Imperial College London

Face Recognition

Binod Bhattarai

Imperial College London

Introduction

Source: *google image*

Source: *google image*

Imperial College London

Applications (Face verification for security)

The end of passports? How Australia plans to make travel documents obsolete by 2020

share

Save 7

New technology could eliminate the need for passports in Australia by 2020 CREDIT: GETTY

Australia is planning to adopt a new contactless passenger identification system that would eliminate the need for passport scanners, paper landing cards and manned immigration desks, the Australian Department of Immigration and Border Protection has announced.

The new system, which is set to be rolled out by 2020, will use facial recognition technology and fingerprint scanners to identify passengers as they pass through Australian airports. People arriving in the country would no longer be required to show their passports and desks fronted by immigration officers would be replaced by automated electronic stations.

Source: Telegraph, 23 Jan 2017

Imperial College London

Applications (Face verification for surveillance)

Credit: AFP/Getty Images

China's police have been testing sunglasses with built-in facial recognition since at least last month to catch suspects and those traveling under false identities. Now China is

Source: The verge

Source: google image

Imperial College London

Face analysis tasks: Face verification

I_1 I_2

Same (+1)

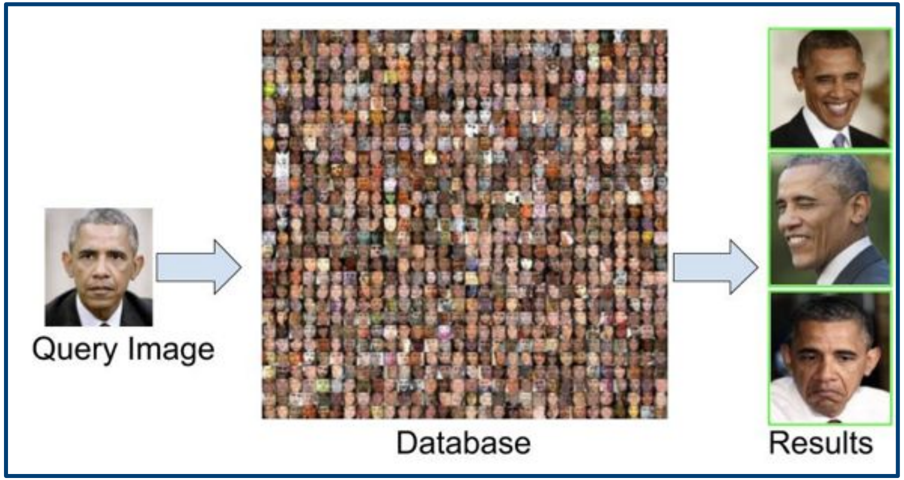
I_2 I_3

Different (-1)

f:(I_i, I_j) -> {+1 , -1}

Imperial College London

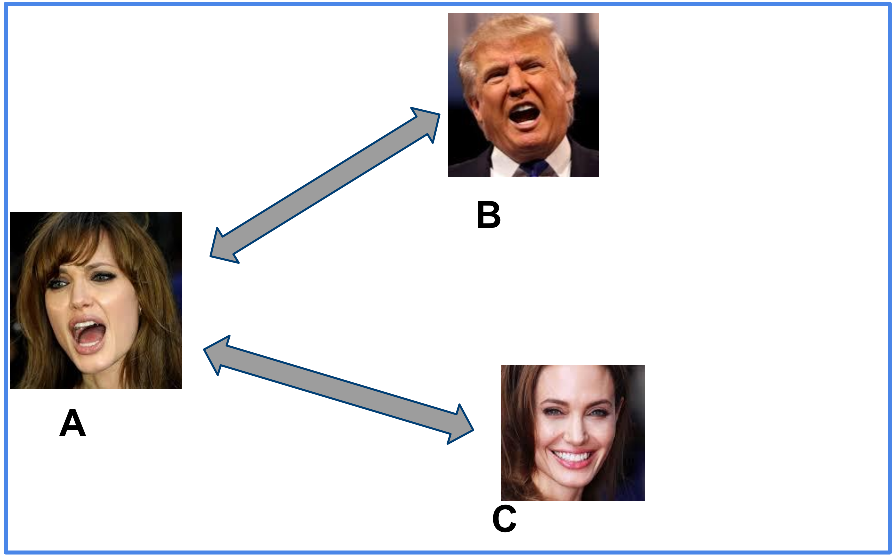
Face retrieval



- Given a query image I_q, compute the ***similarity score*** between the query and the images in the database
- Rank them based on the ***sim_score***
- Retrieve **Top-K** ranked images

Imperial College London

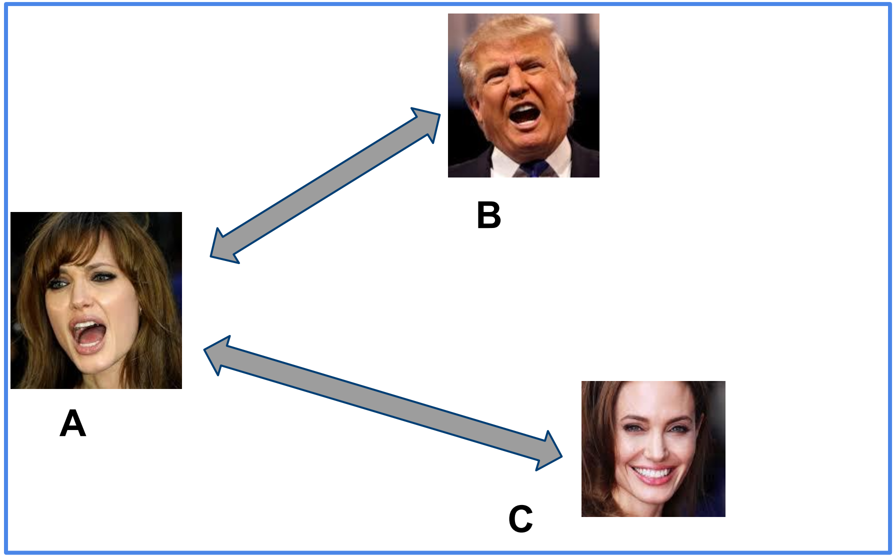
Similarity between the Faces



- **Which pair is more similar?**
  - (A, B)
  - (A, C)

