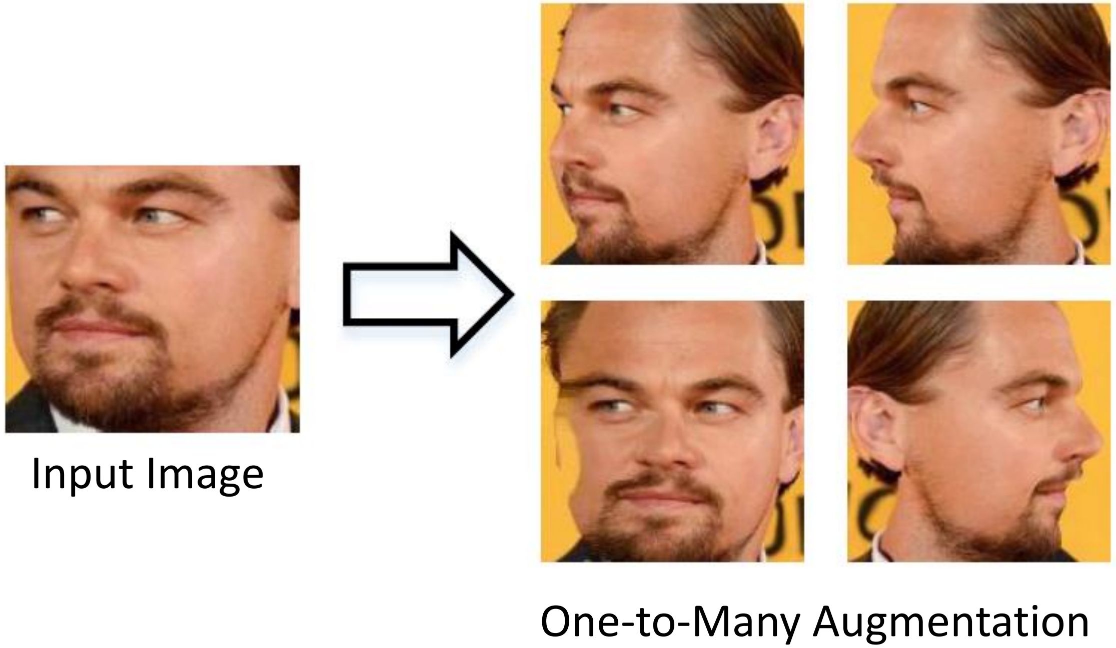
• M. Mehdipour Ghazi and H. Kemal Ekenel, "A comprehensive analysis of deep learning based representation for face recognition," in Proc. IEEE/CVF Conference on Computer vision and Pattern Recognition Workshops, 2016

BiDALab-UAM ©

Face Processing

Approaches:

- One-to-many augmentation: these methods generate many patches or images of the pose (or others) variability from a single image to enable face recognition systems to learn pose-invariant representations.



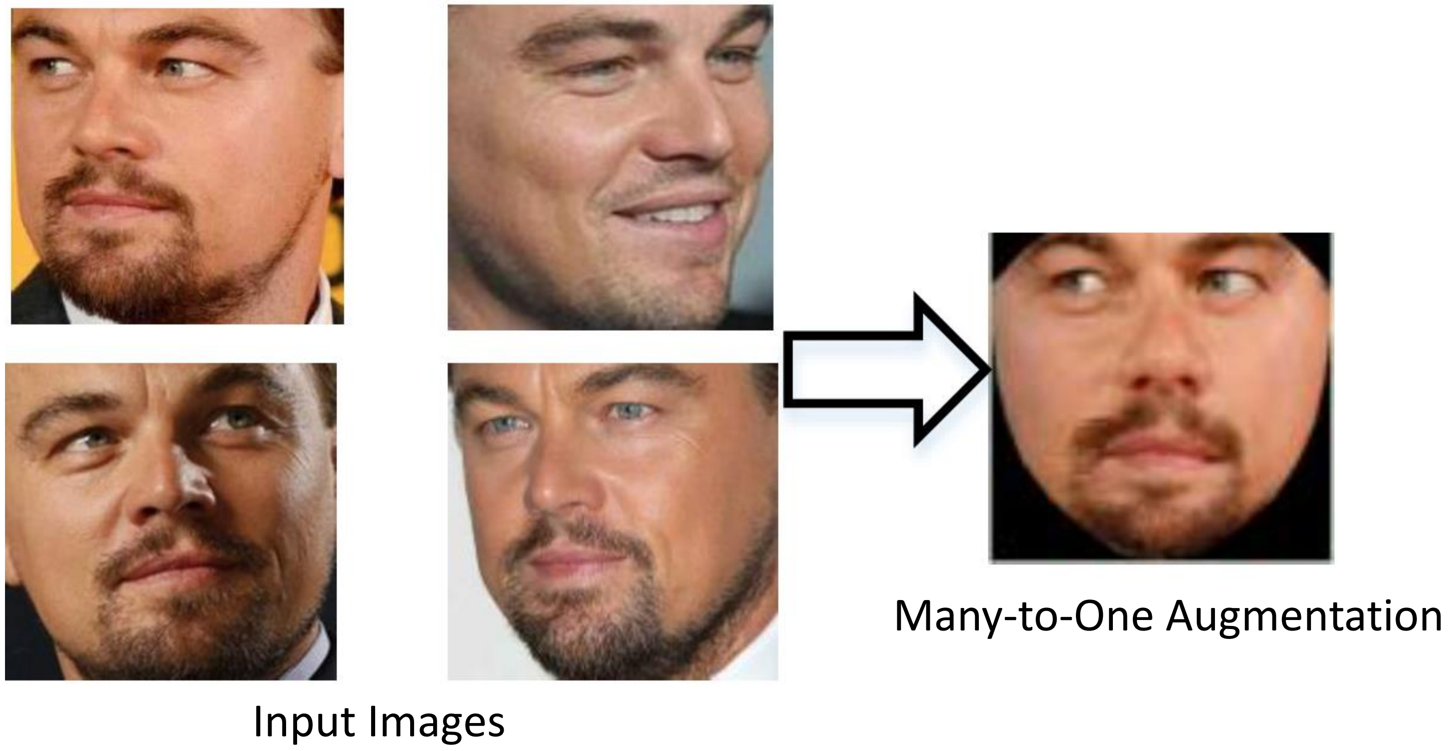
- A. R. Chowdhury, T. Y. Lin, S. Maji, and E. Learned-Miller, “One-to-many face recognition with bilinear cnns” in Proc. IEEE Winter Conference on Applications of Computer Vision, 2016
- E. Richardson, M. Sela, R. Or-El, and R. Kimmel, “Learning detailed face reconstruction from a single image,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017.
- D. Wang, C. Otto, and A. K. Jain, “Face search at scale,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1122–1136, 2016.

BiDALab-UAM ©

Face Processing

Approaches:

- Many-to-one normalization: these methods recover the canonical view of face images from one or many images of a non-frontal view.

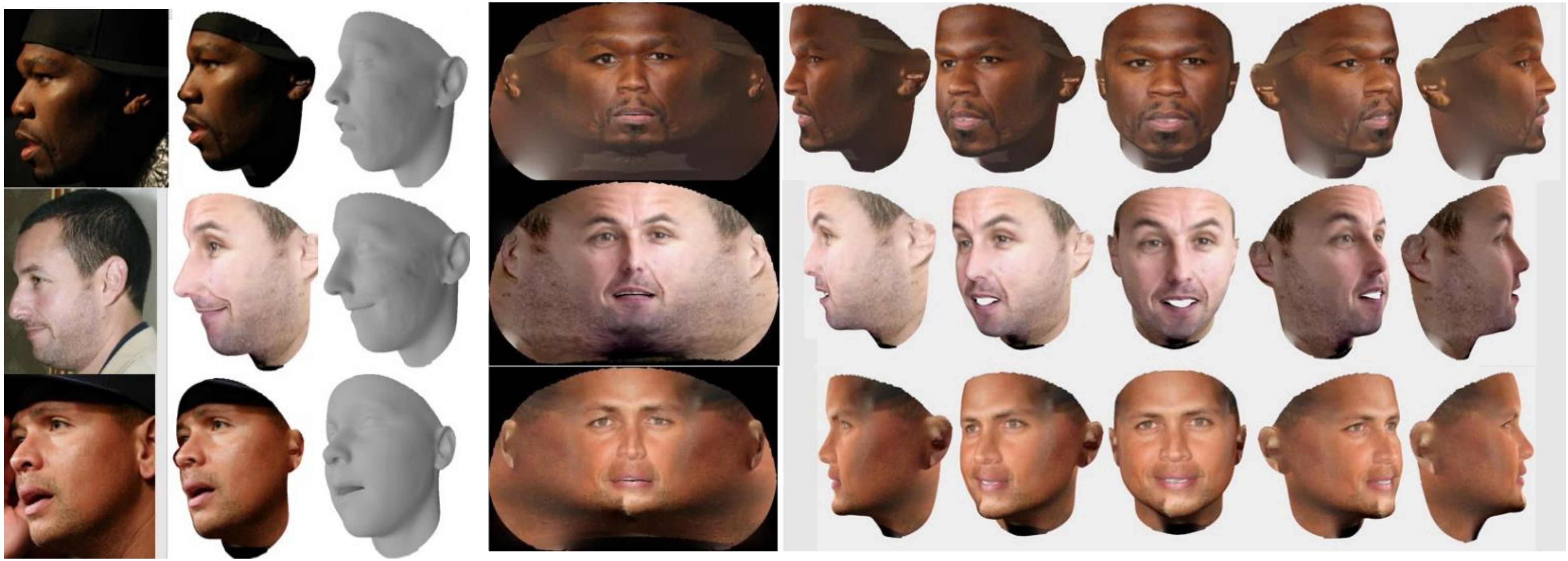


- Y. Zhang, M. Shao, E. K. Wong, and Y. Fu, “Random faces guided sparse many-to-one encoder for pose-invariant face recognition,” in Proc. IEEE/CVF International Conference on Computer Vision, 2013,
- A. R. Chowdhury, T. Y. Lin, S. Maji, and E. Learned-Miller, “One-to-many face recognition with bilinear cnns” in Proc. IEEE Winter Conference on Applications of Computer Vision, 2016
- E. Zhou, Z. Cao, and J. Sun, “Gridface: Face rectification via learning local homography transformations,” in Proc. European Conference on Computer Vision, 2018.

Face Processing

Approaches:

- Many-to-one normalization: these methods recover the canonical view of face images from one or many images of a non-frontal view.



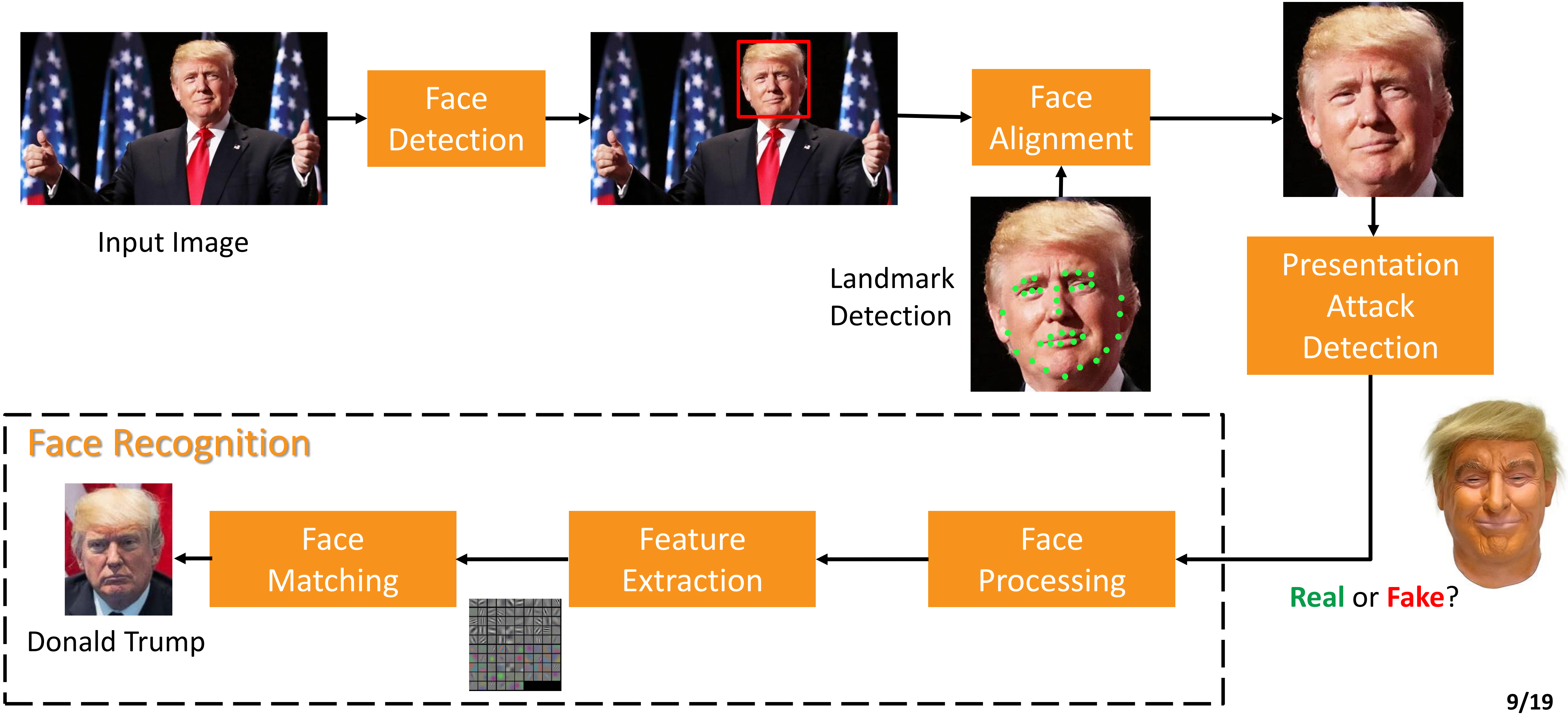
Input Image

Many-to-One Augmentation

- J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, “UV-GAN: Adversarial facial uv map completion for pose-invariant face recognition. In Proc. IEEE/CVF Cconference on Computer Vision and Pattern Recognition, 2018

BiDALab-UAM ©

Automatic Face Recognition: Architecture



Feature Extraction

Extract robust features from the input image for the task of face recognition.

1st Generation (1990~)

Basic Methods

Eigenfaces
Fisherfaces

2nd Generation (2000~)

Linear Subspace Methods

Sparse Representation
Metric Learning

3rd Generation (2010~)

Non-Linear Methods

Deep Learning

BiDALab-UAM ©

Feature Extraction

Extract robust features from the input image for the task of face recognition.

1st Generation (1990~)

Basic Methods

Eigenfaces
Fisherfaces

2nd Generation (2000~)

Linear Subspace Methods

Sparse Representation
Metric Learning

3rd Generation (2010~)

Non-Linear Methods

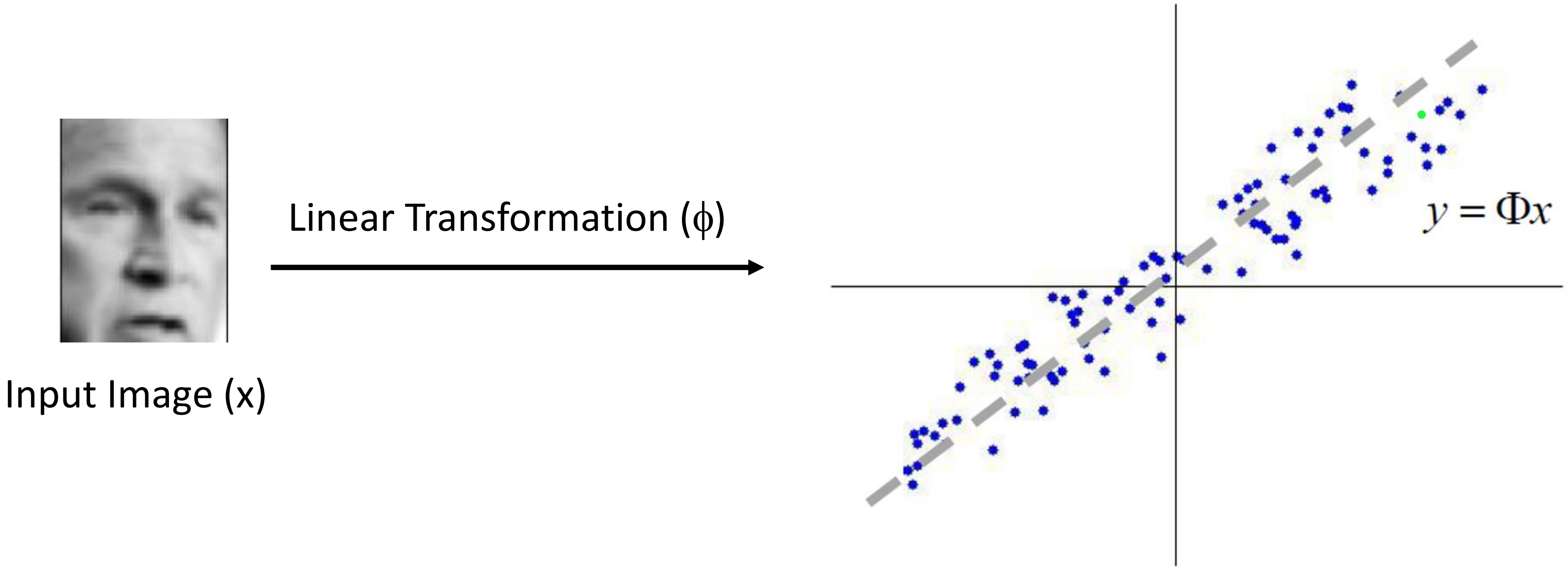
Deep Learning

BiDALab-UAM ©

1^{st} Generation: Eigenfaces

Based on Principal Component Analysis (PCA).

