Organizers:

IAPR/IEEE WINTER SCHOOL ON BIOMETRICS 2023

8 - 12 January 2023 Shenzhen, China

Overview on Biometrics Data Analysis

Zhenan Sun

Center for Research on Intelligent Perception and Computing (CRIPAC)
National Laboratory of Pattern Recognition (NLPR)
Chinese Academy of Sciences’ Institute of Automation (CASIA)

January 9, 2023

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

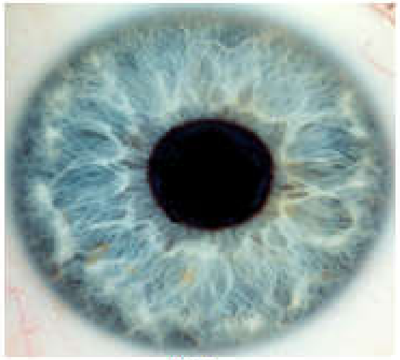
中国科学院
自动化研究所

Biometrics

www.ia.ac.cn

Automated recognition of individuals based on their behavioral and biological characteristics [ISO/IEC JTC1 2382-37:2012]

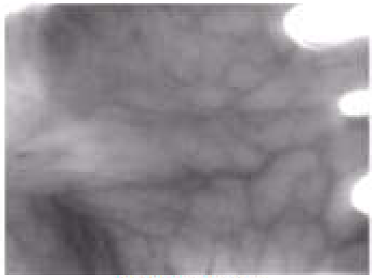
Physiological Modalities

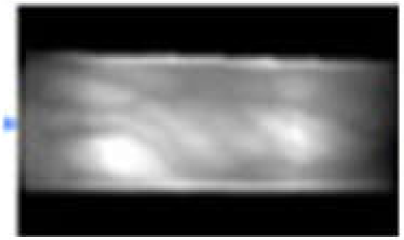

Iris


Face


Fingerprint

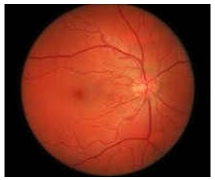

Palmprint

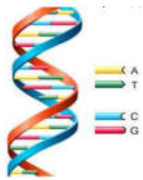

Palm vein


Finger vein


Hand geometry

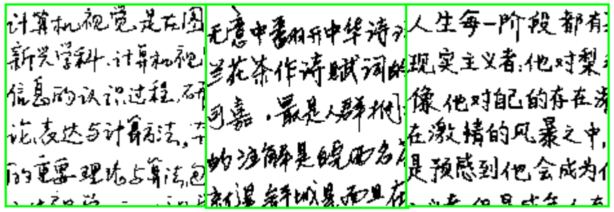

Ear