Imperial College
London

Similarity between the Faces

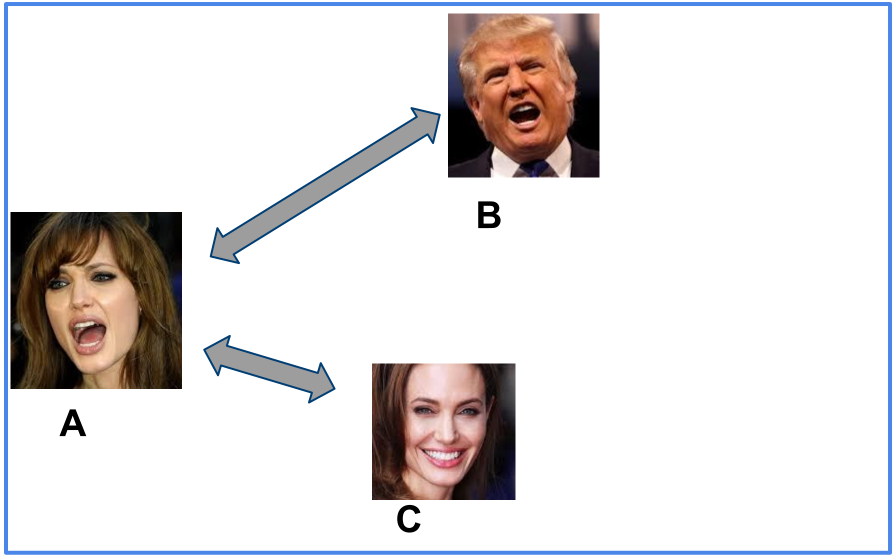


- When we care about expression

$$d_{(mood)}(A, B) < d_{mood}(A, C)$$

Imperial College London

Similarity between the Faces



- When we care about identity

$$d_{(id)}(A, B) > d_{(id)}(A, C)$$

- Similarity between the images depends on the **the task we care about**
- **Need a methodology to learn such metric**

Metric Learning

- Euclidean or L2 distance is probably the most well known metric

$$d_{L2}(x,y) = (x-y)^T (x-y)$$

  - No parameter to learn
- Most common form of learned metrices are Mahalanobis

$$d_M(x,y) = (x-y)^T M(x-y)$$

  - M ∈ $R^{(D \times D)}$ is a semi-definite matrix as it measures distance
  - Generalization of Euclidean metric (setting M=I)
  - M can be decomposed into $L^T L$, L ∈ $R^{(D \times d)}$ d<<D
  - Corresponds to Euclidean metric after linear projection of data

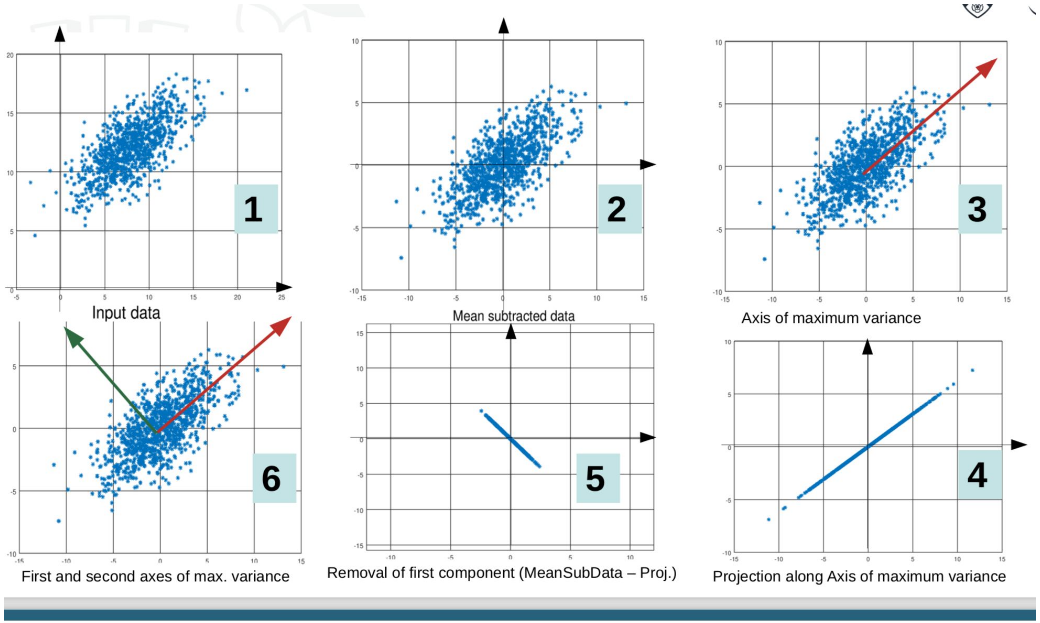
$$d_M(x,y) = (x-y)^T M(x-y) = (x-y)^T L^T L(x-y) = d_{L2}(Lx,Ly)$$

- Reduces the dimension by large-margin

Imperial College London

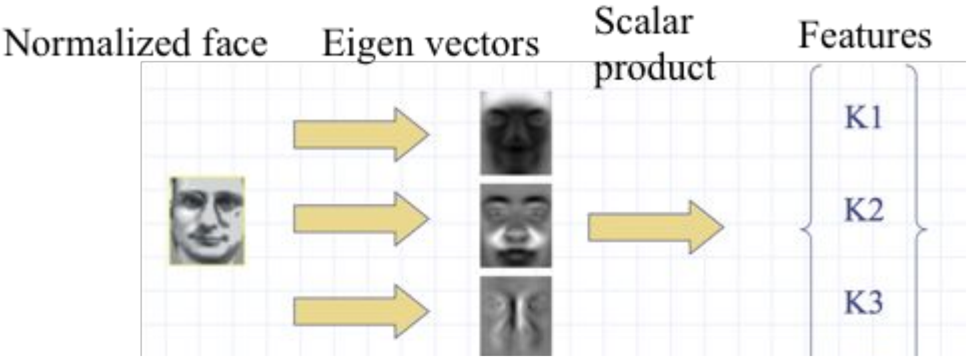
Slide credit: Suresh Manandhar

[Recap] PCA - unsupervised metric learning



Imperial College London

[Recap] on eigenfaces



- Normalized face is projected on the eigenvectors computed by PCA
- Projected on first 3 eigenvectors, the dimension of feature = **3**