- It uses dimensionality reduction and linear algebra concepts to recognize faces (projecting face images on small feature spaces).



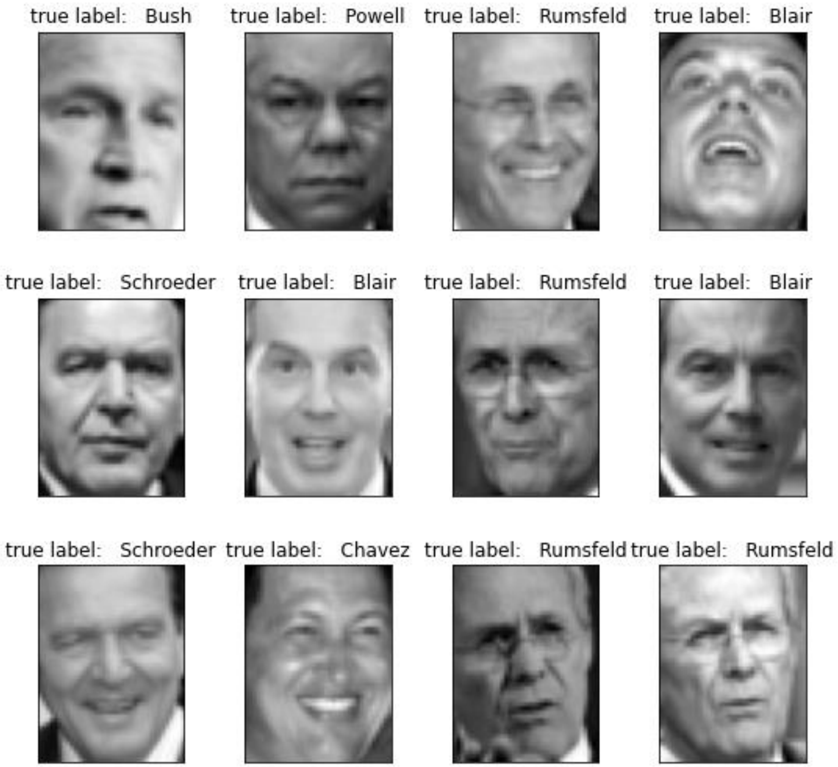
- M. A.Turk, and A. Pentland, “Face recognition using eigenfaces” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991

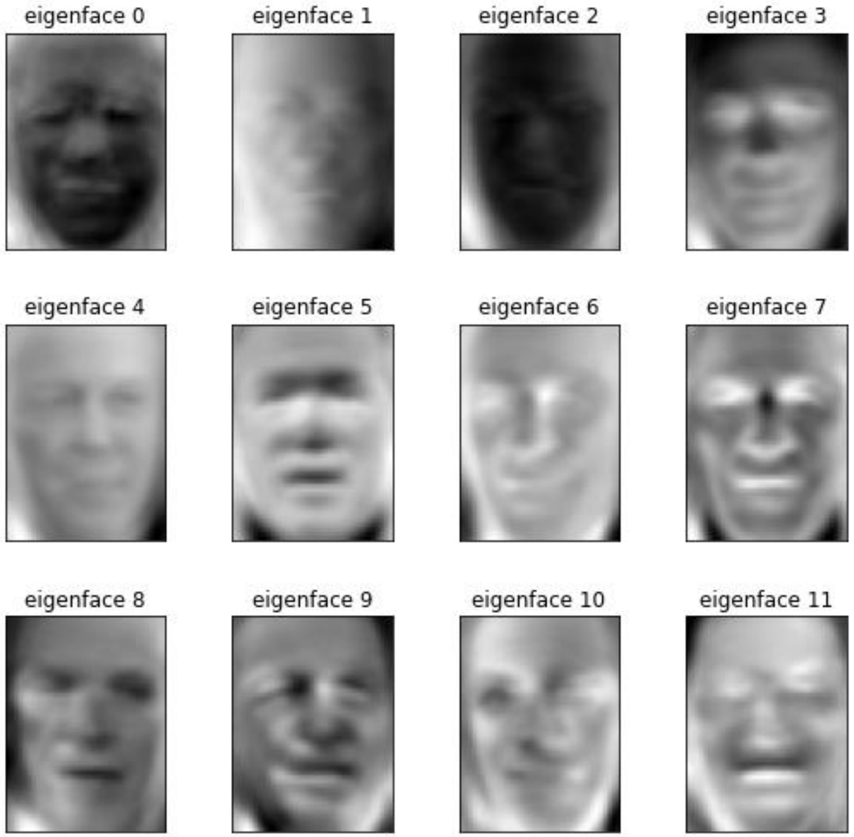
BiDALab-UAM ©

1st Generation: Eigenfaces

Based on Principal Component Analysis (PCA).

- It uses dimensionality reduction and linear algebra concepts to recognize faces (projecting face images on small feature spaces).
- Projection vector is a set of eigenvector of training samples.



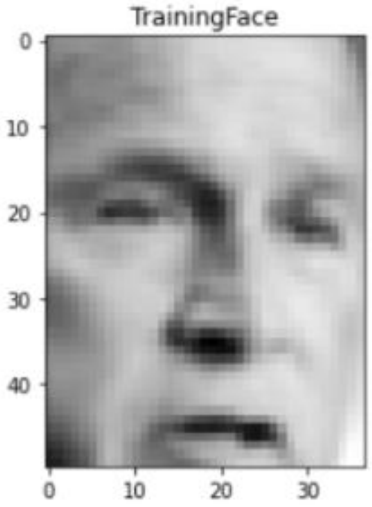


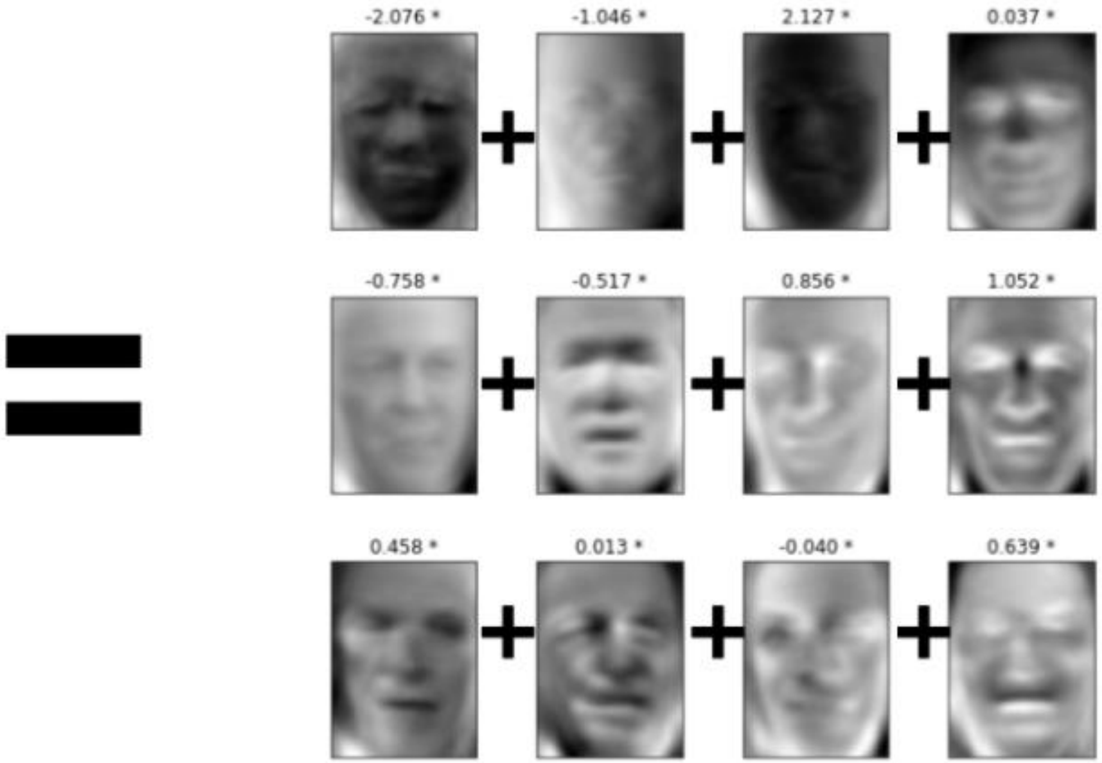
BiDALab-UAM ©

1st Generation: Eigenfaces

Based on Principal Component Analysis (PCA).

- It uses dimensionality reduction and linear algebra concepts to recognize faces (projecting face images on small feature spaces).
- Projection vector is a set of eigenvector of training samples.





- Good visual and comprehensive example: https://www.geeksforgeeks.org/ml-face-recognition-using-eigenfaces-pca-algorithm/

1st Generation: Eigenfaces

Based on Principal Component Analysis (PCA).

- It uses dimensionality reduction and linear algebra concepts to recognize faces (projecting face images on small feature spaces).

- Projection vector is a set of eigenvector of training samples.

- Limitations: PCA projection is optimal for reconstruction of face, but may not be optimal for discrimination (it is only based on simple linear algebra).
    - Unseen users.
    - Variability in the pose, illumination, etc.

1st Generation: Fisherfaces

Based on Linear Discriminant Analysis (LDA).

- Improved version of Eigenfaces insensitive to large variation in lighting direction and facial expression.

- P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-720. 1997.

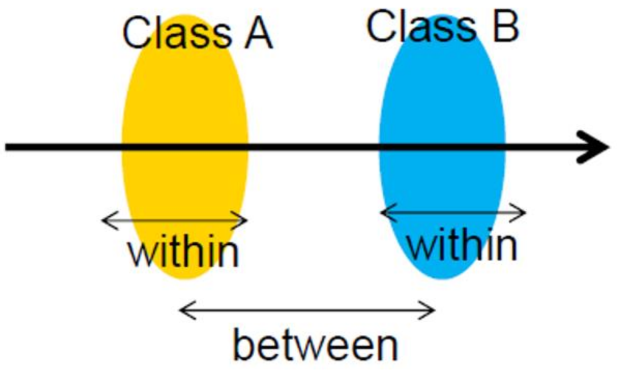
1st Generation: Fisherfaces

Based on Linear Discriminant Analysis (LDA).

- Improved version of Eigenfaces insensitive to large variation in lighting direction and facial expression.
- Optimal subspace is obtained by maximizing the ratio of between and within class scatter matrix:

$$r = \left\| \Phi^T S_b \Phi \right\| / \left\| \Phi^T S_w \Phi \right\|$$

S_b : between class scatter matrix S_w : within class scatter matrix



- P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-720. 1997.

BiDALab-UAM ©

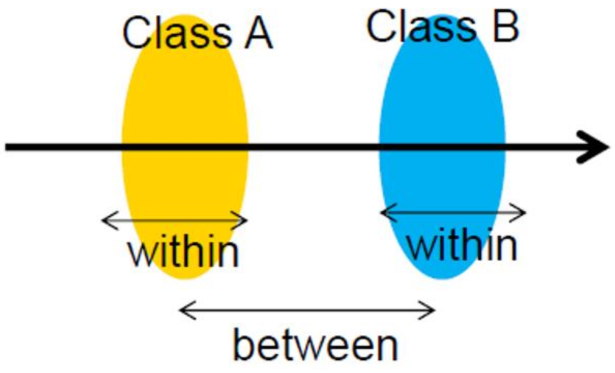
1st Generation: Fisherfaces

Based on Linear Discriminant Analysis (LDA).

- Improved version of Eigenfaces insensitive to large variation in lighting direction and facial expression.
- Optimal subspace is obtained by maximizing the ratio of between and within class scatter matrix:

$$r = \left\| \Phi^T S_b \Phi \right\| / \left\| \Phi^T S_w \Phi \right\|$$

S_b : between class scatter matrix $\quad$ S_w : within class scatter matrix



- Limitations: It is difficult to discriminate faces near the individual boundaries. The feature extraction process is still based on simple linear algebra.

- P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-720. 1997.

Feature Extraction

Extract robust features from the input image for the task of face recognition.

1st Generation (1990~)

Basic Methods

Eigenfaces
Fisherfaces

2nd Generation (2000~)

Linear Subspace Methods

Sparse Representation
Metric Learning

3rd Generation (2010~)

Non-Linear Methods

Deep Learning

BiDALab-UAM ©

2nd Generation: Sparse Representation

Train a Sparse Matrix minimizing the L1 norm with training data.

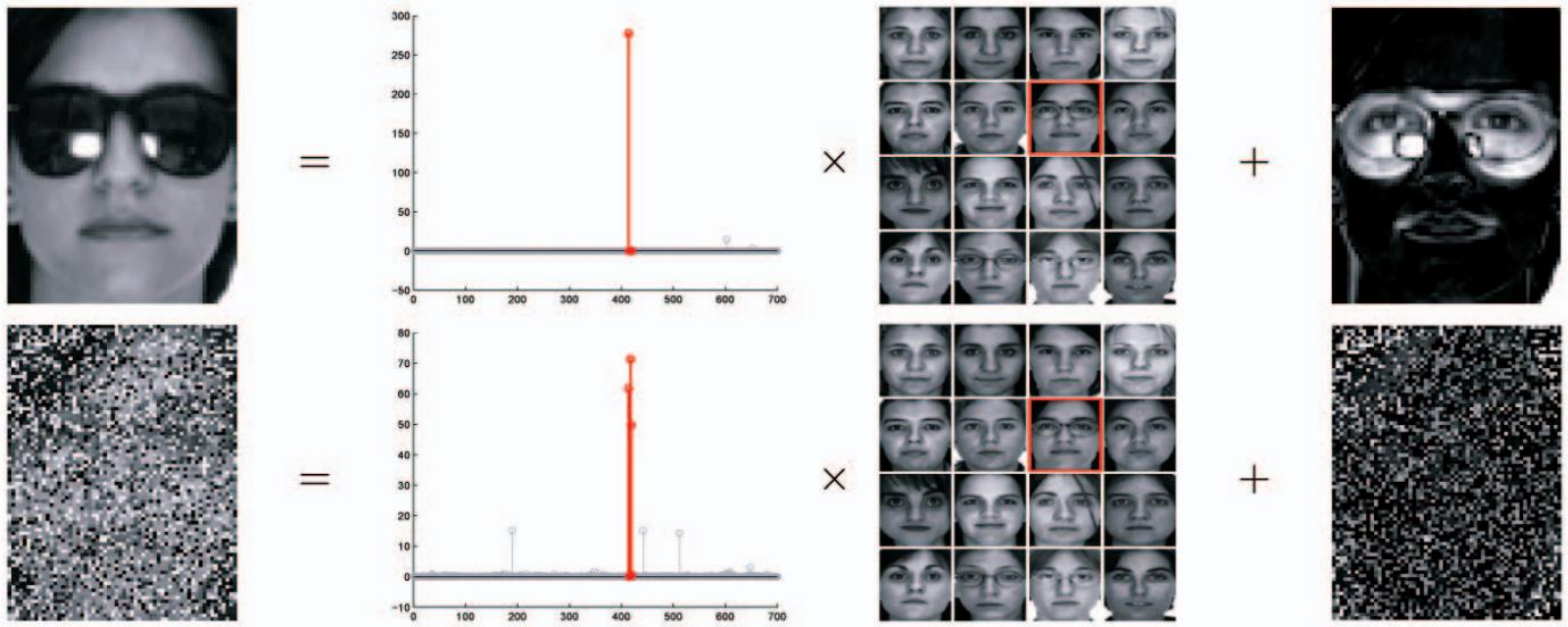
- Very robust method against expression, illumination, and occlusion variations.

- J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), 210-227, 2008.

2nd Generation: Sparse Representation

Train a Sparse Matrix minimizing the L1 norm with training data.

- Very robust method against expression, illumination, and occlusion variations.



- The method represents a test image (left), which can be occluded or corrupt, as a linear sparse combination of all the training images (middle) plus some artefacts (right) due to possible occlusions or corruptions. Red: true identity selected by the algorithm.

- J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), 210-227, 2008.

2nd Generation: Metric Learning

Metric learning aims to automatically construct/learn task-specific distance metrics from supervised data.

- M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," In Proc. IEEE 12th international conference on computer vision, 2009.

BiDALab-UAM ©

2nd Generation: Metric Learning

Metric learning aims to automatically construct/learn task-specific distance metrics from supervised data.

Most methods learn a Mahalanobis distance metric between features xi and xj based on an objective function:

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$$

where M is a symmetric positive definite matrix. Mahalanobis distance: minimize the distances between similarly labeled inputs while maximizing the distances between differently labeled inputs.

- M. Guillaumin, J. Verbeek, and C. Schmid, “Is that you? Metric learning approaches for face identification,” In Proc. IEEE 12th international conference on computer vision, 2009.

BiDALab-UAM ©

2nd Generation: Metric Learning

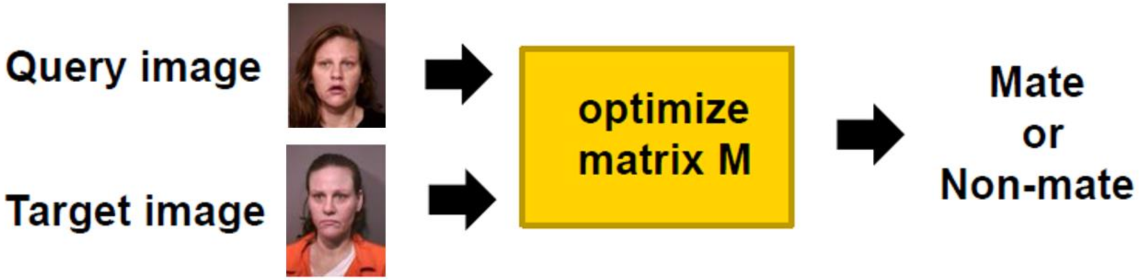
Metric learning aims to automatically construct/learn task-specific distance metrics from supervised data.