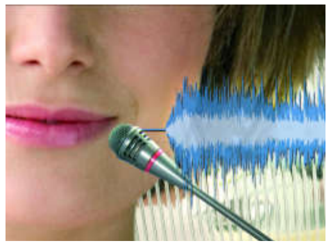

Retina


DNA

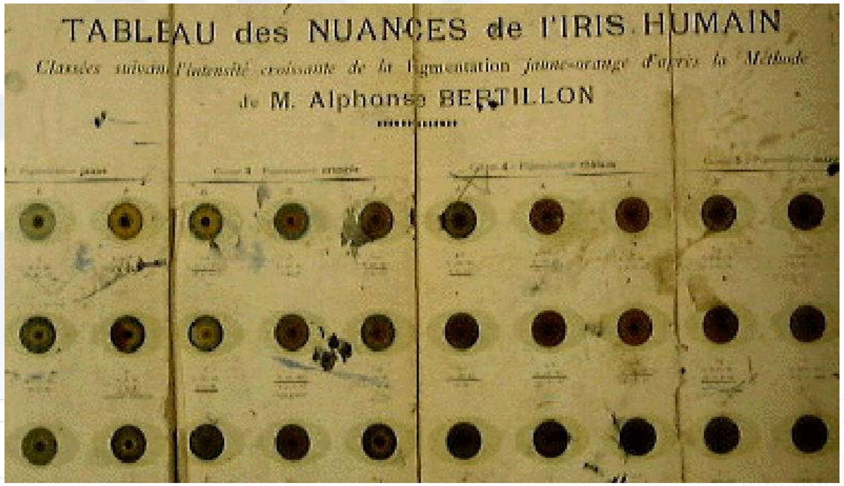
Behavioral Modalities

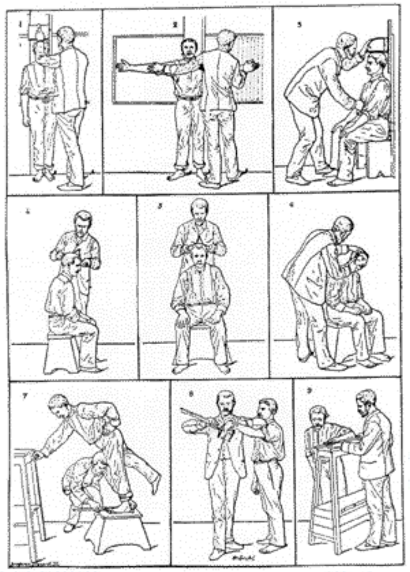

Gait

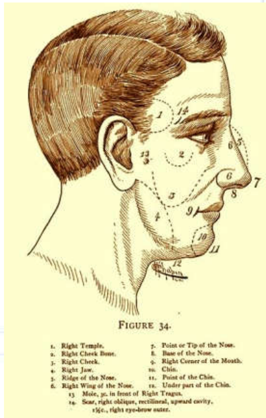

Handwriting

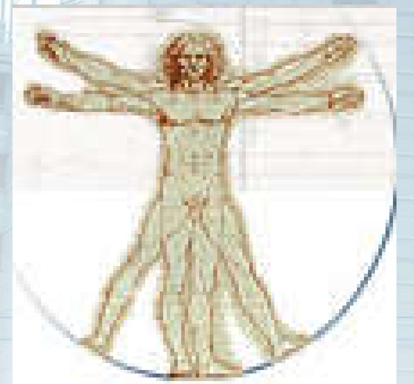

Voiceprint

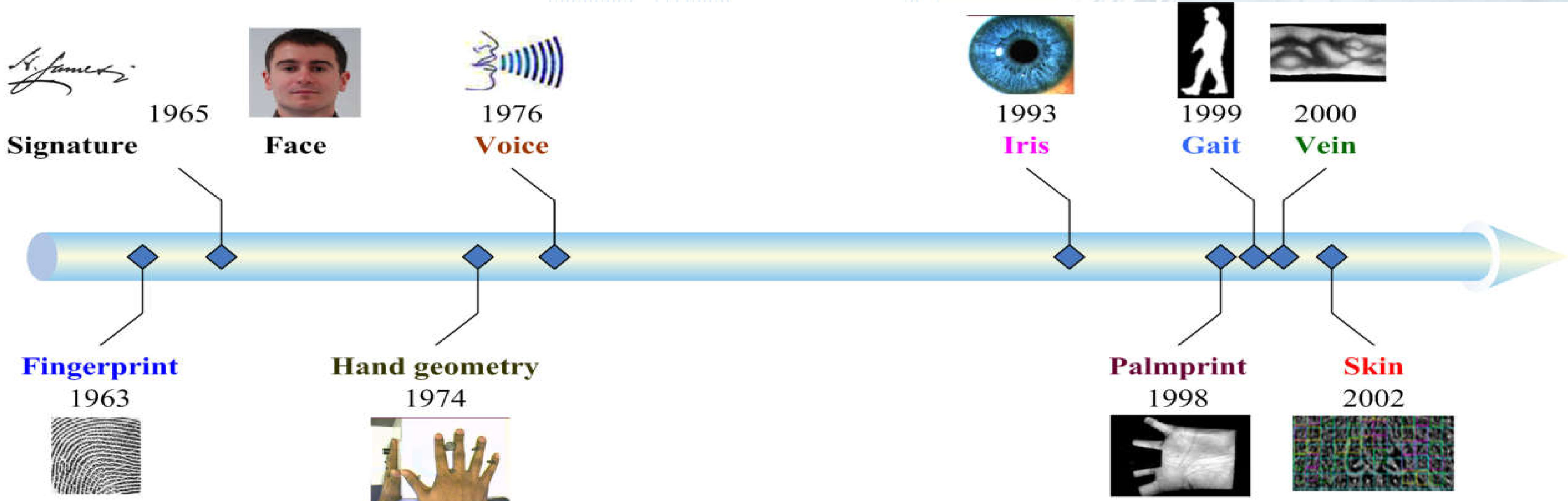
The history of biometrics



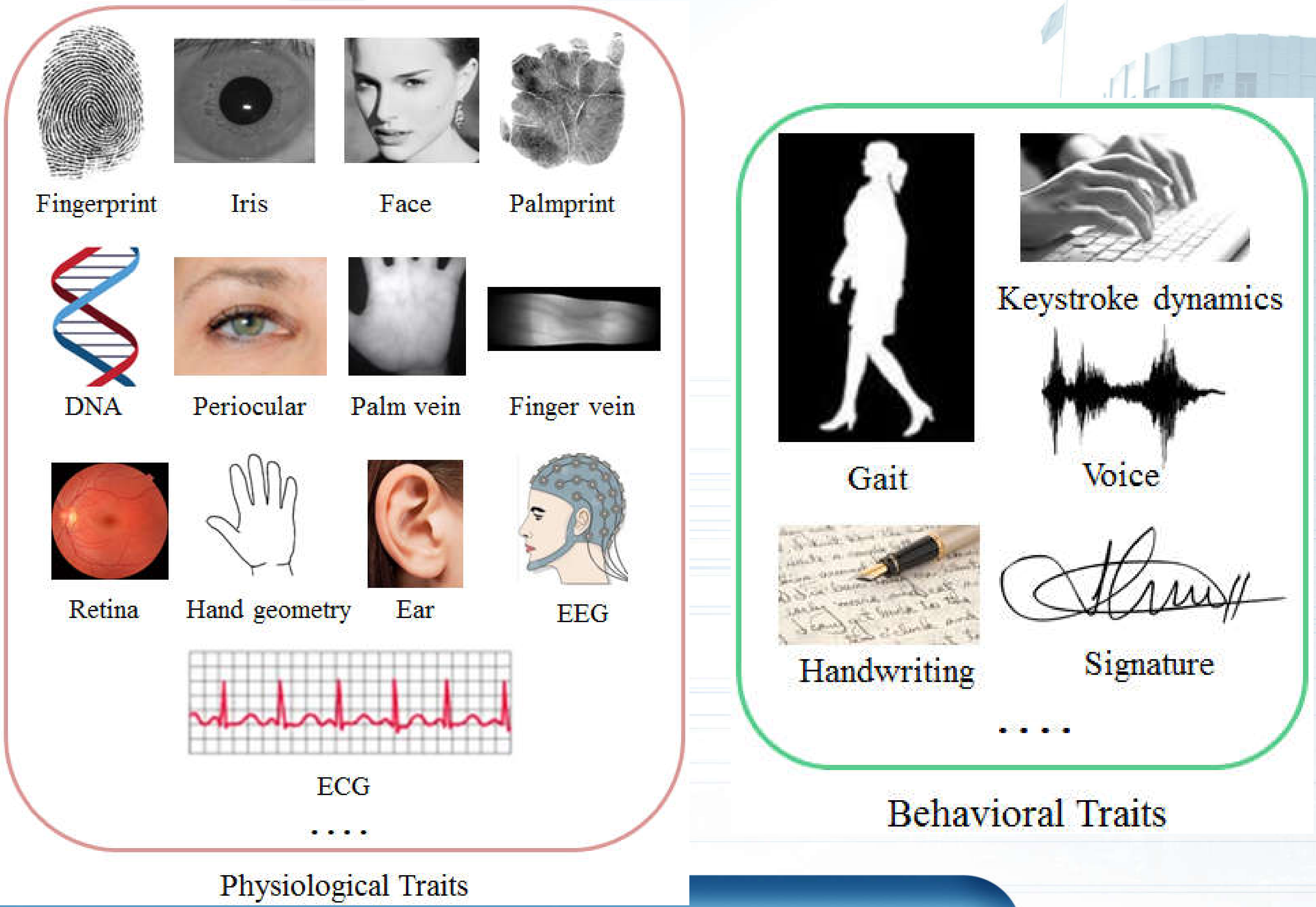




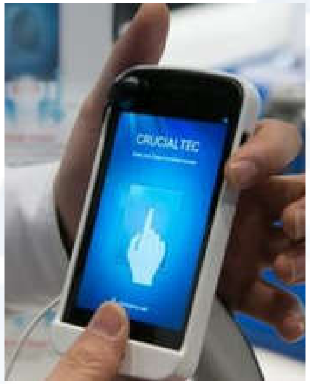




Main biometric modalities