Imperial College London

PCA

- **Merits**
  - Removes redundancies and noise to make feature more discriminative
  - Transform the representations into compact form (dimensionality reduction)
  - Does not require label information

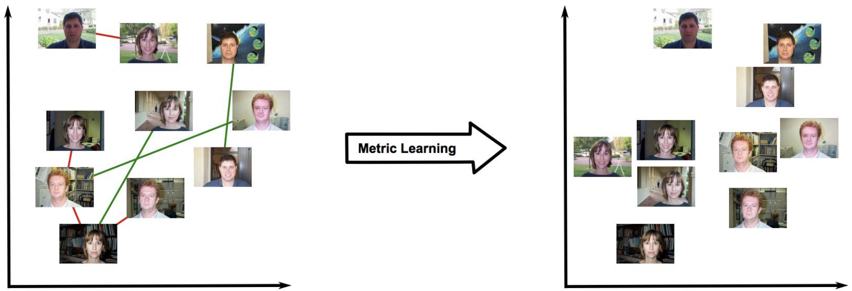
PCA

- **Demerits**
  - Does not embed category information
  - Large variations on illuminations or poses (not necessarily the axis with maximum variance contains discriminative features)

  - Not robust to illuminations, expressions, poses etc.

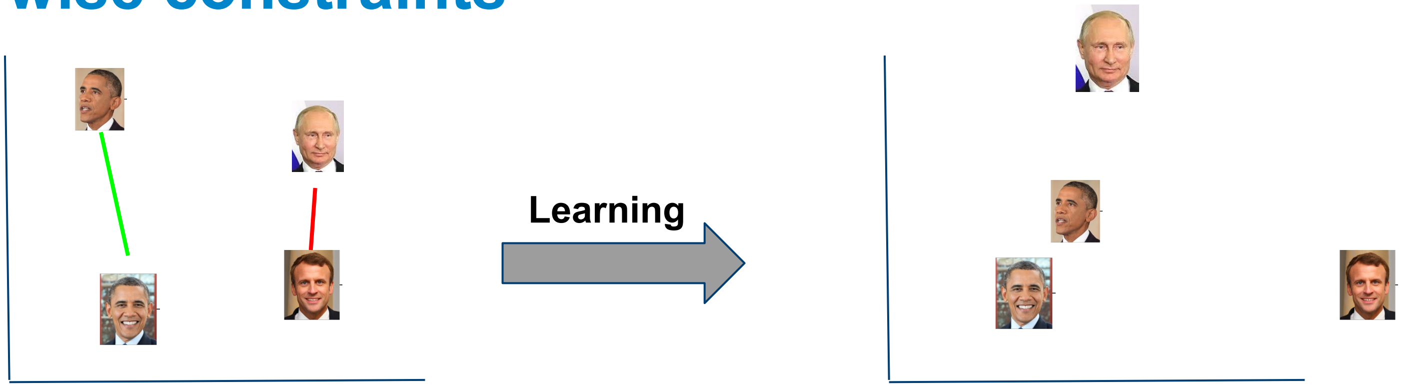
Imperial College
London

Supervised Metric Learning (ML) for face verification



- ❑ Learn a projection matrix where the imposed constraints are **better satisfied**
- ❑ Commonly used constraints are: ***pairwise similarity and dissimilarity*** constraints and ***triplet constraints***

Pairwise constraints

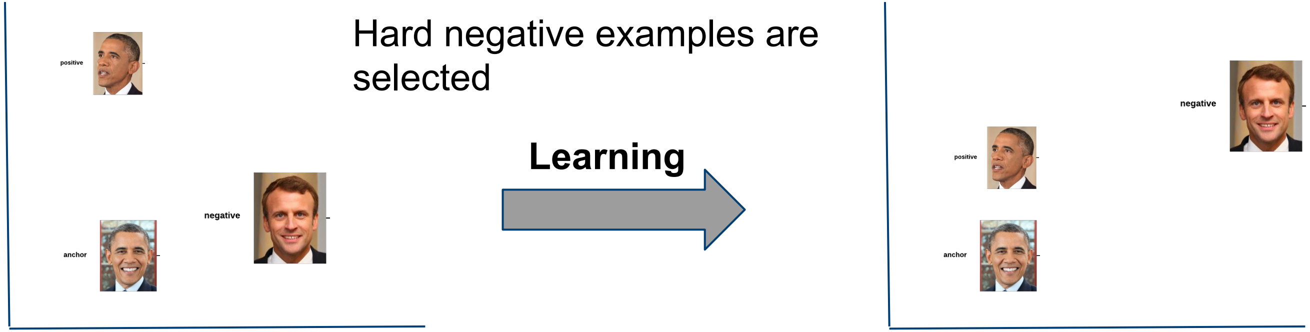


- Must-link / cannot-link constraints (sometimes called positive / negative pairs):

$$
\begin{aligned}
\mathcal{S} &= \{(x_i, x_j) : x_i \text{ and } x_j \text{ should be similar}\}, \\
\mathcal{D} &= \{(x_i, x_j) : x_i \text{ and } x_j \text{ should be dissimilar}\}.
\end{aligned}
$$

Imperial College London

Triplet constraints



- Relative constraints (sometimes called training triplets):

$$\mathcal{R} = \{(x_i, x_j, x_k) : x_i \text{ should be more similar to } x_j \text{ than to } x_k\}.$$

Pairwise dis(similarity) vs triplet constraints

- Pairwise constraints are easy to collect (Weakly supervised)
- Eg. video frames
- Triplet ,requires (hard) negative examples, adds extra layer of difficulty

Imperial College London

Learning projection matrix

- We minimize the max-margin objective function to learn the projection matrix satisfying pairwise dis(similar) constraints (y_{ij} = +/-1)

$$\underset{L}{\operatorname{argmin}} \sum_{t=1}^{t=n} \max\left(m - y_{ij}^{t}(b - d_L^2(x_i^t, x_j^t)),\ 0\right)$$

- Where $d_L^2(x_i, x_j) = ||Lx_i - Lx_j||^2$

- Pushes the examples s.t. Distance of negative pairs is larger by '*m*' than bias '*b*'