Most methods learn a Mahalanobis distance metric between features xi and xj based on an objective function:

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)$$

where M is a symmetric positive definite matrix. Mahalanobis distance: minimize the distances between similarly labeled inputs while maximizing the distances between differently labeled inputs.

Goal: design/learn matrix M to discriminate between Mate and Non-Mate:



- M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," In Proc. IEEE/CVF 12th International Conference on Computer Vision, 2009.

BiDALab-UAM ©

2^{nd} Generation: Metric Learning

State-of-the-Art Approaches:

- Large Margin Nearest Neighbour Metrics: method that learns the matrix M designed to improve results of K Nearest Neighbour (KNN) classification. The objective function comprises two terms:

$$\varepsilon_{\text{pull}}(\mathbf{L}) = \sum_{j \rightsquigarrow i} \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|^2.$$

$$\varepsilon_{\text{push}}(\mathbf{L}) = \sum_{i,j \rightsquigarrow i} \sum_{l} (1 - y_{il}) \left[1 + \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|^2 - \|\mathbf{L}(\vec{x}_i - \vec{x}_l)\|^2\right]_+$$

1^{st} Term: penalizes large distances between each input and its target neighbors.

2^{nd} Term: penalizes small distances between differently labeled examples.

- K. Weinberger, J. Blitzer, and L. Saul, "Distance metric learning for large margin nearest neighbor classification," in Proc. NIPS, 2006.

BiDALab-UAM ©

2nd Generation: Metric Learning

State-of-the-Art Approaches:

- Large Margin Nearest Neighbour Metrics: method that learns the matrix M designed to improve results of K Nearest Neighbour (KNN) classification. The objective function comprises two terms:

$$\varepsilon_{\mathrm{pull}}(\mathbf{L}) = \sum_{j \rightsquigarrow i} \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|^2.$$

$$\varepsilon_{\mathrm{push}}(\mathbf{L}) = \sum_{i,j \rightsquigarrow i} \sum_{l} (1 - y_{il}) \left[1 + \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|^2 - \|\mathbf{L}(\vec{x}_i - \vec{x}_l)\|^2\right]_+$$

1st Term: penalizes large distances between each input and its target neighbors.

2nd Term: penalizes small distances between differently labeled examples.

- Information Theoretic Metric Learning: optimize the matrix M under a wide range of possible contraints and prior knowledge on the Mahalanobis distance. The objective function is Kullbach-Leiber divergence

- K. Weinberger, J. Blitzer, and L. Saul, "Distance metric learning for large margin nearest neighbor classification," in Proc. NIPS, 2006.
- J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon, "Information theoretic metric learning," in ICML, 2007.

BiDALab-UAM ©

2^{nd} Generation: Metric Learning

State-of-the-Art Approaches:

- Large Margin Nearest Neighbour Metrics: method that learns the matrix M designed to improve results of K Nearest Neighbour (KNN) classification. The objective function comprises two terms:

$$\varepsilon_{\text{pull}}(\mathbf{L}) = \sum_{j \rightsquigarrow i} \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|^2.$$

$$\varepsilon_{\text{push}}(\mathbf{L}) = \sum_{i,j \rightsquigarrow i} \sum_{l} (1 - y_{il}) \left[1 + \|\mathbf{L}(\vec{x}_i - \vec{x}_j)\|^2 - \|\mathbf{L}(\vec{x}_i - \vec{x}_l)\|^2\right]_+$$

1^{st} Term: penalizes large distances between each input and its target neighbors.

2^{nd} Term: penalizes small distances between differently labeled examples.

- Information Theoretic Metric Learning: optimize the matrix M under a wide range of possible contraints and prior knowledge on the Mahalanobis distance. The objective function is Kullbach-Leiber divergence

Limitations: robustness against the complex non-linear facial appearance variations.

- K. Weinberger, J. Blitzer, and L. Saul, “Distance metric learning for large margin nearest neighbor classification,” in Proc. NIPS, 2006.
- J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon, “Information theoretic metric learning,” in ICML, 2007.

BiDALab-UAM ©

Feature Extraction

Extract robust features from the input image for the task of face recognition.

1st Generation (1990~)

Basic Methods

Eigenfaces
Fisherfaces

2nd Generation (2000~)

Linear Subspace Methods

Sparse Representation
Metric Learning

3rd Generation (2010~)

Non-Linear Methods

Deep Learning

BiDALab-UAM ©

Deep Learning: Motivation?

Availability of large-scale face databases:

- **LFW Database:**
  - G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep., 2007.
- **IJB-A/B/C:**
  - 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 Proc. IEEE International Conference on Biometrics (ICB), 2018
- **Megaface:**
  - I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, “The megaface benchmark: 1 million faces for recognition at scale,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016.
- **Megaface 2:**
  - A. Nech, and I. Kemelmacher-Shlizerman, “Level playing field for million scale face recognition,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017.
- **MS-Celeb-1M:**
  - Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, “MS-Celeb-1M: A dataset and benchmark for large-scale face recognition,” in Proc. European Conference on Computer Vision, 2016.
- **VGGFace2:**
  - Q. Cao, L. Shen, W. Xie, O.M. Parkhi, and A. Zisserman, “VGGFace2: A dataset for recognising faces across pose and age,” in Proc. IEEE International Conference on Automatic Face & Gesture Recognition, 2018

Deep Learning: Motivation?

Availability of large-scale face databases:

Datasets	# of subjects	# of images	# of images per subject	manual identity labelling	pose	age	year
LFW [10]	5,749	13,233	1/2.3/530	-	-	-	2007
YTF [24]	1,595	3,425 videos	-	-	-	-	2011
CelebFaces+ [21]	10,177	202,599	19.9	-	-	-	2014
CASIA-WebFace [26]	10,575	494,414	2/46.8/804	-	-	-	2014
IJB-A [13]	500	5,712 images, 2,085 videos	11.4	-	-	-	2015
IJB-B [23]	1,845	11,754 images, 7,011 videos	36.2	-	-	-	2017
IJB-C [14]	3,531	31,334 images, 11,779 videos	36.3	-	-	-	2018
VGGFace [17]	2,622	2.6 M	1,000/1,000/1,000	-	-	Yes	2015
MegaFace [12]	690,572	4.7 M	3/7/2469	-	-	-	2016
MS-Celeb-1M [7]	100,000	10 M	100	-	-	-	2016
UMDFaces [5]	8,501	367,920	43.3	Yes	Yes	Yes	2016
UMDFaces-Videos [4]	3,107	22,075 videos	-	-	-	-	2017
VGGFace2 (**this paper**)	9,131	3.31 M	80/362.6/843	Yes	Yes	Yes	2018

- High variability:
  - Age
  - Gender
  - Ethnic
  - Pose, expression
  - Quality...

- Q. Cao, L. Shen, W. Xie, O.M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in Proc. IEEE International Conference on Automatic Face & Gesture Recognition, 2018

Deep Learning: Network Architectures and Loss Functions

Learning invariant and discriminative feature representations is a critical step in face recognition.

Deep larning methods have shown that compact and discriminative representations can be learned using Deep Convolutional Neural Networks (CNN) trained with very large datasets.

Deep Learning: Network Architectures and Loss Functions

Learning invariant and discriminative feature representations is a critical step in face recognition.

Deep larning methods have shown that compact and discriminative representations can be learned using Deep Convolutional Neural Networks (CNN) trained with very large datasets.

Key aspects to improve the performance of the system:

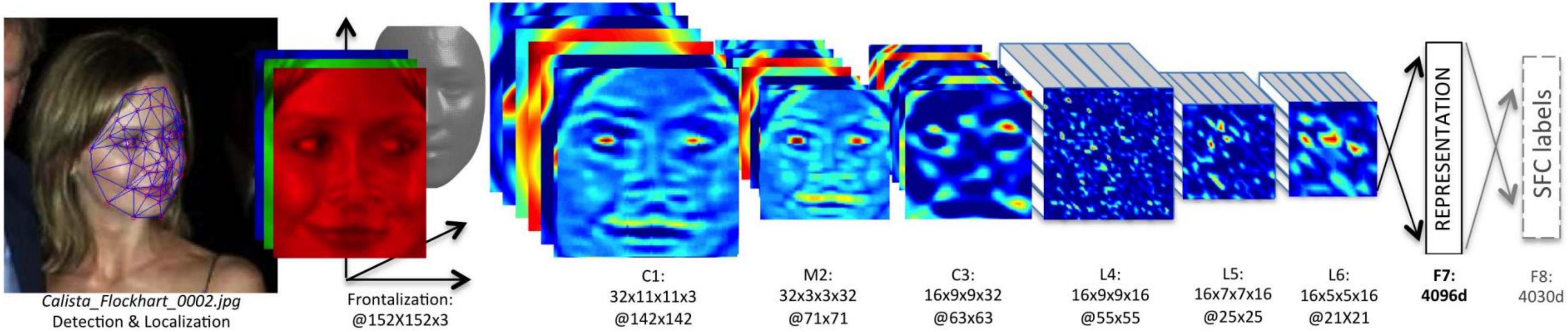
- Increase the # of subjects (and images/subject) in the training process to improve generalization. Face recognition is considered as a zero-shot learning task (system is not trained for new unseen subjects).
- Network architecture: backbone and assembled networks.
- Loss function: classical softmax loss or novel loss functions (e.g. angular/cosine distances).

BiDALab-UAM ©

Deep Learning: Network Architectures

DeepFace: Closing the Gap to Human-Level Performance in Face Verification.

- One of the earlier DCNN approaches for face recognition.
- # layers: 9 (120 million parameters).
- Dataset (Facebook, proprietary): 4M facial images belonging to more than 4K identities.



Convolution Layer: capture spatial dependencies

Max-pooling Layer: learn translation invariance

Convolution Layer: capture spatial dependencies

Locally-Connected Layers: similar to convolution layers but every location learns a different set of filters

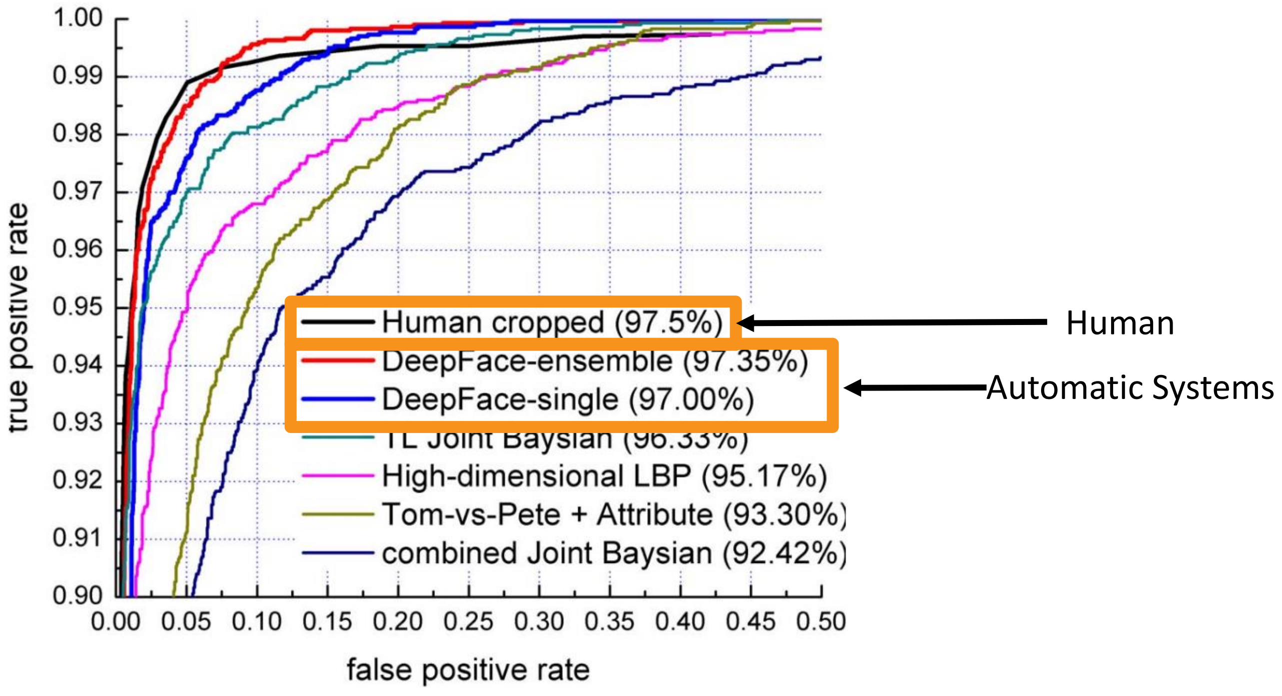
Fully-Connected Layers: capture correlations between features

- Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Proc. IEEE/CVF Conf. Computer Vision Pattern Recognition, 2014.

Deep Learning: Network Architectures

DeepFace: Closing the Gap to Human-Level Performance in Face Verification.

- One of the earlier DCNN approaches for face recognition.
- # layers: 9 (120 million parameters).
- Dataset (Facebook, proprietary): 4M facial images belonging to more than 4K identities.

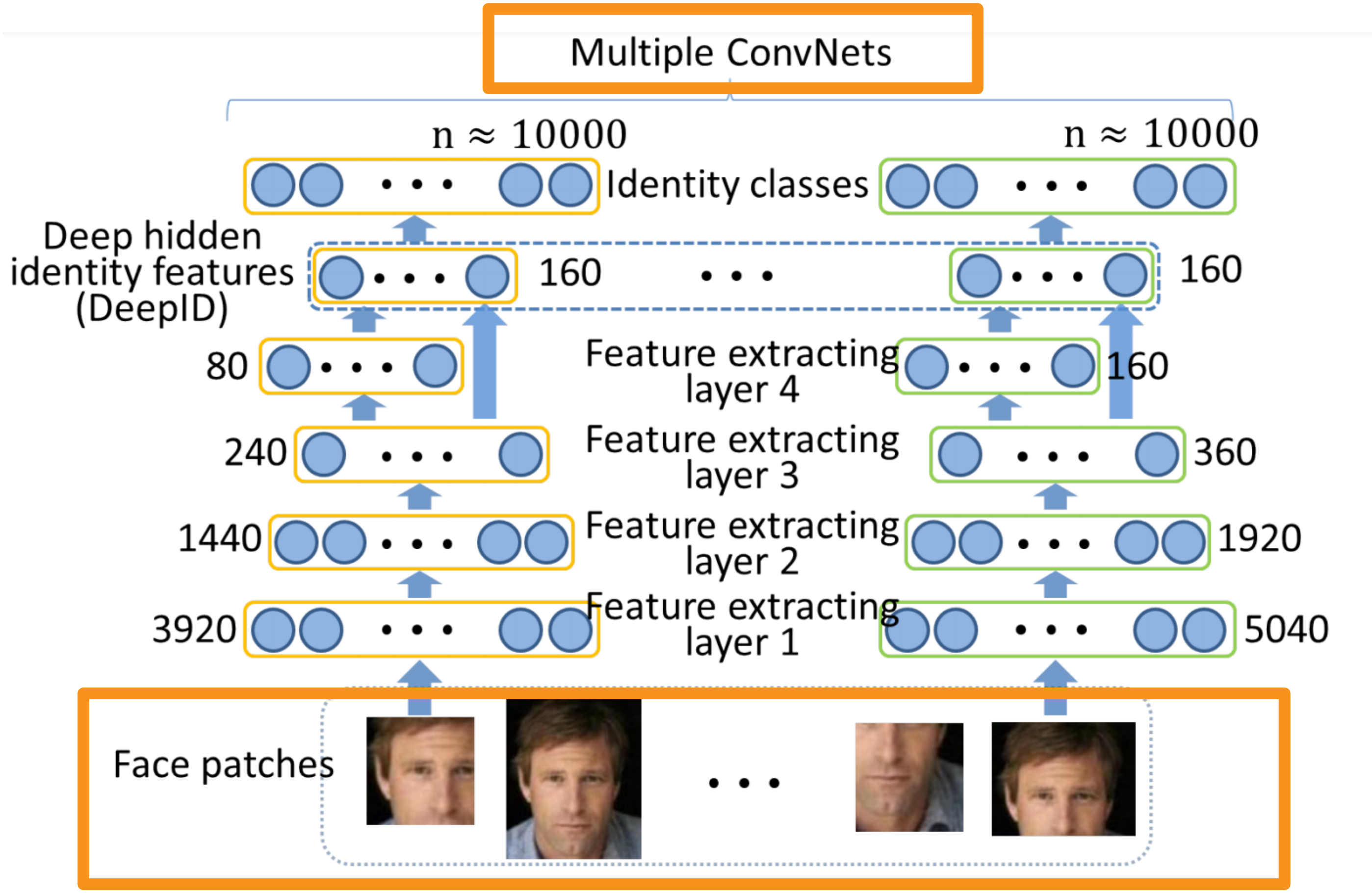


ROC curves on the LFW database

- Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Proc. IEEE/CVF Conf. Computer Vision Pattern Recognition, 2014.