Applications of Biometrics



Fingerprint recognition for
mobile authentication

Face recognition for
border control

Iris recognition for
coal miner identification

Finger vein recognition for
ATM authentication

Voiceprint recognition
for payment

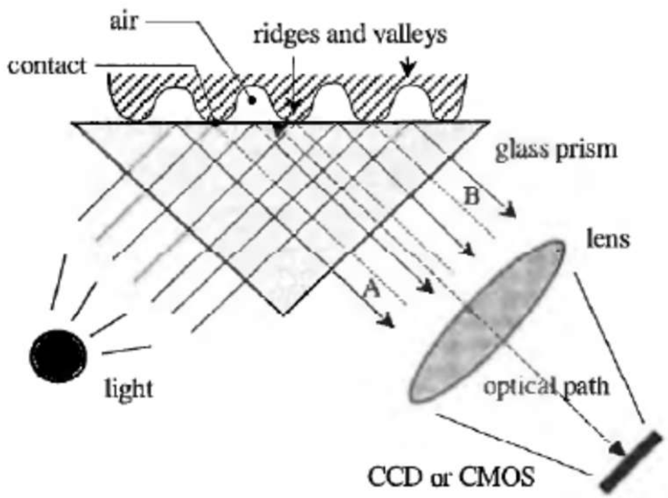
Signature verification
for credit card security

Fast Growing Market of Biometric Recognition

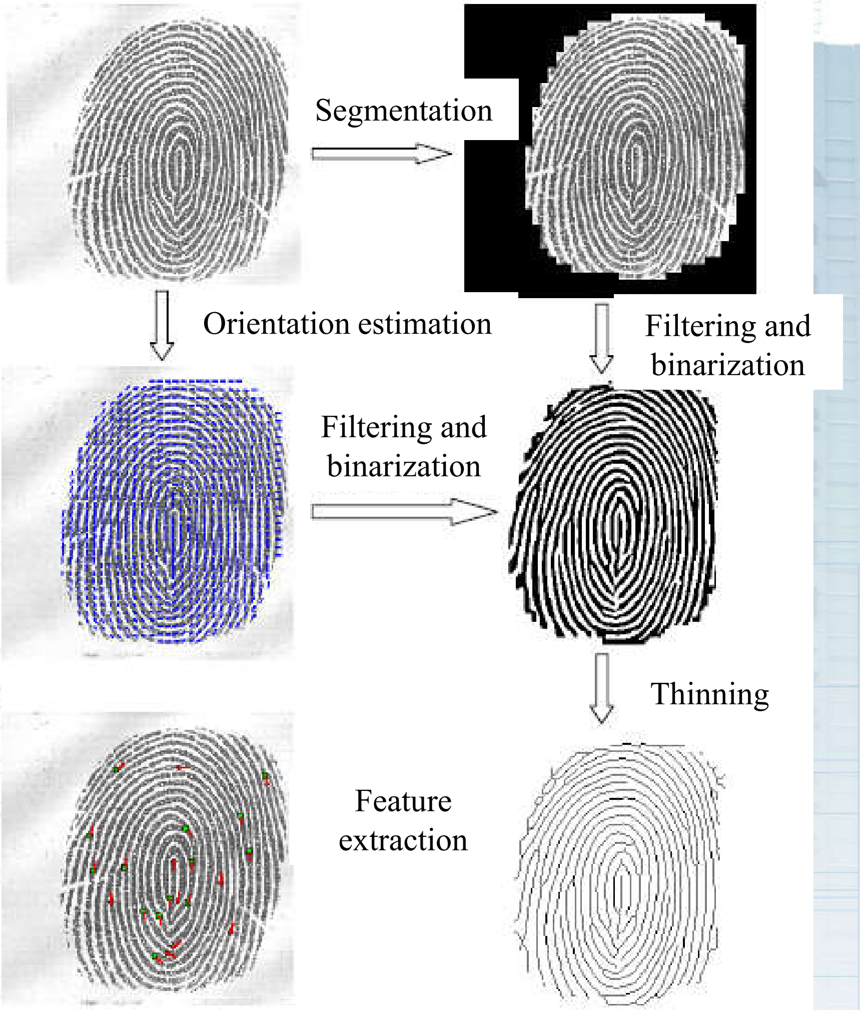
Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