Imperial College London

Learning the parameters of the projection matrix

- We use stochastic gradient descent

$$\frac{d_L^2(x_i, x_j)}{dL} = L(x_i - x_j)(x_i - xj)^T$$

- Update rule

if $y_{ij}(b - d_L^2(x_i, x_j)) < m$ **then**
  $L \leftarrow L - \eta y_{ij} L(x_i - x_j)(x_i - x_j)^T$
else
  `no update`
end if

Imperial College London

Performance comparison

- Database
  - LFW: Labeled Faces in the Wild, contains *13K* of *5K* identities
  - Standard benchmark for face analysis task
- Performed face retrieval task
- Metric used is *1-call@K* (2, 5, 10)

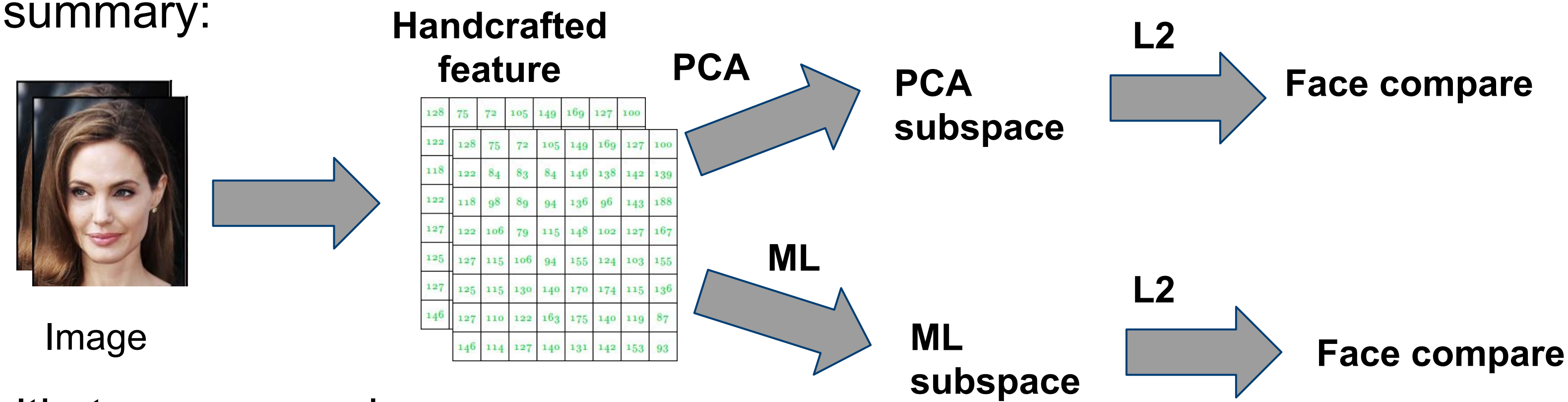
Method	K=2	K=5	K=10	K=20
PCA	30.0	37.4	43.3	51.3
ML	**38.1**	**51.1**	**60.5**	**69.3**

Source: ***Bhattarai et al CVPR 2016***

Imperial College London

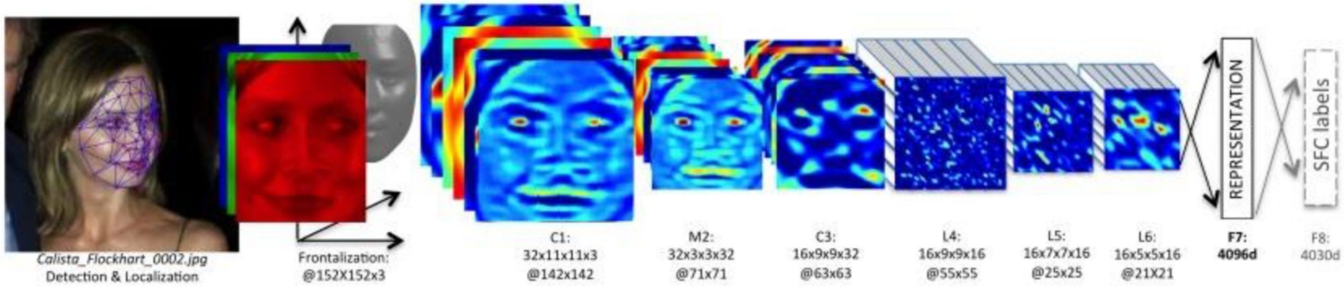
Limitations of parameterized distance

- In summary:



- Multi-stage approach
- Features are not optimal for end task (no feedback mechanism to propagate error to input)

Deepface

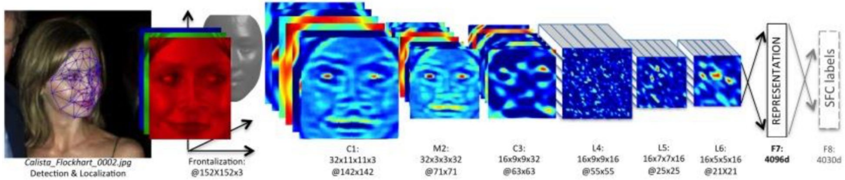


- ❑ Train set size: 4M images from 4K identities
- ❑ Minimize cross-entropy loss to learn the parameters

Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." *CVPR 2014*

Imperial College London

Deepface



- F8 calculates probability with softmax $p_k = \exp(o_k) / \sum_h \exp(o_h)$
- Cross-entropy loss function: $L = -\sum_k \log(p_k)$
- Computed using SGD and performs backpropagation

Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." *CVPR 2014*

Imperial College London

Experiments

- DeepFace was evaluated in LFW

- Human cropped: 97.5% vs Deepface 97.35%

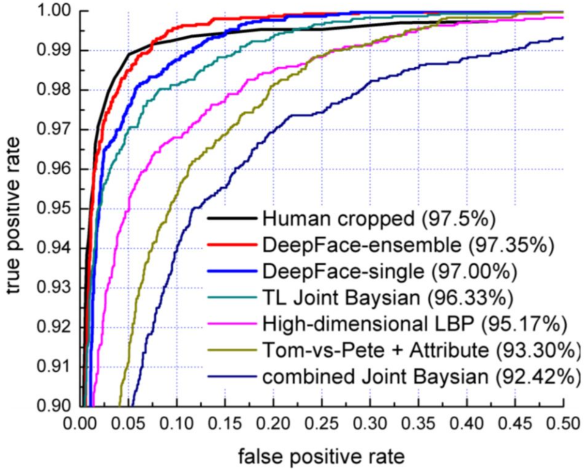
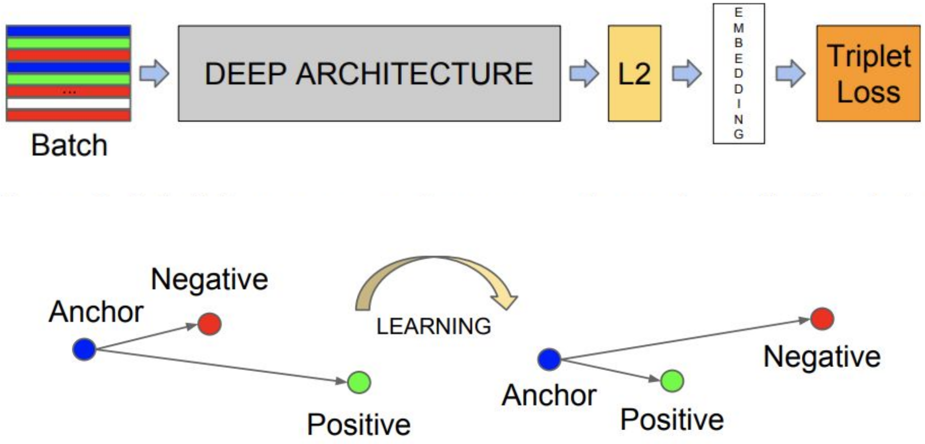


Fig. Roc Curve on LFW

Imperial College London

FaceNet



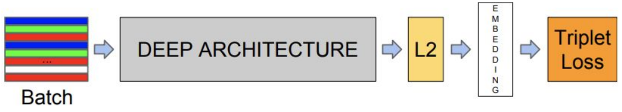
- Objective of this architecture is to minimize L2 distance between same identity's faces representations
- Directly transforms image representations at a low dimensional feature space (128D vs 4096D (Deepface)) rather than bottleneck intermediate representations

Schroff, Florian, Dmitry Kalenichenko, and James Philbin. **"Facenet: A unified embedding for face recognition and clustering."** *CVPR 2015*

APA

Imperial College London

FaceNet



- Uses triplet loss
- Minimize the max-margin objective

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

s.t.

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 ,$$

$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} .$$

- This ensures all positive examples are nearer than negative examples
- Very useful for clustering of faces

Schroff, Florian, Dmitry Kalenichenko, and James Philbin. **"Facenet: A unified embedding for face recognition and clustering."** *CVPR 2015*

APA

Experiments

- Dataset: LFW
- Accuracy: 0.9963 Vs 0.9735 (Deepface) Vs 0.975 (Human)

Imperial College London

Mis-classified examples

False reject

False accept

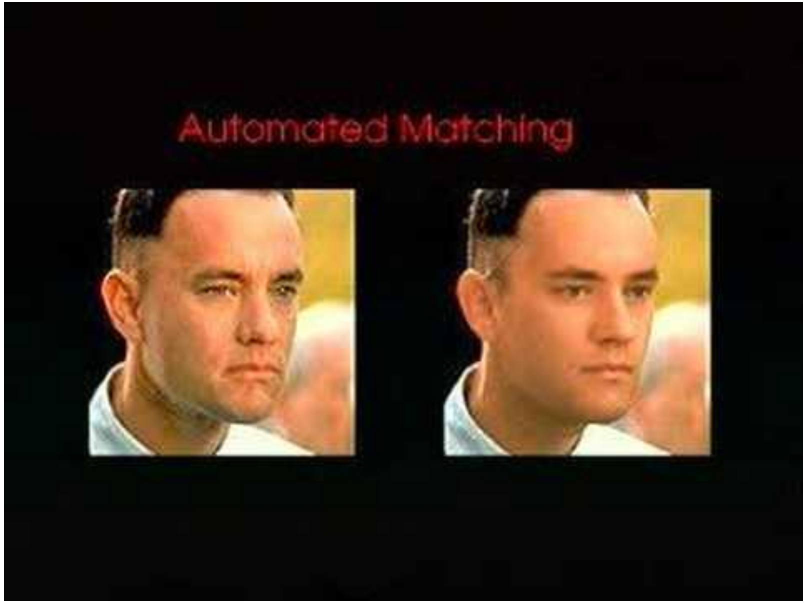
Imperial College London

Bottlenecks in face analysis using deep learning

- Computing resource
- <u>Data hungry</u>

Imperial College London

3D Morphable Model



Imperial College London

Synthetic Face Images by 3DMM

- + Easy to manipulate attributes such as identity, pose, expression, and lighting
- + Can generate millions of images with controlled attributes
- - Domain gap with real face images

Synthetic images generated by 3DMM

Imperial College London

Problem Definition

- Generating photorealistic face images 3DMM rendered faces of new identities with arbitrary poses, expressions, and illuminations

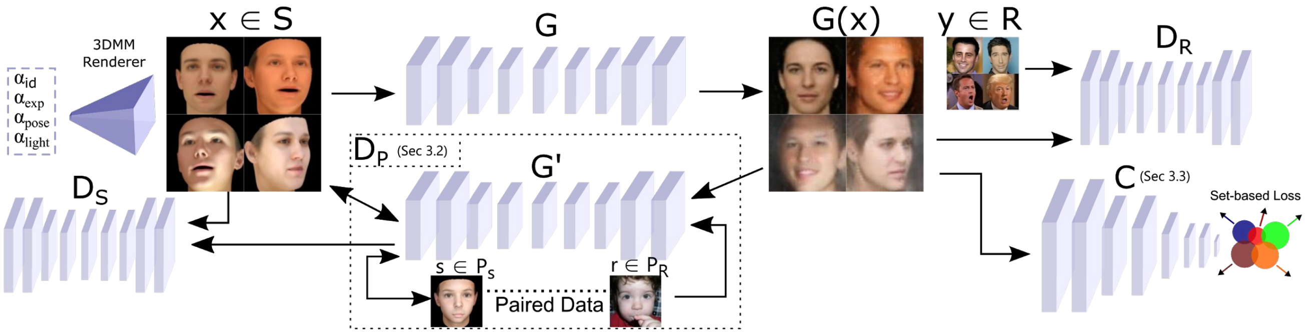
Imperial College London

Problem Definition

- Generating photorealistic face images 3DMM rendered faces of new identities with arbitrary poses, expressions, and illuminations
- We formulate this problem as domain adaptation problem/ style transfer problem (3DMM -> Real)
    - *Pixel2Pixel (Isola et al 2017)*
    - *CycleGAN(Zhu et al 2017)*

- How can we benefit small amount of paired data in unsupervised style transfer GAN?
- How to prove identity consistency of generated images?