Deep Learning: Network Architectures

DeepID: Ensemble of shallower and smaller DCNN (each DCNN consists of 4 convolutional layers). Face patches of sizes 39 x 31 x 1 are considered as input.



Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10000 classes," in Proc. IEEE/CVF Conf. Computer Vision Pattern Recognition, 2014

BiDALab-UAM ©

Deep Learning: Network Architectures

DeepID: Ensemble of shallower and smaller DCNN (each DCNN consists of 4 convolutional layers). Face patches of sizes 39 x 31 x 1 are considered as input.

Results on LFW database:

DeepFace [32]	97.25 (o+u)	6 + 67	4,400,000 + 3,000,000	4096 $\times$ 4
DeepID on CelebFaces	**96.05** (o)	5	87,628	150
DeepID on CelebFaces+	**97.20** (o)	5	202,599	150
DeepID on CelebFaces+ & TL	**97.45** (o+u)	5	202,599	150

Fewer # training images needed

Smaller feature dimension

Similar results compared with Human recognition (97.5%).

Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10000 classes," in Proc. IEEE/CVF Conf. Computer Vision Pattern Recognition, 2014

BiDALab-UAM ©

Deep Learning: Network Architectures

DeepID2: Ensemble of shallower and smaller DCNN similar to DeepID.

Key aspect: training process is based on identification and verification signals.

- Identification signal: increases the inter-personal variations. The network is trained to minimize the cross-entropy loss (e.g. using 8192 identities).

$$\mathrm{Ident}(f, t, \theta_{id}) = -\sum_{i=1}^{n} -p_i \log \hat{p}_i = -\log \hat{p}_t ,$$

- Verification signal: reduces the intra-personal variations

$$\mathrm{Verif}(f_i, f_j, y_{ij}, \theta_{ve}) = \begin{cases} \frac{1}{2} \left\| f_i - f_j \right\|_2^2 & \text{if } y_{ij} = 1 \\ \frac{1}{2} \max\left(0, m - \left\| f_i - f_j \right\|_2\right)^2 & \text{if } y_{ij} = -1 \end{cases} ,$$

Y. Sun, X. Wang, and X. Tang, “Deep learning face representation from predicting 10000 classes,” in Proc. IEEE/CVF Conf. Computer Vision Pattern Recognition, 2014

BiDALab-UAM ©

Deep Learning: Network Architectures

DeepID2: Ensemble of shallower and smaller DCNN similar to DeepID.

Results on LFW database:

method	accuracy (%)
high-dim LBP [4]	95.17 ± 1.13
TL Joint Bayesian [2]	96.33 ± 1.08
DeepFace [22]	97.35 ± 0.25
DeepID [21]	97.45 ± 0.26
GaussianFace [14]	98.52 ± 0.66
DeepID2	99.15 ± 0.13

First automatic system that outperforms Human recognition (99.15% vs. 97.50% accuracy).

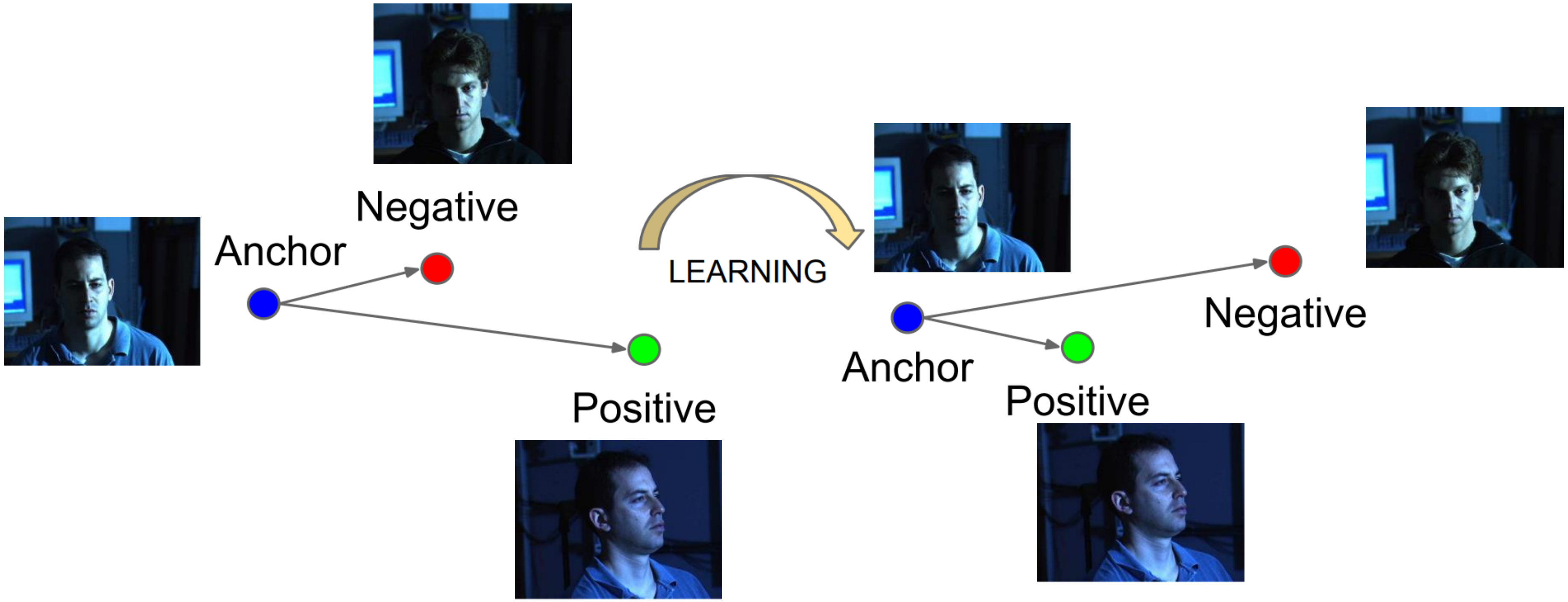
Improvement? The training process is the key!!! (DeepID2 same architecture than DeepID).

- Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” In Proc. Advances in neural information processing systems, 2014.

Deep Learning: Network Architectures

FaceNet: Training using triplets of roughly aligned matching/non-matching face patches.

- DCNN trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer.
- Better generalization to new unseen identities and with smaller feature dimension (128 vs. 4,096).



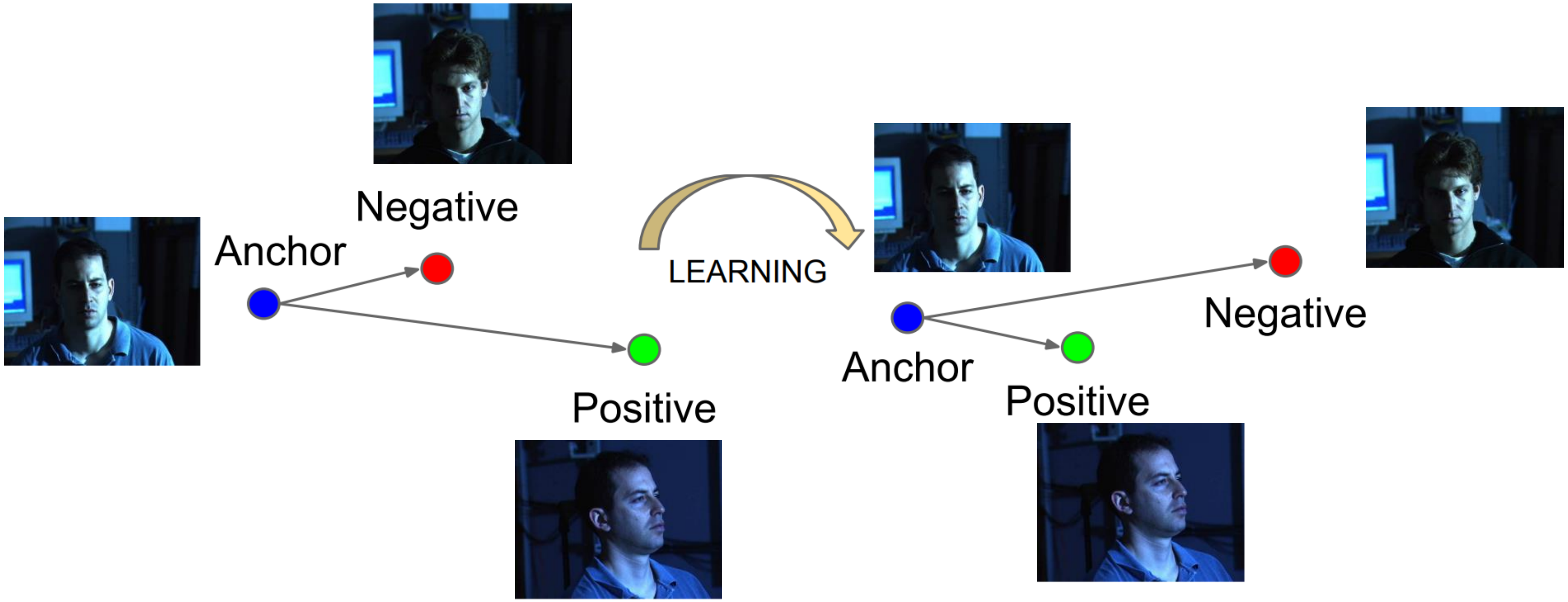
- F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proc. IEEE/CVF Conference on Computer Vision and Patter Recognition, 2015

BiDALab-UAM ©

Deep Learning: Network Architectures

FaceNet: Training using triplets of roughly aligned matching/non-matching face patches.

- DCNN trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer.
- Better generalization to new unseen identities and with smaller feature dimension (128 vs. 4,096).



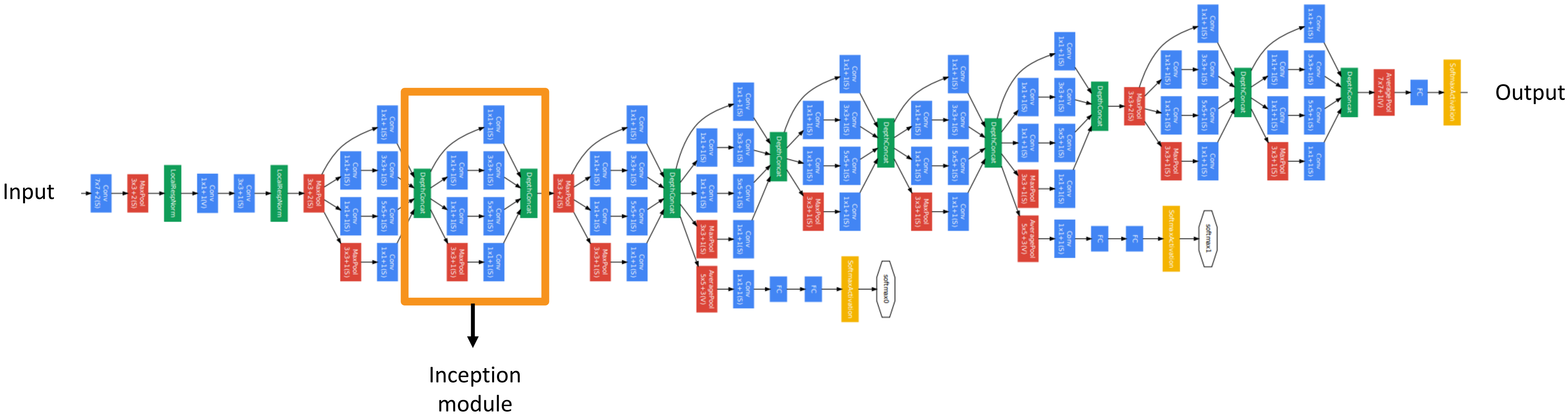
Key: selecting hard triplets, that are active and can therefore contribute to improve the model.

- F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in Proc. IEEE/CVF Conference on Computer Vision and Patter Recognition, 2015

Deep Learning: Network Architectures

FaceNet: They considered a deeper CNN (Inception GoogLeNet).

- Trained using 100-200 million faces of about 8 million different identities (Google private dataset).



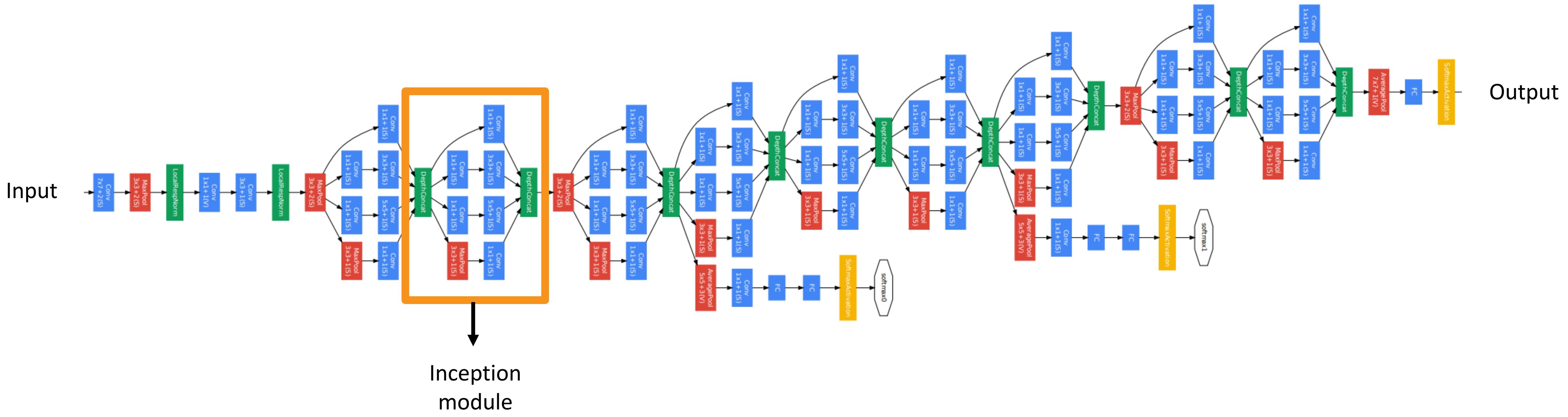
- C. Szegedy, W. Liu, Y. Jia, et al.,"Going deeper with convolutions." in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2015.

BiDALab-UAM ©

Deep Learning: Network Architectures

FaceNet: They considered a deeper CNN (Inception GoogLeNet).

- Trained using 100-200 million faces of about 8 million different identities (Google private dataset).

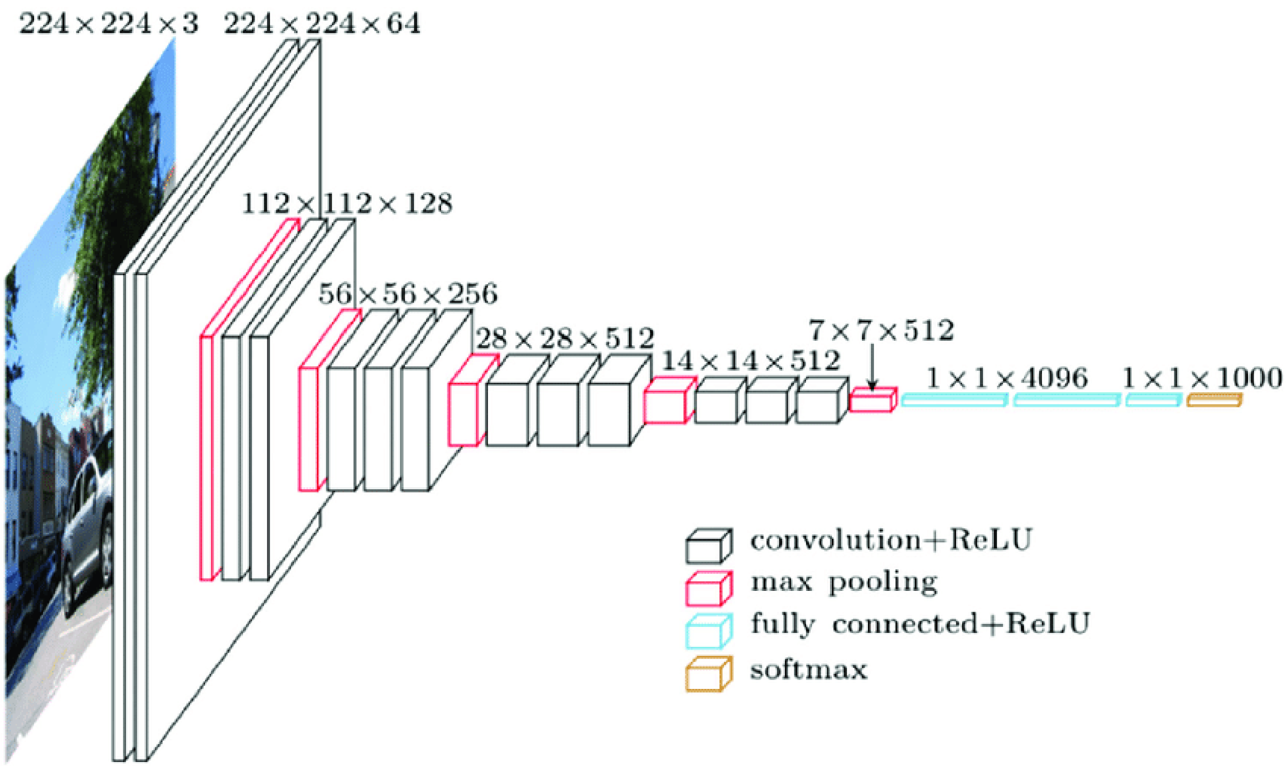


Results on LFW database: 99.63% accuracy (DeepID=99.15%).

• C. Szegedy, W. Liu, Y. Jia, et al.,"Going deeper with convolutions." in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2015.