Photorealistic identity synthesis (Gecer, Bhattarai, Kittler, and Kim *ECCV'2018*)



- ❖ Randomly generated 3DMM images with random pose, expression and lighting attributes for the new IDs.
- ❖ Unsupervised training with forward cycle consistency.
- ❖ Adversarial Pair Matching network G' by the help of a limited number of paired data.
- ❖ ID preservation by a set-based supervision through a pre-trained classification network C.

Experiments (Quantitative)

VGG(%100)	1.8M	-	96×96	94.8
VGG(%100) + GANFaces-500K	1.8M	500K	96×96	**94.9**
VGG(%100) + GANFaces-5M	1.8M	5M	96×96	**95.2**

Tab. Verification accuracy on LFW benchmark

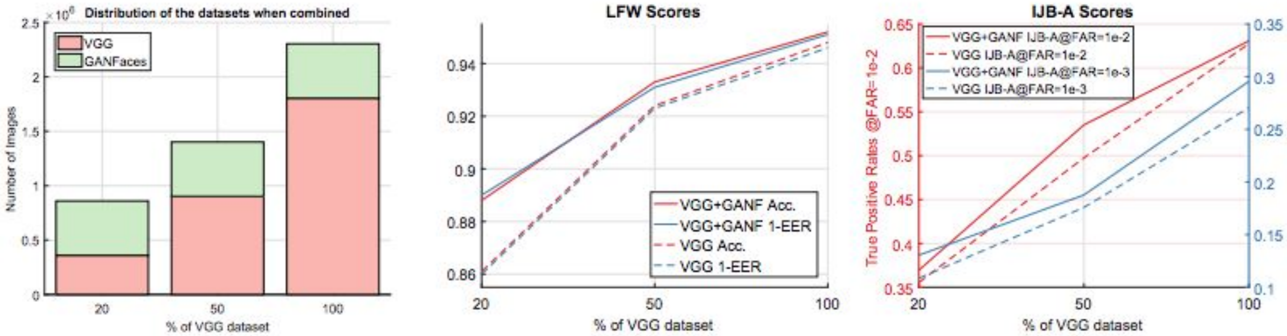


Fig. Verification on LFW and IJB-A database with different size of original and synthetic data

Experiments (Qualitative)

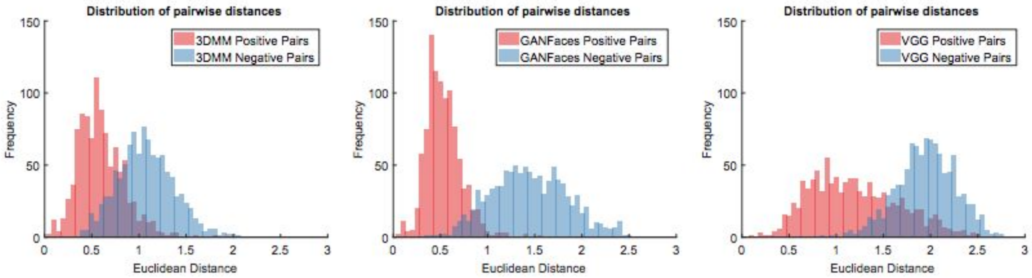
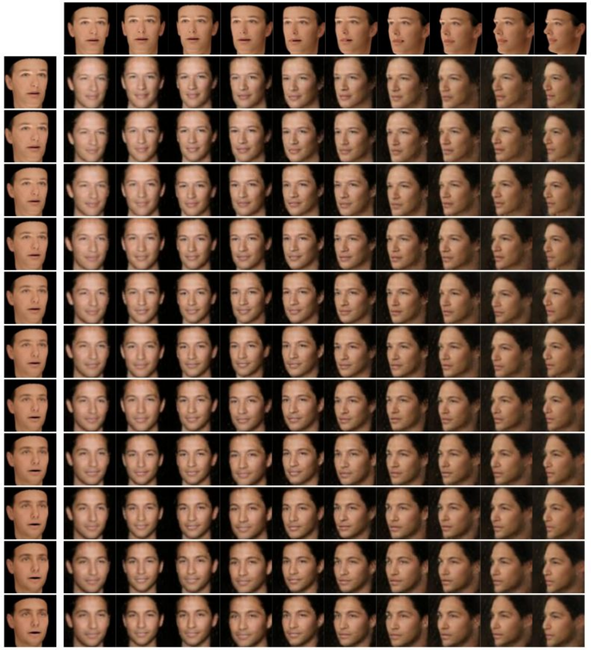


Fig. Face pairs euclidean distance distribution

Experiments (Qualitative)

Fig. Face generated by the proposed method conditioned on 3DMM identity, expression and pose parameters

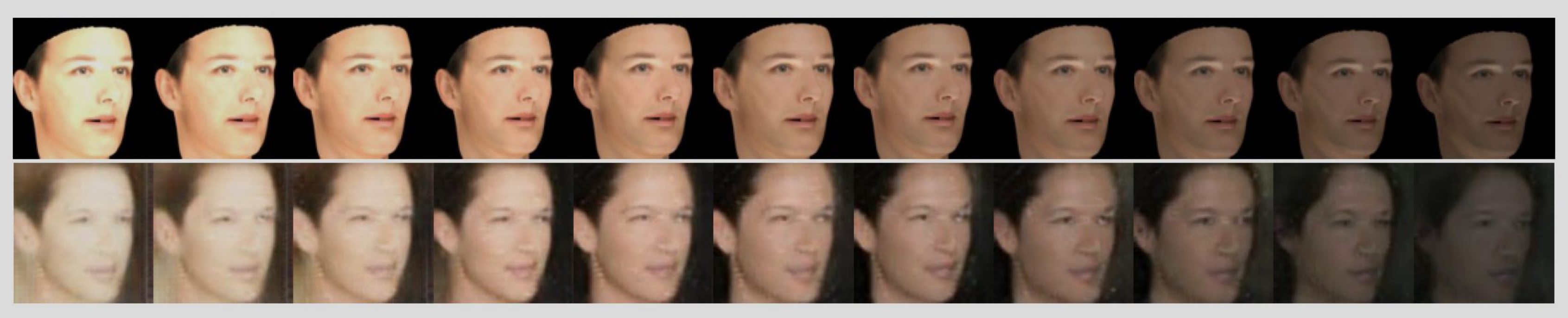
Experiments (Qualitative)



- Interpolation in identity space
- Smooth transition from one identity space to another identity space shows that manifold of image generator is smooth.

Imperial College London

Experiments (illumination preservation)



Our model is also preserving the lighting conditioned by 3DMM.

Imperial College London

Experiments(Qualitative)

- ❖ The nearest images from the training set in terms of identity features for the images
- ❖ Variation in the nearest images shows diversity of GANFaces in the embedding space while they bear similar higher order attributes such as gender, shape of face etc.

Imperial College London

Comparison with existing methods

A) 3DMM

B) CycleGAN

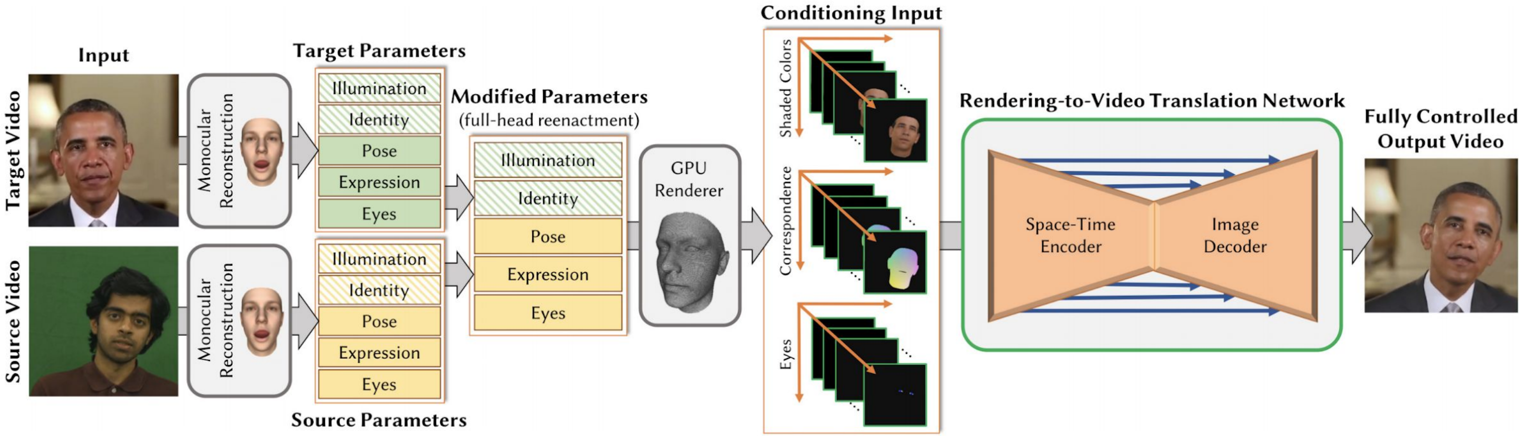
C) Our approach

Imperial College London

Deep video portrait for video translation

Imperial College London

Deep video portrait for video translation

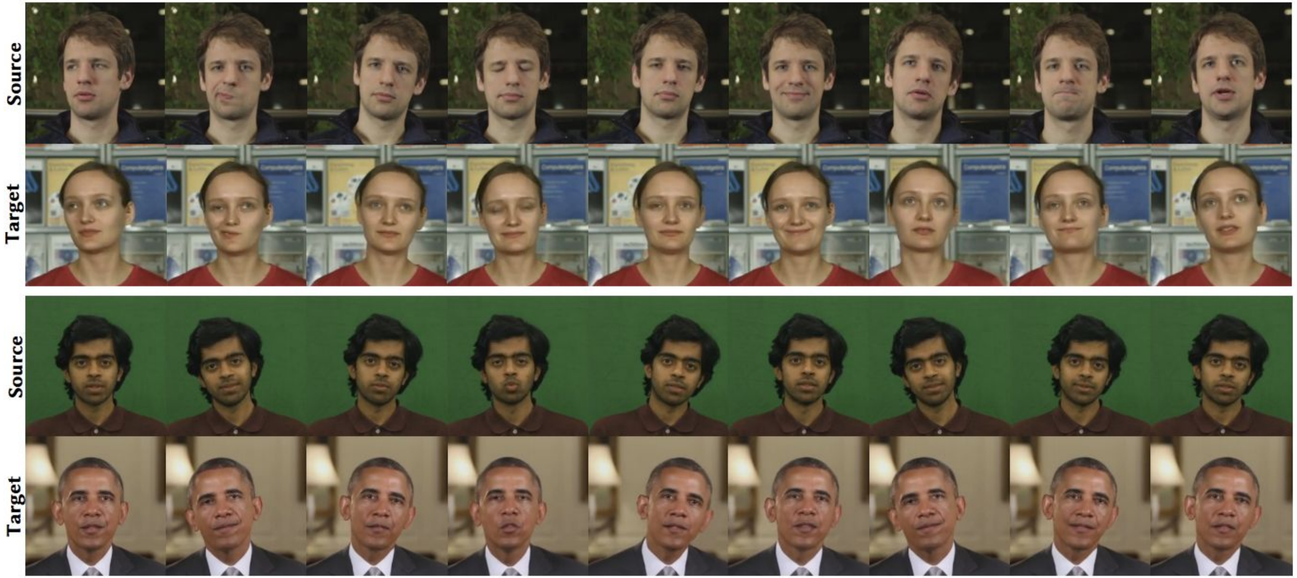


Kim, Hyeongwoo, et al. "Deep Video Portraits." *SIGGRAPH 2018*

APA

Imperial College London

Qualitative results



Imperial College London

Native Dubbing

Imperial College
London

Imperial College London

Thank You

Imperial College London

PCA

- **Merits**
  - Removes redundancies and noise to make feature more discriminative
  - Transform the representations into compact form (dimensionality reduction)
  - Does not require label information

- **Demerits**
  - Does not embed category information
  - Imagine large variations is due to illuminations due external sources (not necessarily the axis with maximum variance contains discriminative features)
  - Sub-optimal for specific task(s)

Imperial College London

References

1. Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Gray scale and rotation invariant texture classification with local binary patterns." *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2000.
2. Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12 (2006): 2037-2041.
3.