Deep Learning: Network Architectures

VGGNet: They trained a DCNN based on the well-known VGGNet for object recognition, followed by triplet embedding for face verification.



- O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in Proc. British Machine Vision Conference, 2015.
- K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conference on Learning Representations 2014.

Deep Learning: Network Architectures

VGGNet: They trained a DCNN based on the well-known VGGNet for object recognition, followed by triplet embedding for face verification.

They presented and used for training VGGFace database.

Dataset	Identities	Images
Ours	2,622	2.6M
FaceBook [29]	4,030	4.4M
Google [17]	8M	200M

Public

Private

Celebrities from Internet

- O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in Proc. British Machine Vision Conference, 2015.
- K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conference on Learning Representations 2014.

BiDALab-UAM ©

Deep Learning: Network Architectures

VGGNet: They trained a DCNN based on the well-known VGGNet for object recognition, followed by triplet embedding for face verification.

They presented and used for training VGGFace database.

Results on LFW database:

No.	Method	Images	Networks	Acc.
1	Fisher Vector Faces [21]	-	-	93.10
2	DeepFace [29]	4M	3	97.35
3	Fusion [30]	500M	5	98.37
4	DeepID-2,3		200	99.47
5	FaceNet [17]	200M	1	98.87
6	FaceNet [17] + Alignment	200M	1	99.63
7	Ours	2.6M	1	98.95

- O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. British Machine Vision Conference, 2015.
- K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conference on Learning Representations 2014.

BiDALab-UAM ©

Deep Learning: Network Architectures

ArcFace: incorporate margins in well-stablished loss functions in order to maximize face class separability. They proposed an Additive Angular Margin Loss (ArcFace) to obtain highly discriminative features for face recognition.

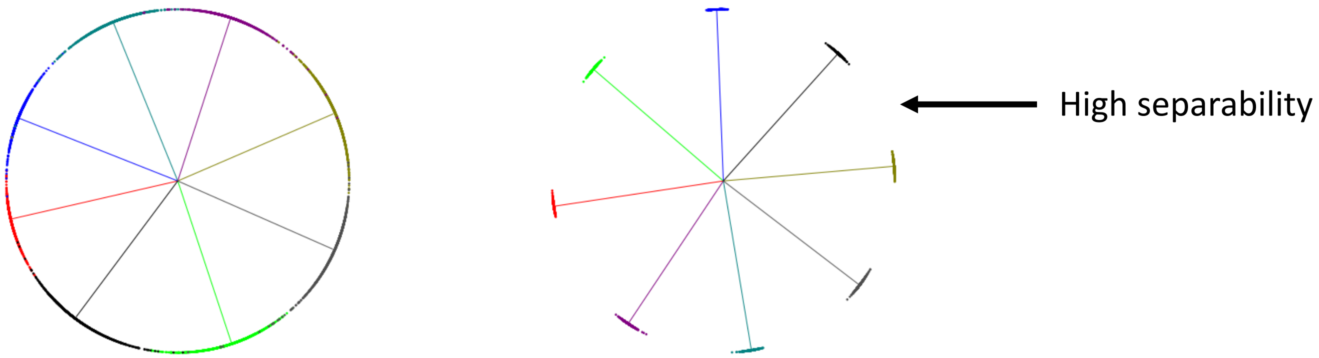
- J. Deng, J. Guo, N. Xue, S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition” In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

BiDALab-UAM ©

Deep Learning: Network Architectures

ArcFace: incorporate margins in well-stablished loss functions in order to maximize face class separability. They proposed an Additive Angular Margin Loss (ArcFace) to obtain highly discriminative features for face recognition.

ArcFace has a clear geometric interpretation due to geodesic distance on a hypersphere.



(a) Softmax (b) ArcFace

Toy examples under the softmax and ArcFace loss on 8 identities with 2D features. Dots indicate samples and lines refer to the centre direction of each identity. Based on the feature normalization, all face features are pushed to the arc space with a fixed radius. The geodesic distance gap between closest classes becomes evident as the additive angular margin penalty is incorporated.

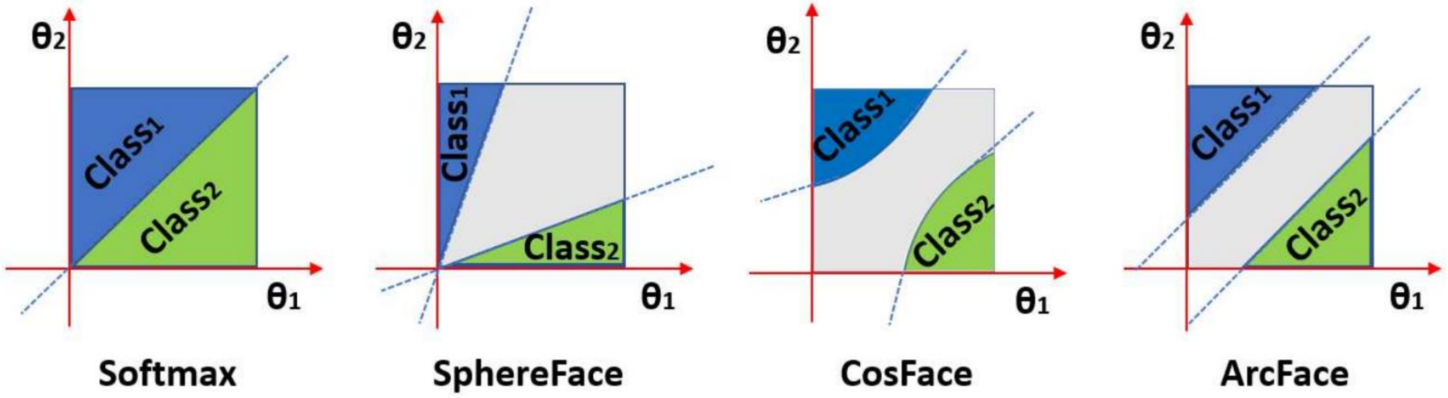
• J. Deng, J. Guo, N. Xue, S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition" In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

BiDALab-UAM ©

Deep Learning: Network Architectures

ArcFace: incorporate margins in well-stablished loss functions in order to maximize face class separability. They proposed an Additive Angular Margin Loss (ArcFace) to obtain highly discriminative features for face recognition.

ArcFace has a constant linear angular margin throughout the whole interval. By contrast, other loss functions such as SphereFace and CosFace only have a nonlinear angular margin.



- J. Deng, J. Guo, N. Xue, S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition" In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

BiDALab-UAM ©

Deep Learning: Network Architectures

ArcFace: They considered deeper DCNN (ResNet50 and ResNet100) popular in object recognition.

Deeper networks are more difficult to train:

- Degradation problem: with the network depth increasing, accuracy gets saturated (which might be unsurprising) and then degrades rapidly (and not caused by overfitting).

- K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016

Deep Learning: Network Architectures

ArcFace: They considered deeper DCNN (ResNet50 and ResNet100) popular in object recognition.

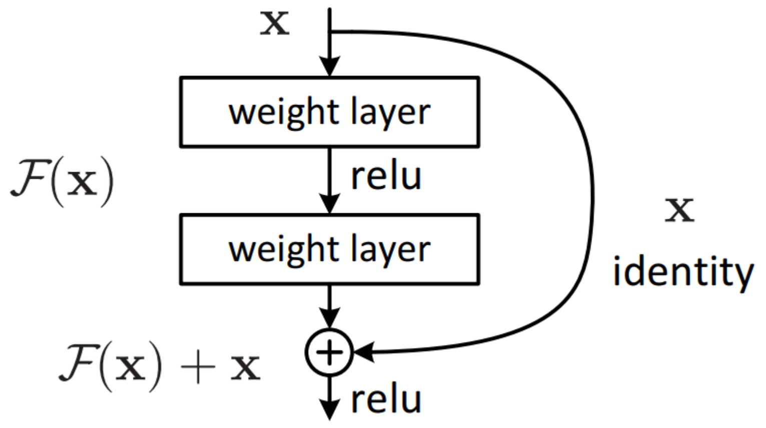
Deeper networks are more difficult to train:

- Degradation problem: with the network depth increasing, accuracy gets saturated (which might be unsurprising) and then degrades rapidly (and not caused by overfitting).

ResNet50 and ResNet100 presented a residual learning framework to ease the training of the networks.

Adding shortcut connections (skipping one or more layers):

- Simply perform identity mapping, and their outputs are added to the outputs of the stacked layers.



Residual learning: a building block

- K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016

Deep Learning: Network Architectures

ArcFace: They considered deeper DCNN (ResNet50 and ResNet100) popular in object recognition.

Results:

Method	#Image	LFW	YTF
DeepID [30]	0.2M	99.47	93.20
Deep Face [31]	4.4M	97.35	91.4
VGG Face [22]	2.6M	98.95	97.30
FaceNet [27]	200M	99.63	95.10
Baidu [13]	1.3M	99.13	-
Center Loss [36]	0.7M	99.28	94.9
Range Loss [43]	5M	99.52	93.70
Marginal Loss [6]	3.8M	99.48	95.98
SphereFace [15]	0.5M	99.42	95.0
SphereFace+ [14]	0.5M	99.47	-
CosFace [35]	5M	99.73	97.6
MS1MV2, R100, ArcFace	5.8M	**99.83**	**98.02**

Method	LFW	CALFW	CPLFW
HUMAN-Individual	97.27	82.32	81.21
HUMAN-Fusion	99.85	86.50	85.24
Center Loss [36]	98.75	85.48	77.48
SphereFace [15]	99.27	90.30	81.40
VGGFace2 [3]	99.43	90.57	84.00
MS1MV2, R100, ArcFace	**99.82**	**95.45**	**92.08**

- J. Deng, J. Guo, N. Xue, S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition" In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.

BiDALab-UAM ©

Deep Learning: Network Architectures

And many more...

Method	Public. Time	Loss	Architecture	Number of Networks	Training Set	Accuracy±Std(%)
DeepFace [20]	2014	softmax	Alexnet	3	Facebook (4.4M,4K)	97.35±0.25
DeepID2 [21]	2014	contrastive loss	Alexnet	25	CelebFaces+ (0.2M,10K)	99.15±0.13
DeepID3 [36]	2015	contrastive loss	VGGNet-10	50	CelebFaces+ (0.2M,10K)	99.53±0.10
FaceNet [38]	2015	triplet loss	GoogleNet-24	1	Google (500M,10M)	99.63±0.09
Baidu [58]	2015	triplet loss	CNN-9	10	Baidu (1.2M,18K)	99.77
VGGface [37]	2015	triplet loss	VGGNet-16	1	VGGface (2.6M,2.6K)	98.95
light-CNN [85]	2015	softmax	light CNN	1	MS-Celeb-1M (8.4M,100K)	98.8
Center Loss [101]	2016	center loss	Lenet+-7	1	CASIA-WebFace, CACD2000, Celebrity+ (0.7M,17K)	99.28
L-softmax [104]	2016	L-softmax	VGGNet-18	1	CASIA-WebFace (0.49M,10K)	98.71
Range Loss [82]	2016	range loss	VGGNet-16	1	MS-Celeb-1M, CASIA-WebFace (5M,100K)	99.52
L2-softmax [109]	2017	L2-softmax	ResNet-101	1	MS-Celeb-1M (3.7M,58K)	99.78
Normface [110]	2017	contrastive loss	ResNet-28	1	CASIA-WebFace (0.49M,10K)	99.19
CoCo loss [112]	2017	CoCo loss	-	1	MS-Celeb-1M (3M,80K)	99.86
vMF loss [115]	2017	vMF loss	ResNet-27	1	MS-Celeb-1M (4.6M,60K)	99.58
Marginal Loss [116]	2017	marginal loss	ResNet-27	1	MS-Celeb-1M (4M,80K)	99.48
SphereFace [84]	2017	A-softmax	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.42
CCL [113]	2018	center invariant loss	ResNet-27	1	CASIA-WebFace (0.49M,10K)	99.12
AMS loss [105]	2018	AMS loss	ResNet-20	1	CASIA-WebFace (0.49M,10K)	99.12
Cosface [107]	2018	cosface	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.33
Arcface [106]	2018	arcface	ResNet-100	1	MS-Celeb-1M (3.8M,85K)	99.83
Ring loss [117]	2018	Ring loss	ResNet-64	1	MS-Celeb-1M (3.5M,31K)	99.50

- M. Wang, and W. Deng, "Deep Face Recognition: A Survey," arXiv preprint arXiv:1804.06655, 2020.

SOFT BIOMETRICS

AGING

OTHER RESEARCH LINES RELATED TO FACE RECOGNITION

FAIRNESS

DEEPFAKES

Fairness in Face Recognition

Face recognition systems usually have gender and racial bias:

CBS NEWS

NEWS | 2020 ELECTIONS | SHOWS | LIVE

Amazon face-detection technology shows gender and racial bias, researchers say

JANUARY 25, 2019 / 8:25 PM / AP

Facial-detection technology that Amazon is marketing to law enforcement often misidentifies women, particularly those with darker skin, according to researchers from MIT and the University of Toronto. Privacy and civil rights advocates have called on Amazon to stop marketing its Rekognition service because of worries about discrimination against minorities.

Fairness in Face Recognition

Face recognition systems usually have gender and racial bias:

CBS NEWS

NEWS 2020 ELECTIONS SHOWS LIVE

Amazon face-detection technology shows gender and racial bias, researchers say

JANUARY 25, 2019 / 8:25 PM / AP

Facial-detection technology that Amazon is marketing to law enforcement often misidentifies women, particularly those with darker skin, according to researchers from MIT and the University of Toronto. Privacy and civil rights advocates have called on Amazon to stop marketing its Rekognition service because of worries about discrimination against minorities.

Why? Training databases are usually unbalanced.

BiDALab-UAM ©

Fairness in Face Recognition

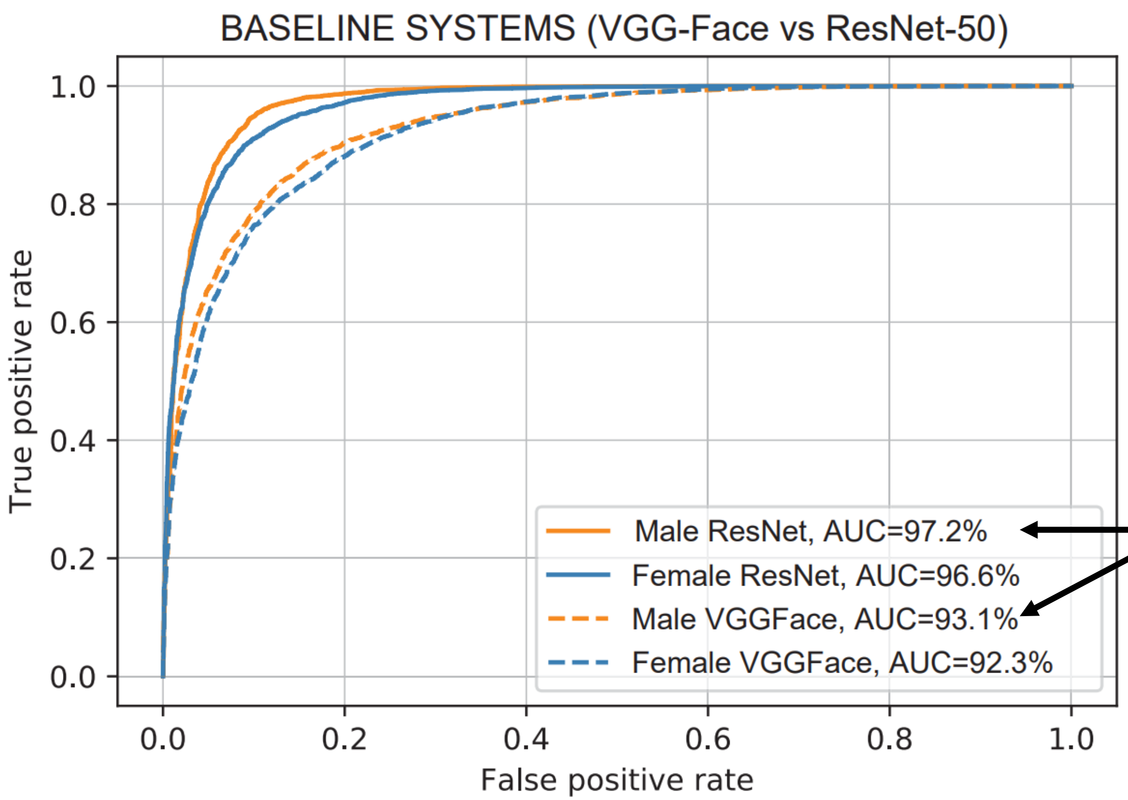
FaceGenderID: Gender-dependent training approach.

- R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proença, “FaceGenderID: exploiting gender information in DCNNs face recognition systems”. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019

Fairness in Face Recognition

FaceGenderID: Gender-dependent training approach.

- DCNN: VGGFace and ResNet50 (pre-trained models).



Male: better results

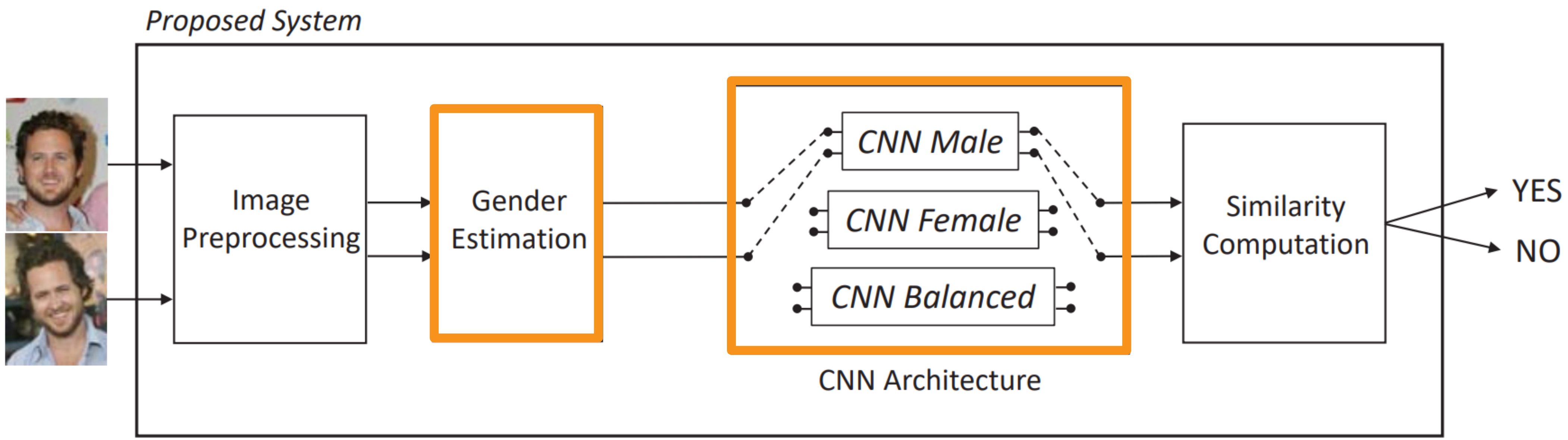
- O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in Proc. of the British Machine Vision Conference, 2015.
- Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “Vggface2: A dataset for recognising faces across pose and age,” in Proc. International Conf. on Automatic Face and Gesture Recognition, 2018.

Fairness in Face Recognition

Biometrics & Data Pattern Analytics Lab

FaceGenderID: Gender-dependent training approach.

- **DCNN**: VGGFace and ResNet50 (pre-trained models).



Architecture of the proposed approach for gender specific DCNNs models for face verification

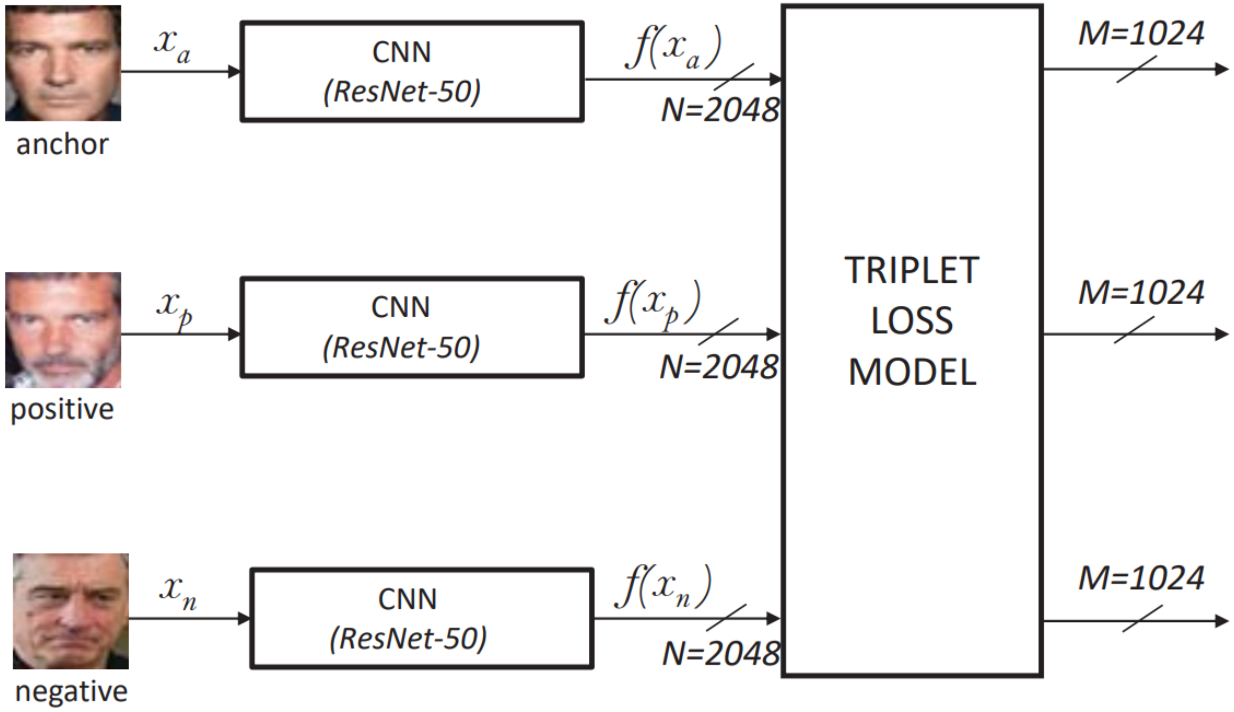
- R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proença, "FaceGenderID: exploiting gender information in DCNNs face recognition systems". In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019

Fairness in Face Recognition

BiDA Lab
Biometrics & Data Pattern Analytics Lab

FaceGenderID: Gender-dependent training approach.

- DCNN: VGGFace and ResNet50 (pre-trained models).
- Gender-dependent DCNN: trained using Triplet Loss over the embeddings.



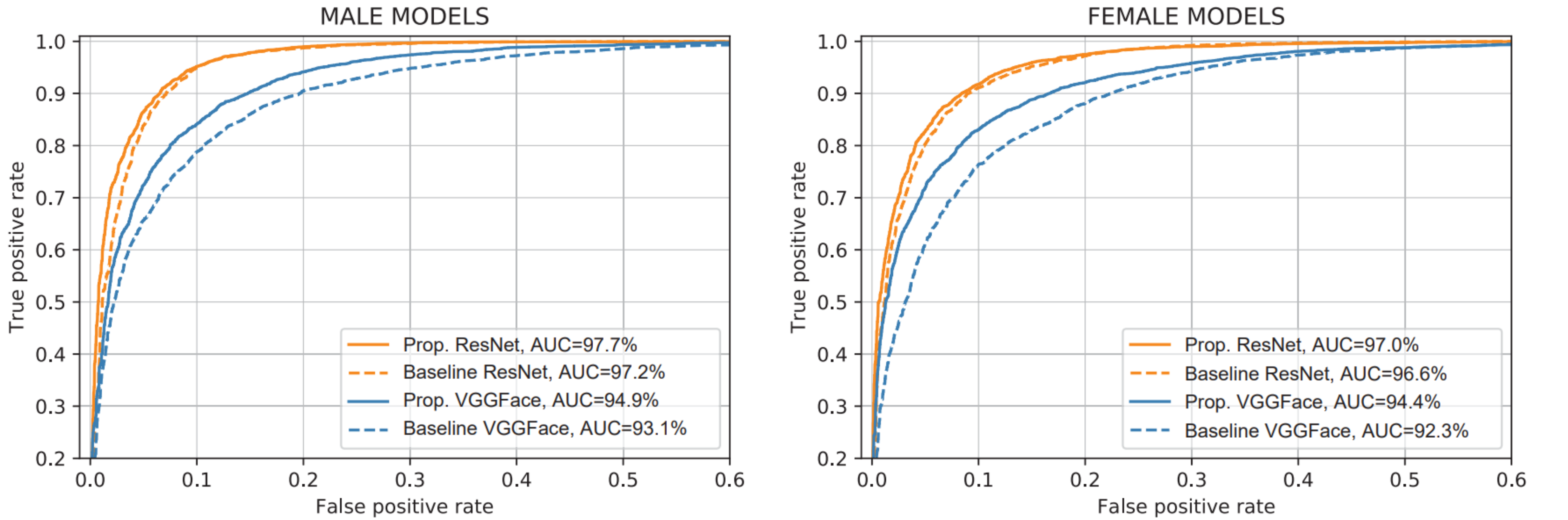
- R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proença, "FaceGenderID: exploiting gender information in DCNNs face recognition systems". In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019

Fairness in Face Recognition

BiDA Lab
Biometrics & Data Pattern Analytics Lab

FaceGenderID: Gender-dependent training approach.

- DCNN: VGGFace and ResNet50 (pre-trained models).
- Gender-dependent DCNN: trained using Triplet Loss over the embeddings.



(a) Male Gender (b) Female Gender

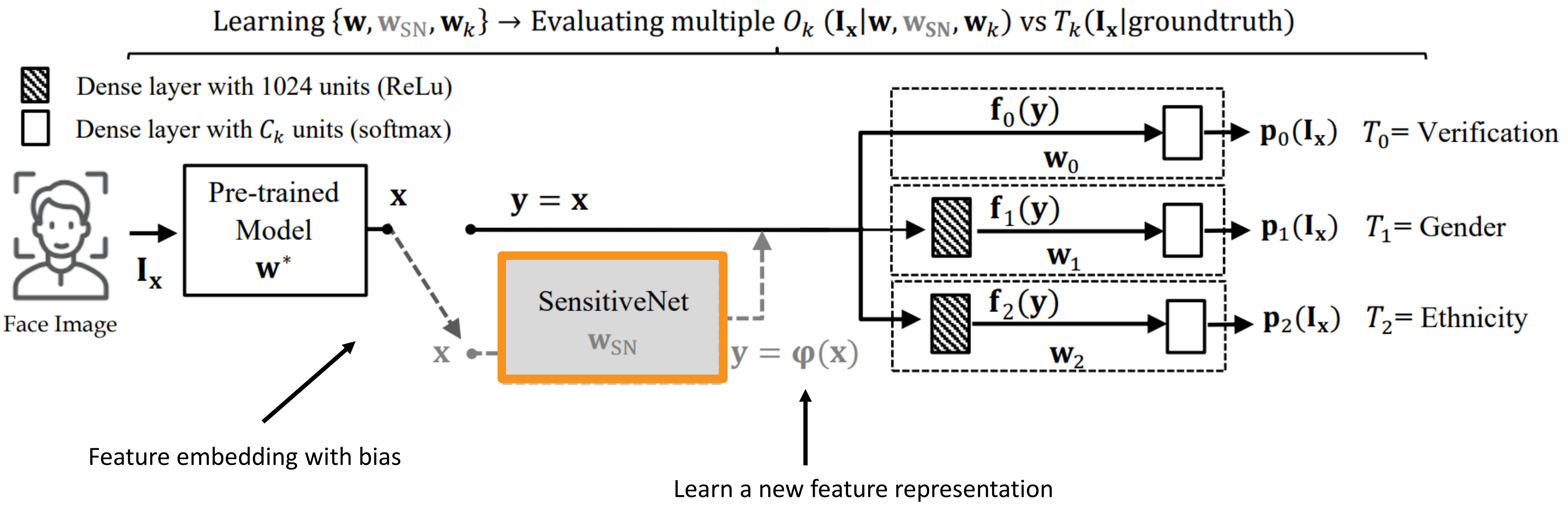
Better results in terms of performance and gender bias!

- R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proença, "FaceGenderID: exploiting gender information in DCNNs face recognition systems". In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019

Fairness in Face Recognition

SensitiveNets: Learning agnostic representations.

- Novel privacy-preserving neural network feature representation to suppress the sensitive information of a learned space while maintaining the utility of the data.



- A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning agnostic representations with application to face images" IEEE Transactions on Pattern Analysis and Machine Intelligence.

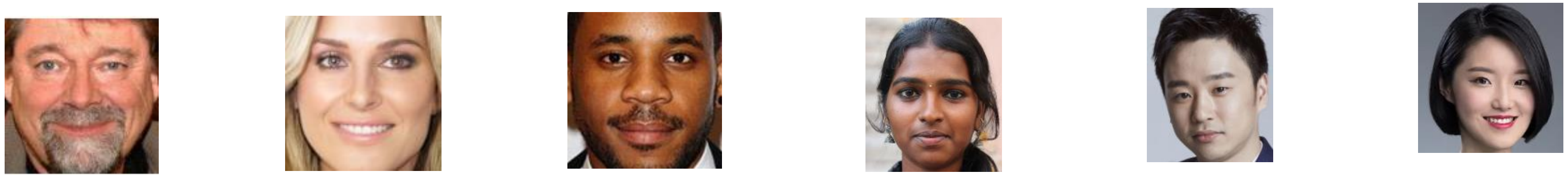
Fairness in Face Recognition

BiDA Lab
Biometrics & Data Pattern Analytics Lab

SensitiveNets: Learning agnostic representations.

- DiveFace database: dataset for diversity-aware face recognition (generated from Megaface2).
  - Gender: Male/Female.
  - Ethnic physical characteristics: 3 groups.

Baseline experiment: VGGFace and ResNet50



Model	Group 1		Group 2		Group 3	
	Male	Female	**Male**	Female	**Male**	Female
VGG-Face	7.99	9.38 (↑17%)	12.03 (↑50%)	13.95 (↑76%)	18.43 (↑131%)	23.66 (↑196%)
ResNet-50	1.60	1.96 (↑22%)	2.15 (↑34%)	3.61 (↑126%)	3.25 (↑103%)	5.07 (↑217%)

Table 1: Performance (False Match Rate in % @ False Non-Match Rate = 0.1%) of Face Recognition Models on the DiveFace dataset. We show in brackets the relative error growth rates with respect to the best class (*Group 1 Male*).

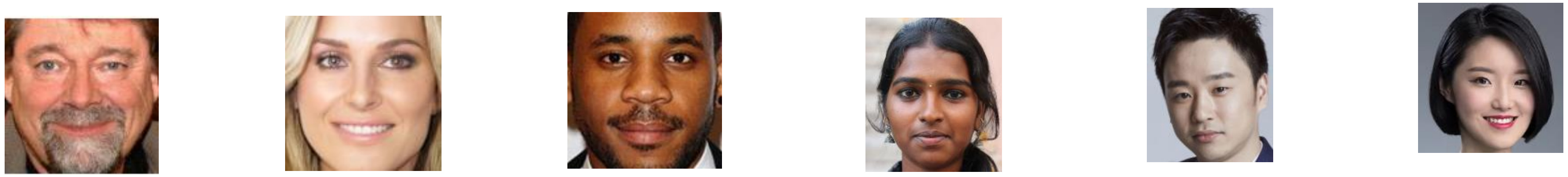
- I. Serna, A. Morales, J. Fierrez, et al., "Algorithmic Discrimination: Formulation and Exploration in Deep Learning-based Face Biometrics," in Proc. of AAAI Workshop on Artificial Intelligence Safety, 2020.

Fairness in Face Recognition

SensitiveNets: Learning agnostic representations.

- DiveFace database: dataset for diversity-aware face recognition (generated from Megaface2).
  - Gender: Male/Female.
  - Ethnicity groups: 3 groups.

Baseline experiment: VGGFace and ResNet50.



Model	Group 1		Group 2		Group 3	
	Male	Female	**Male**	Female	**Male**	Female
VGG-Face	7.99	9.38 (↑17%)	12.03 (↑50%)	13.95 (↑76%)	18.43 (↑131%)	23.66 (↑196%)
ResNet-50	1.60	1.96 (↑22%)	2.15 (↑34%)	3.61 (↑126%)	3.25 (↑103%)	5.07 (↑217%)

Table 1: Performance (False Match Rate in % @ False Non-Match Rate = 0.1%) of Face Recognition Models on the DiveFace dataset. We show in brackets the relative error growth rates with respect to the best class (*Group 1 Male*).

High difference in the system performance among gender and demographic groups!

- I. Serna, A. Morales, J. Fierrez, et al., "Algorithmic Discrimination: Formulation and Exploration in Deep Learning-based Face Biometrics," in Proc. of AAAI Workshop on Artificial Intelligence Safety, 2020.

Fairness in Face Recognition

SensitiveNets: Learning agnostic representations.

- Triplet loss learning with an adversarial sensitive regularizer used to measure the amount of sensitive information present in the model.

$$\mathcal{L}_{\mathrm{SN}} = \min_{\mathbf{w}_{\mathrm{SN}}} \sum_{i \in \mathcal{T}} \left[\mathcal{L}'_0 \left(\boldsymbol{\varphi}(\mathbf{x}_{\mathrm{A}}^i, \mathbf{x}_{\mathrm{P}}^i, \mathbf{x}_{\mathrm{N}}^i | \mathbf{w}_{\mathrm{SN}}) \right) + \right.$$
$$\left. + \Lambda_{\mathrm{A}}^i + \Lambda_{\mathrm{P}}^i + \Lambda_{\mathrm{N}}^i \right],$$

SensitiveNets

$$\mathcal{L}'_0 = \|\boldsymbol{\varphi}(\mathbf{x}_{\mathrm{A}}^i) - \boldsymbol{\varphi}(\mathbf{x}_{\mathrm{P}}^i)\|^2 - \|\boldsymbol{\varphi}(\mathbf{x}_{\mathrm{A}}^i) - \boldsymbol{\varphi}(\mathbf{x}_{\mathrm{N}}^i)\|^2 + \alpha,$$

Triplet loss function

$$\Lambda^i(\mathbf{x}^i) = \log\left(1 + \left|0.9 - P_k\left(D^i \middle| \boldsymbol{\varphi}\left(\mathbf{x}^i \middle| \mathbf{w}^*, \mathbf{w}_{\mathrm{SN}}\right), \mathbf{w}_k^*\right)\right|\right)$$

Adversarial regularizer

- A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "SensitiveNets: Learning agnostic representations with application to face images" IEEE Transactions on Pattern Analysis and Machine Intelligence.