Imperial College London

Properties of pseudo-distance

1. $d_M(\boldsymbol{x}, \boldsymbol{x}') \geq 0$ (nonnegativity),
2. $d_M(\boldsymbol{x}, \boldsymbol{x}) = 0$ (identity),
3. $d_M(\boldsymbol{x}, \boldsymbol{x}') = d(\boldsymbol{x}', \boldsymbol{x})$ (symmetry),
4. $d_M(\boldsymbol{x}, \boldsymbol{x}'') \leq d(\boldsymbol{x}, \boldsymbol{x}') + d(\boldsymbol{x}', \boldsymbol{x}'')$ (triangle inequality).

Imperial College London

Additional Applications

Security

Surveillance

Robotics

All these applications are face matching tasks

Metric Learning

- Euclidean or L2 distance is probably the most well known

$$d_{L2}(x,y) = (x-y)^T (x-y)$$

- Most common form of learned metrics are Mahalanobis

$$d_M(x,y) = (x-y)^T M(x-y)$$

  - M is a positive definite matrix
  - Generalization of Euclidean metric (setting M=I)
  - Corresponds to Euclidean metric after linear projection of data

$$d_M(x,y) = (x-y)^T M(x-y) = (x-y)^T L^T L(x-y) = d_{L2}(Lx,Ly)$$

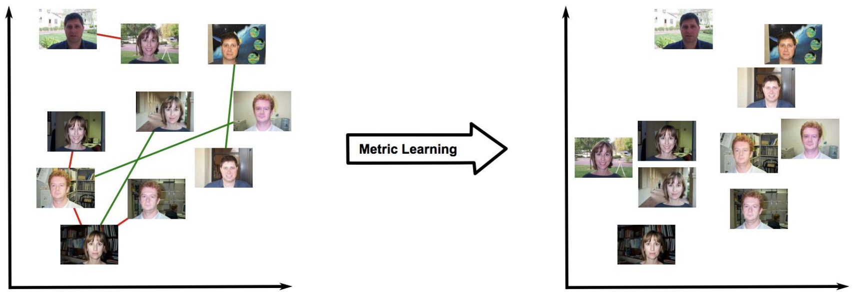
$,Ly)$

tation, eg

- Not all methods fit this formulation of fixed vectorial data representation, eg based on matching image regions **[Nowak & Jurie 2007]**

Imperial College London

Metric Learning for face recognition



$$d_M^2(x_i, x_j) = (x_i - x_j)^\top M (x_i, x_j)$$

$$M = L^\top L; L \in \mathbb{R}^{d \times D}; d \ll D$$

$$d_L^2(x_i, x_j) = ||Lx_i - Lx_j||^2$$

- ❑ Pairwise constraints (same person vs different persons) are imposed
- ❑ Projection matrix is learned to better satisfy the imposed constraints

Imperial College London

Recap (Image representation)

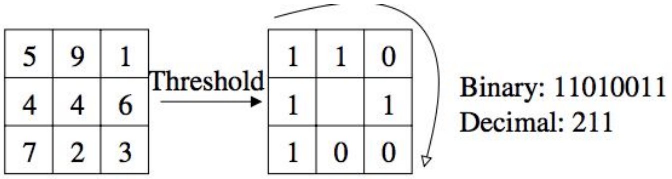
- Pixel Intensity
  - Not robust to scaling, illuminations, rotations etc

- SIFT
  - invariant to scaling, rotation and translations and partially illuminations

Imperial College London

Local Binary Patterns

- Ojala et al. (PAMI’ 94) proposed a simple but efficient method for texture classification
- Successfully applied for face recognition [2]
- Compare ***neighbouring pixel intensities***
- If central pixel intensity is lower assign 1 or else 0
- Map the binary to decimal
- Compute 256D histogram as image Feature

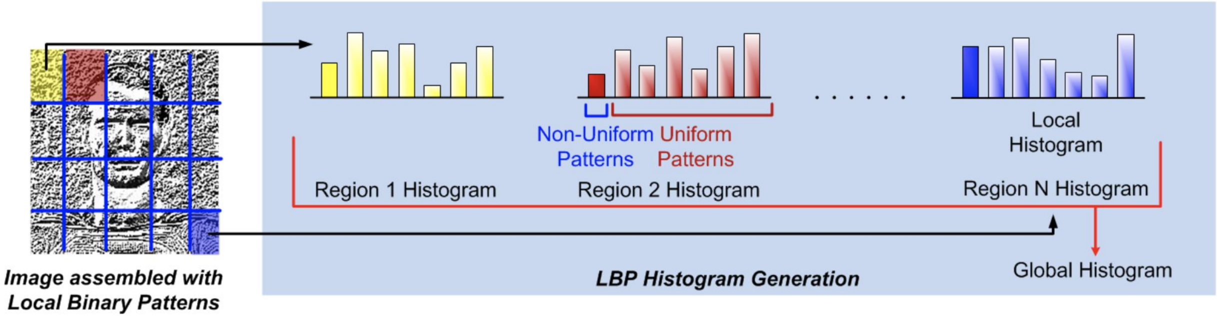


Pixel Intensity Thresholding

Imperial College London

Local Binary Patterns for face recognition

- Ojala et al. (PAMI' 94) proposed a simple but efficient method for texture classification
- Successfully applied for face recognition [2]



Imperial College London

Our contributions

- End-to-end adversarial training framework to generate photorealistic face images on new identities conditioned on synthetic 3DMM images
- Semi-supervised adversarial style transfer framework
- Set-based loss function to preserve consistency among unknown identities