Fairness in Face Recognition

BiDA Lab

Biometrics & Data Pattern Analytics Lab

SensitiveNets: Learning agnostic representations.

- Triplet loss learning with an adversarial sensitive regularizer used to measure the amount of sensitive information present in the model.

- Results on LFW database: keeping similar recognition performance while reducing gender/ethnicity bias!

TABLE I. CLASSIFICATION ACCURACIES FOR EACH TASK BEFORE AND AFTER APPLYING THE PROJECTION INTO THE NEW FEATURE REPRESENTATION. RECOGNITION REPRESENTS FACE VERIFICATION ACCURACY (IN %).

Task	Before	After	Reduction*	Random
Recognition	98.4%	95.8%	5.4%	50%
Neural Network (NN)				
Gender	97.7%	58.8%	81.5%	50%
Ethnicity	98.8%	55.1%	66.4%	33%
Support Vector Machine (SVM)				
Gender	96.2%	56.3%	86.4%	50%
Ethnicity	98.2%	54.1%	67.6%	33%
Random Forest (RF)				
Gender	95.1%	54.6%	89.8%	50%
Ethnicity	97.3%	53.5%	68.1%	33%

*Reduction = (Before-After)/(Before-Random)

- A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, “SensitiveNets: Learning agnostic representations with application to face images” IEEE Transactions on Pattern Analysis and Machine Intelligence.

Aging in Face Recognition

One of the challenges in automatic face recognition is to achieve temporal invariance. In other words, the goal is to come up with a representation and matching scheme that is robust to changes due to facial aging.

- U. Park, Y. Tong, and A.K. Jain, “Age-invariant face recognition” IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 947-954, 2010.
- M.M. Sawant, and K.M. Bhurchandi, “Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging,” Artificial Intelligence Review, 52(2), 981-1008, 2019.

Aging in Face Recognition

One of the challenges in automatic face recognition is to achieve temporal invariance. In other words, the goal is to come up with a representation and matching scheme that is robust to changes due to facial aging.

Madeleine McCann
Disappeared in 2003 at the age of 9

?

How she looks actually?

- U. Park, Y. Tong, and A.K. Jain, “Age-invariant face recognition” IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 947-954, 2010.
- M.M. Sawant, and K.M. Bhurchandi, “Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging,” Artificial Intelligence Review, 52(2), 981-1008, 2019.

BiDALab-UAM ©

Aging in Face Recognition

One of the challenges in automatic face recognition is to achieve temporal invariance. In other words, the goal is to come up with a representation and matching scheme that is robust to changes due to facial aging.

Madeleine McCann
Disappeared in 2003 at the age of 3

?

How she looks actually?

Forensic artist's impression of how she may have looked in 2012, aged 9

- U. Park, Y. Tong, and A.K. Jain, "Age-invariant face recognition" IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 947-954, 2010.
- M.M. Sawant, and K.M. Bhurchandi, "Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging," Artificial Intelligence Review, 52(2), 981-1008, 2019.

BiDALab-UAM ©

Aging in Face Recognition

Kinship verification: It is defined as a relationship between two persons who are biologically related with overlapping genes (father, mother, son and daughter).

Madeleine McCann
Disappeared in 2003 at the age of 9

?

How she looks actually?

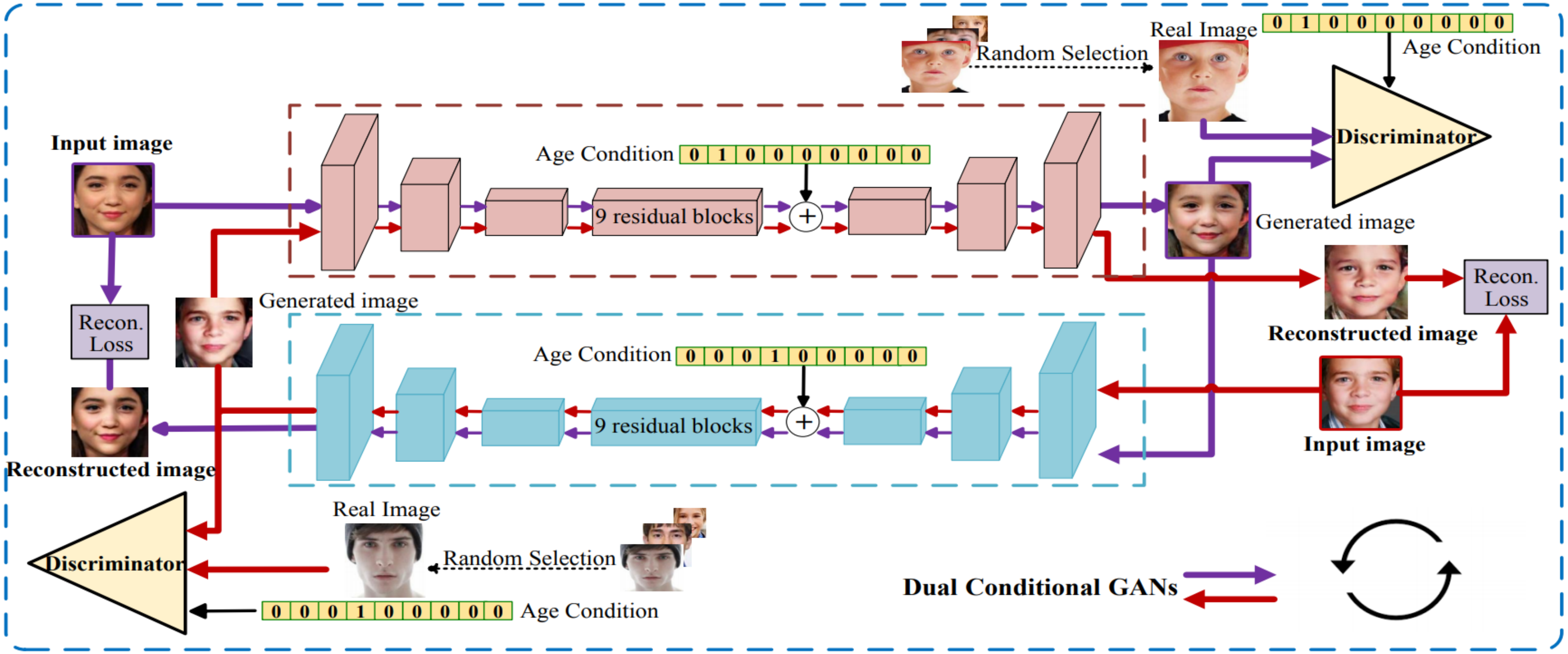
She looks similar to her parents

- U. Park, Y. Tong, and A.K. Jain, "Age-invariant face recognition" IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 947-954, 2010.
- M.M. Sawant, and K.M. Bhurchandi, "Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging," Artificial Intelligence Review, 52(2), 981-1008, 2019.

BiDALab-UAM ©

Aging in Face Recognition

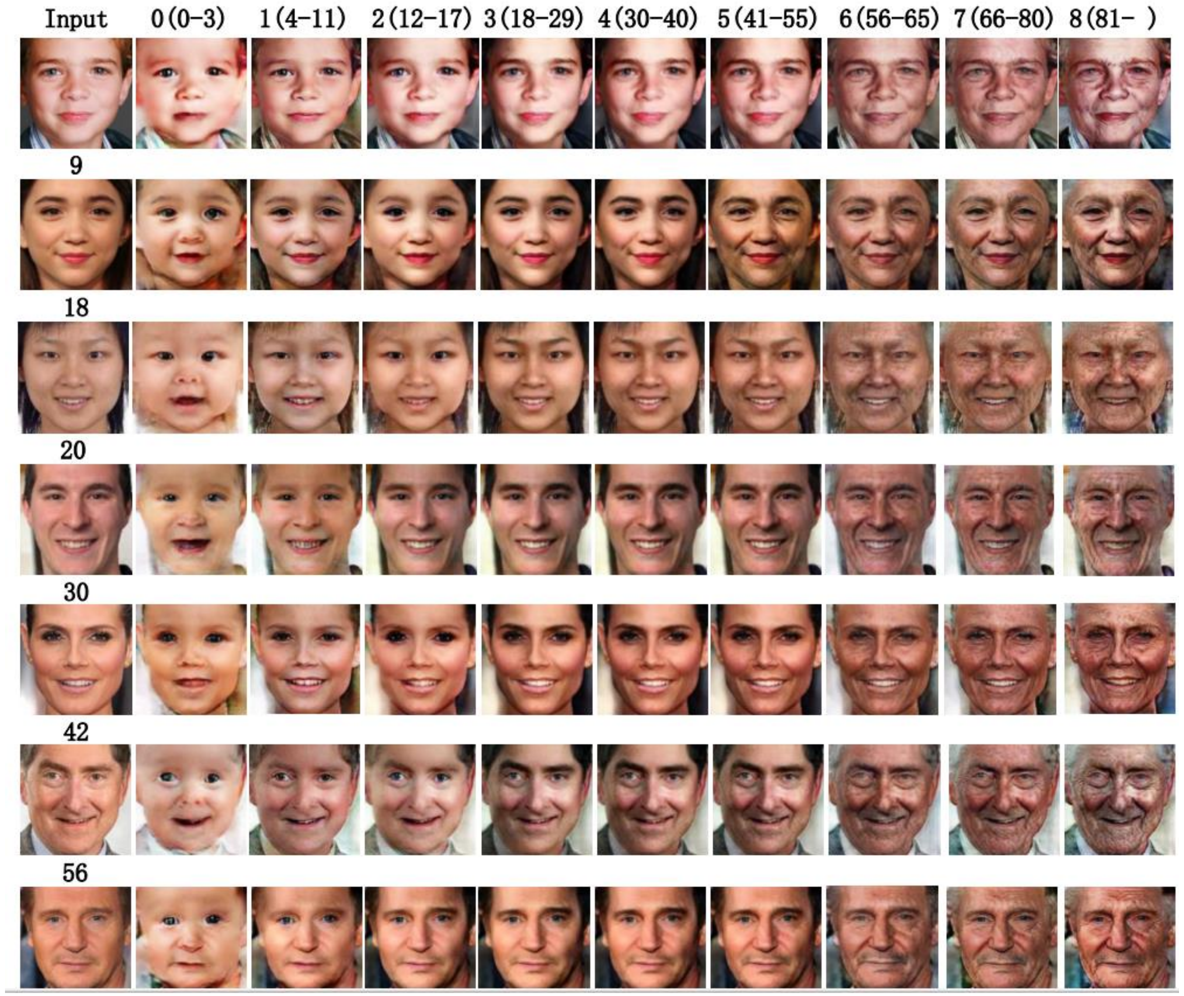
Generative Adversarial Networks: based on a Generator and Discriminator



- J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, "Dual Conditional GANs for Face Aging and Rejuvenation," in Proc. International Joint Conference on Artificial Intelligence, 2018

Aging in Face Recognition

Generative Adversarial Networks: based on a Generator and Discriminator



- J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, "Dual Conditional GANs for Face Aging and Rejuvenation," in Proc. International Joint Conference on Artificial Intelligence, 2018

Soft Biometrics

Personal attributes like gender, ethnicity, age, height, weight, eye color, glasses, scars, marks, and tatoos are examples of soft biometric traits.

SOFT BIOMETRICS

- E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, “Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation,” IEEE Transactions on Information Forensics and Security, 13(8), 2001-2014, 2018.

BiDALab-UAM ©

Soft Biometrics

Applications:

- Surveillance (unconstrained scenarios).
- Complementary to face recognition.

- E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, “Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation,” IEEE Transactions on Information Forensics and Security, 13(8), 2001-2014, 2018.

Soft Biometrics

BiDA Lab
Biometrics & Data Pattern Analytics Lab

Results on LFW database in terms of Equal Error Rate (%):

Performance of Soft Biometrics		Face		Fusion	
Set of Soft Biometrics		Face++	VGG-face	Face++	VGG-face
Age	50.6 ± 3.1	12.7 ± 1.4	7.8 ± 1.2	10.9 ± 1.4	7.1 ± 0.7
Age Ethnicity	31.1 ± 3.9			9.0 ± 1.2	5.8 ± 0.5
Age Ethnicity Gender	19.1 ± 3.3			8.4 ± 1.3	4.9 ± 0.6
Age Ethnicity Gender Moustache	14.4 ± 2.6			7.7 ± 1.5	4.8 ± 0.5
Age Ethnicity Gender Moustache Glasses	11.9 ± 2.2			7.7 ± 1.5	4.8 ± 0.7
Age Ethnicity Gender Moustache Glasses Beard	12.0 ± 2.2			8.3 ± 1.7	5.4 ± 0.9
Age Ethnicity Gender Moustache Glasses*	11.2 ± 2.1			7.6 ± 1.4	4.4 ± 0.5
Age Ethnicity Gender Moustache Glasses* Beard	11.1 ± 2.1			8.0 ± 1.7	5.2 ± 0.7

- E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, "Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation," IEEE Transactions on Information Forensics and Security, 13(8), 2001-2014, 2018.

DeepFakes

Fake images and videos including facial information generated by digital manipulation. Reasons:

- 1) Free access to large-scale public databases.
- 2) Fast progress of deep learning techniques such as Generative Neural Networks, Autoencoders and Variatonal Autoencoders.

- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Deepfakes and beyond: A survey of face manipulation and fake detection,” Information Fusion, vol. 64, pp. 131-148, 2020.

BiDALab-UAM ©

DeepFakes

Fake images and videos including facial information generated by digital manipulation. Reasons:

- 1) Free access to large-scale public databases.
- 2) Fast progress of deep learning techniques such as Generative Neural Networks, Autoencoders and Variatonal Autoencoders.

Applications: fake news, fake pornography, hoaxes, financial fraud, etc.

- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131-148, 2020.

BiDALab-UAM ©

DeepFakes

Fake images and videos including facial information generated by digital manipulation. Reasons:

twitter facebook Instagram

Real: Robert De Niro

Fake: Al Pacino

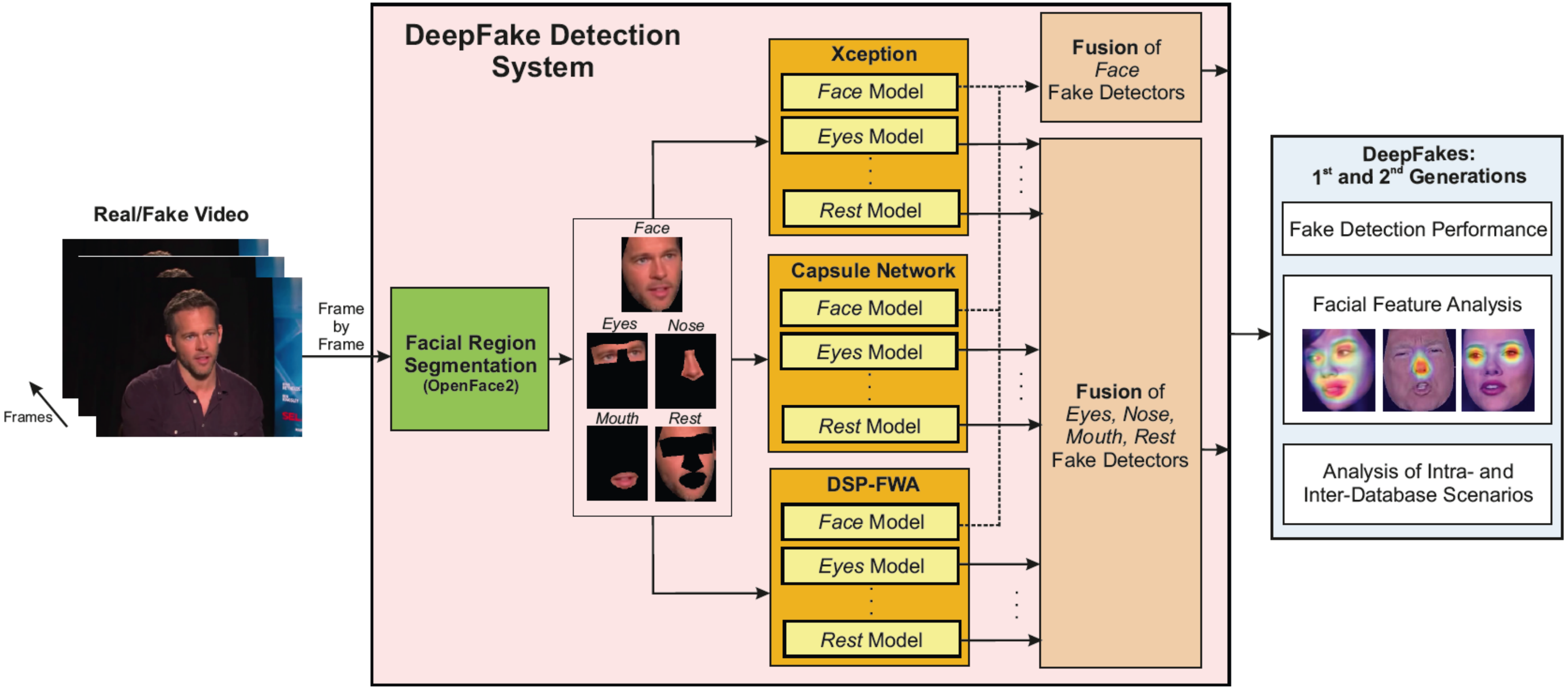
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Deepfakes and beyond: A survey of face manipulation and fake detection,” Information Fusion, vol. 64, pp. 131-148, 2020.

DeepFakes

BiDA Lab
Biometrics & Data Pattern Analytics Lab

Real/Fake Detection:

- Based on facial regions, state-of-the-art DCNNs, and fusion of systems.



- R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance," in Proc. International Conference on Pattern Recognition Workshops, 2020.

DeepFakes

Real/Fake Detection:

- Based on facial regions, state-of-the-art DCNNs, and fusion of systems.

Study	Method	Classifiers	AUC Results (%)			
			UADFV [5]	FF++ [11]	Celeb-DF v2 [6]	DFDC-Preview [7]
Yang *et al.* [32]	Head Pose Features	SVM	89.0	*47.3*	*54.6*	*55.9*
Li *et al.* [6]	Face Warping Features	CNN	97.7	*93.0*	*64.6*	*75.5*
Afchar *et al.* [33]	Mesoscopic Features	CNN	*84.3*	84.7	*54.8*	*75.3*
Sabir *et al.* [14]	Image + Temporal Features	CNN + RNN	-	96.3	-	-
Li *et al.* [5]	Eye Blinking Features	LRCN	99.0%	-	-	-
Dang *et al.* [13]	Deep Learning Features	CNN + Attention Mechanism	98.4	-	*71.2*	-
Present Study	**Deep Learning Features**	**Fusion of CNN**	**100**	**99.79**	**99.04**	**90.61**

Comparison in terms of AUC (%) of different state-of-the-art fake detectors with the present study.
The best results achieved for each database are remarked in bold. Results in italics indicate that the evaluated database was not used for training [6].

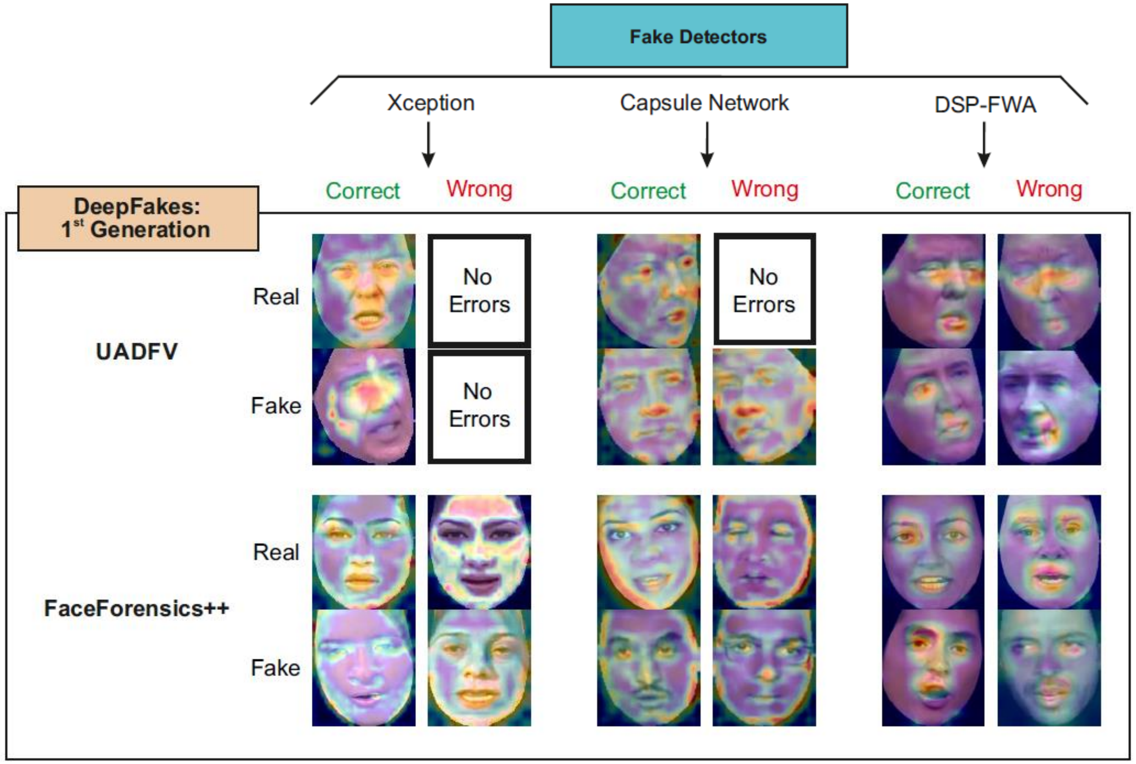
- R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance," in Proc. International Conference on Pattern Recognition Workshops, 2020.

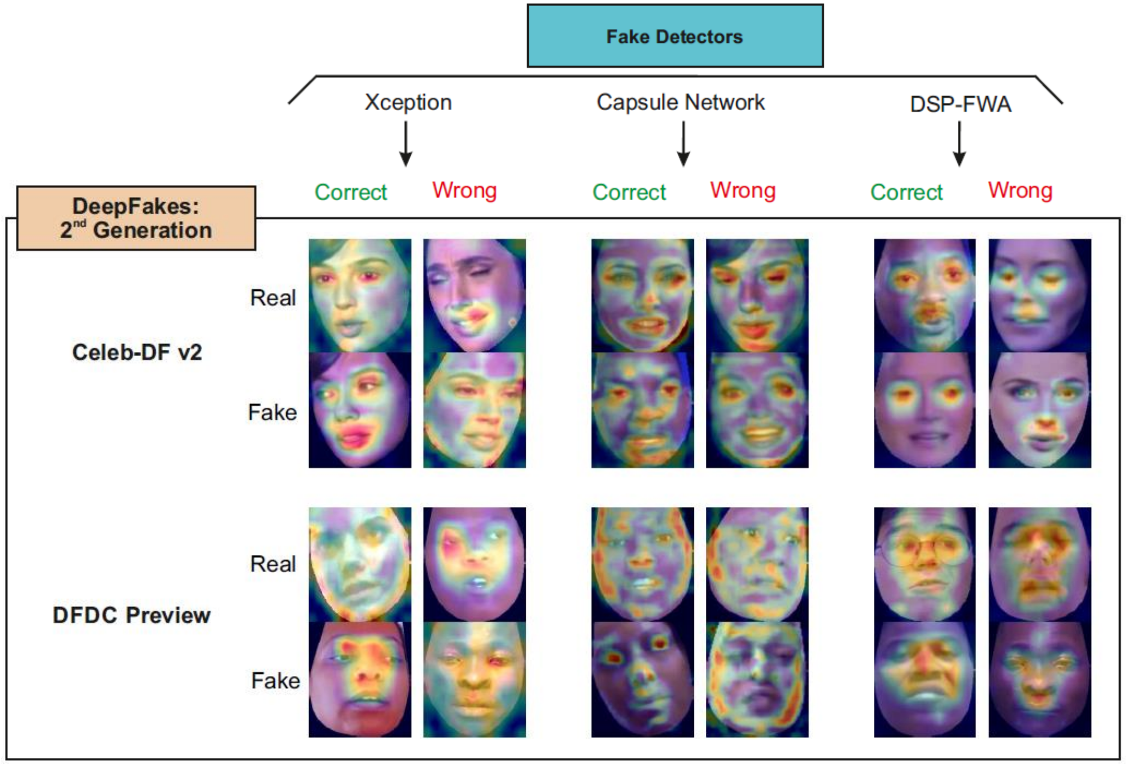
BiDALab-UAM ©

DeepFakes

Real/Fake Detection:

- Based on facial regions, state-of-the-art DCNNs, and fusion of systems.



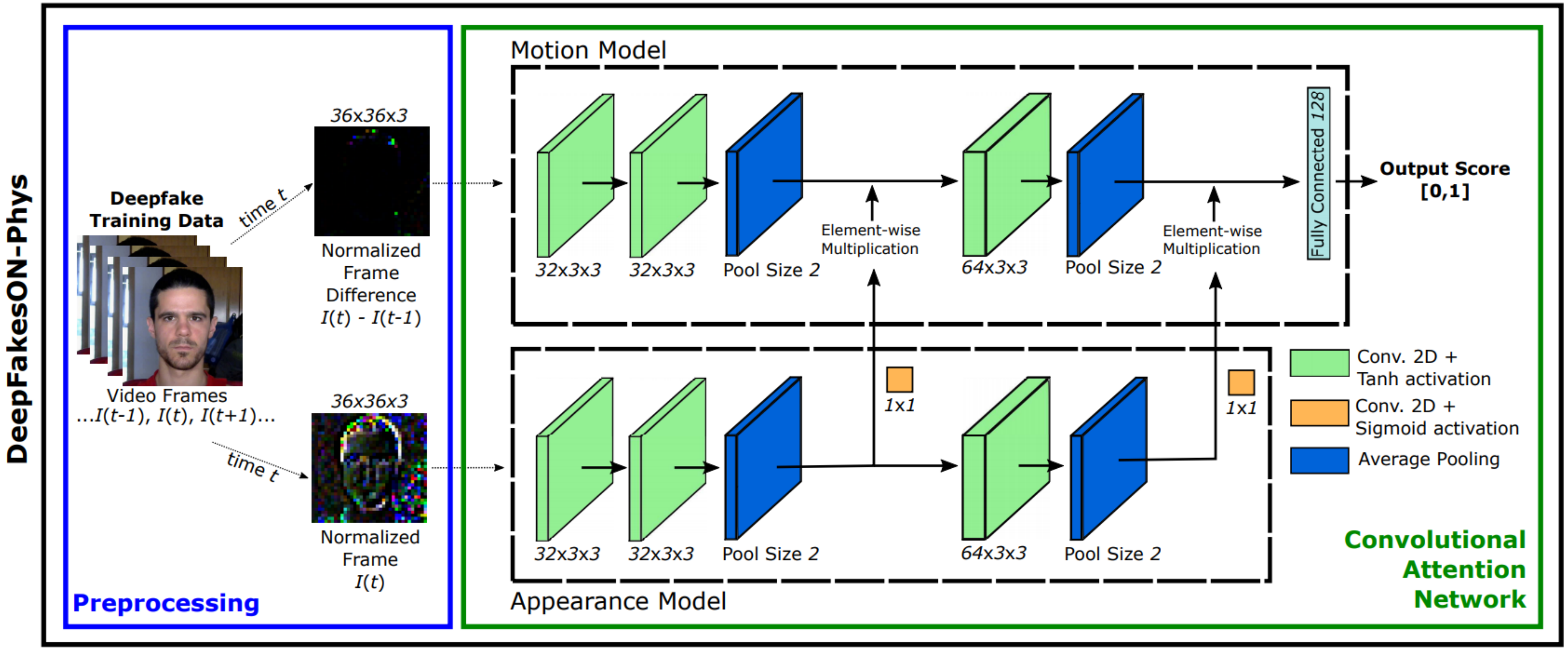


- R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, “DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance,” in Proc. International Conference on Pattern Recognition Workshops, 2020.

DeepFakes

Real/Fake Detection:

- Based on physiological information (heart rate) and state-of-the-art DCNNs.



- J. Hernandez-Ortega, R. Tolosana, J. Fierrez, and A. Morales, "DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation," in Proc. AAAI Conference on Artificial Intelligence Workshops, 2021.

DeepFakes

BiDA Lab

Biometrics & Data Pattern Analytics Lab

Real/Fake Detection:

- Based on physiological information (heart rate) and state-of-the-art DCNNs.

Study	Method	Classifiers	AUC Results (%)	
			Celeb-DF [25]	**DFDC** [30]
Yang *et al.* (2019) [42]	Head Pose Features	SVM	*54.6*	*55.9*
Li *et al.* (2020) [25]	Face Warping Features	CNN	*64.6*	*75.5*
Afchar *et al.* (2018) [43]	Mesoscopic Features	CNN	*54.8*	*75.3*
Dang *et al.* (2020) [29]	Deep Learning Features	CNN + Attention Mechanism	*71.2*	-
Tolosana *et al.* (2020) [7]	Deep Learning Features	CNN	83.6	91.1
Qi *et al.* (2020) [36]	Physiological Features	CNN + Attention Mechanism	-	*Acc. = 64.1*
Ciftci *et al.* (2020) [34]	Physiological Features	SVM/CNN	Acc. = 91.5	-
DeepFakesON-Phys [Present Paper]	**Physiological Features**	**CNN + Attention Mechanism**	**AUC = 99.9 Acc. = 98.7**	**AUC = 98.2 Acc. = 94.4**

Comparison in terms of AUC (%) of different state-of-the-art fake detectors with the present study. The best results achieved for each database are remarked in bold. Results in italics indicate that the evaluated database was not used for training [6].

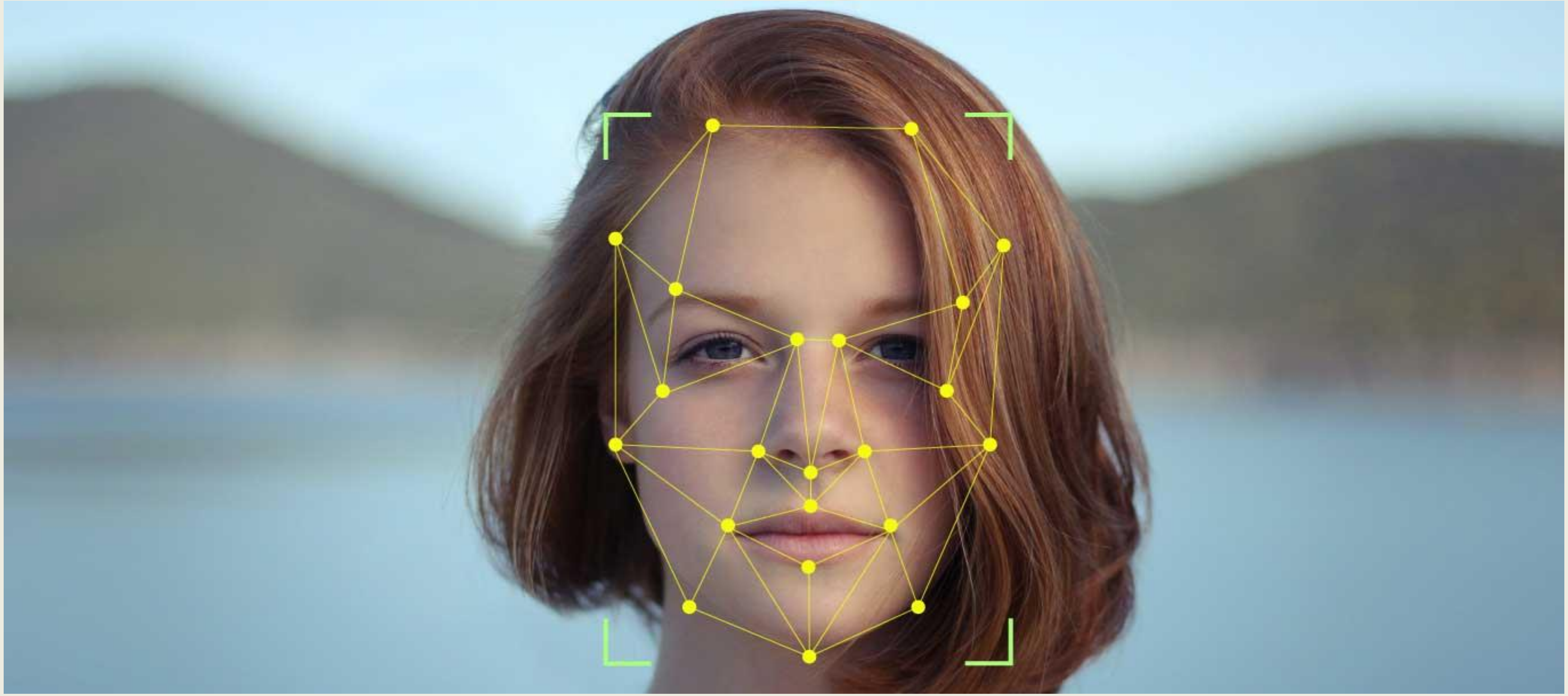
- J. Hernandez-Ortega, R. Tolosana, J. Fierrez, and A. Morales, "DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation," in Proc. AAAI Conference on Artificial Intelligence Workshops, 2021.

Key References

- W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” ACM Computing Surveys (CSUR), 35(4), 399-458., 2003.
- R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J. C. Chen, V. M. Patel, C. D. Castillo, and R Chellappa, “Deep learning for understanding faces: Machines may be just as good, or better, than humans,” IEEE Signal Processing Magazine, 35(1), 66-83, 2018.
- I. Masi, Y. Wu, T. Hassner, and P. Natarajan, “Deep face recognition: A survey,” in Proc. Conference on Graphics, Patterns and Images, 2018.
- M. Wang, and W. Deng, “Deep Face Recognition: A Survey,” arXiv preprint arXiv:1804.06655, 2020.
- X. Jin, and X. Tan,“Face alignment in-the-wild: A survey,” Computer Vision and Image Understanding, 162, 1-22, 2017.
- S. Marcel, M.S. Nixon, J. Fierrez, N. Evans, “ Handbook of Biometric Anti-Spoofing (2nd Edition),” Springer, 2019.
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Deepfakes and beyond: A survey of face manipulation and fake detection,” Information Fusion, vol. 64, pp. 131-148, 2020.
- A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, “SensitiveNets: Learning agnostic representations with application to face images” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
- E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, “Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation,” IEEE Transactions on Information Forensics and Security, 13(8), 2001-2014, 2018.
- U. Park, Y. Tong, and A.K. Jain, “Age-invariant face recognition” IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 947-954, 2010.
- M.M. Sawant, and K.M. Bhurchandi, “Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging,” Artificial Intelligence Review, 52(2), 981-1008, 2019.

BiDALab-UAM ©

Face Recognition



Ruben Tolosana
ruben.tolosana@uam.es

BiDA Lab
Biometrics & Data Pattern Analytics Lab

UAM
Universidad Autónoma de Madrid