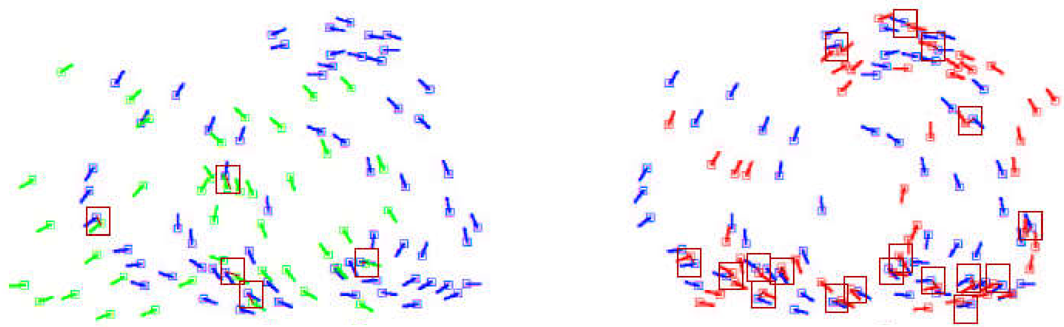
Fingerprint Recognition



Imaging



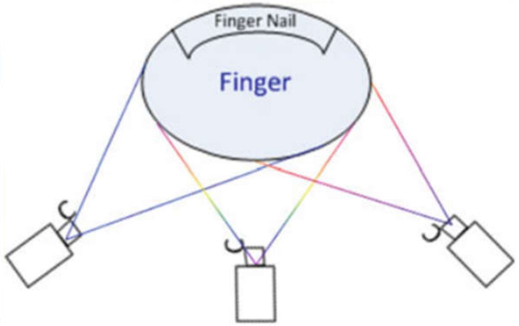
Preprocessing and feature extraction



Minutiae matching

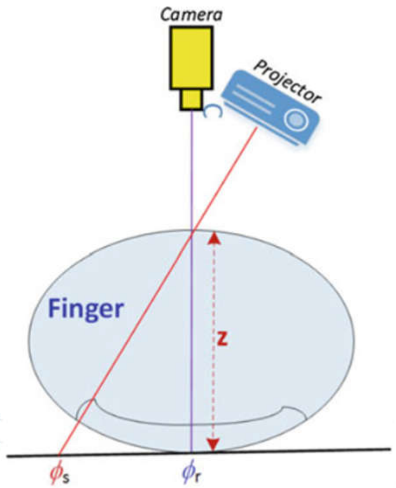
Fingerprint sensing

- 3D fingerprint



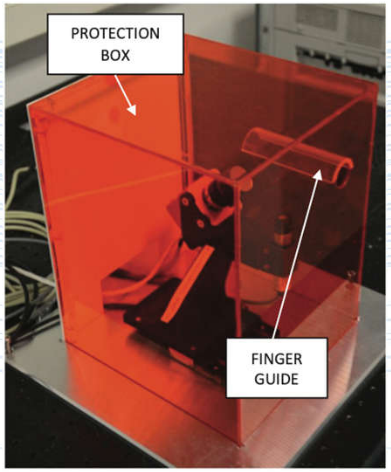
Multiple cameras

Liu and Zhang PR 2014
Labati et al. TSMC-S 2016



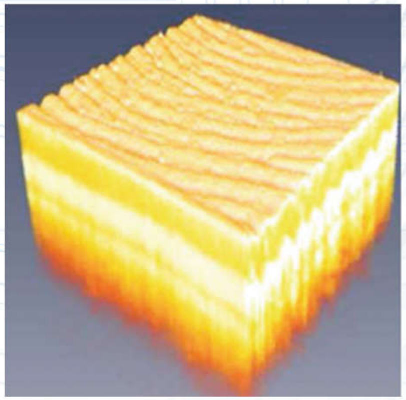
Structured lighting illumination

Wang et al. TIFS 2010
Huang et al. Opt Laser 2014
Chatterjee et al. Opt Laser 2017



Laser sensing

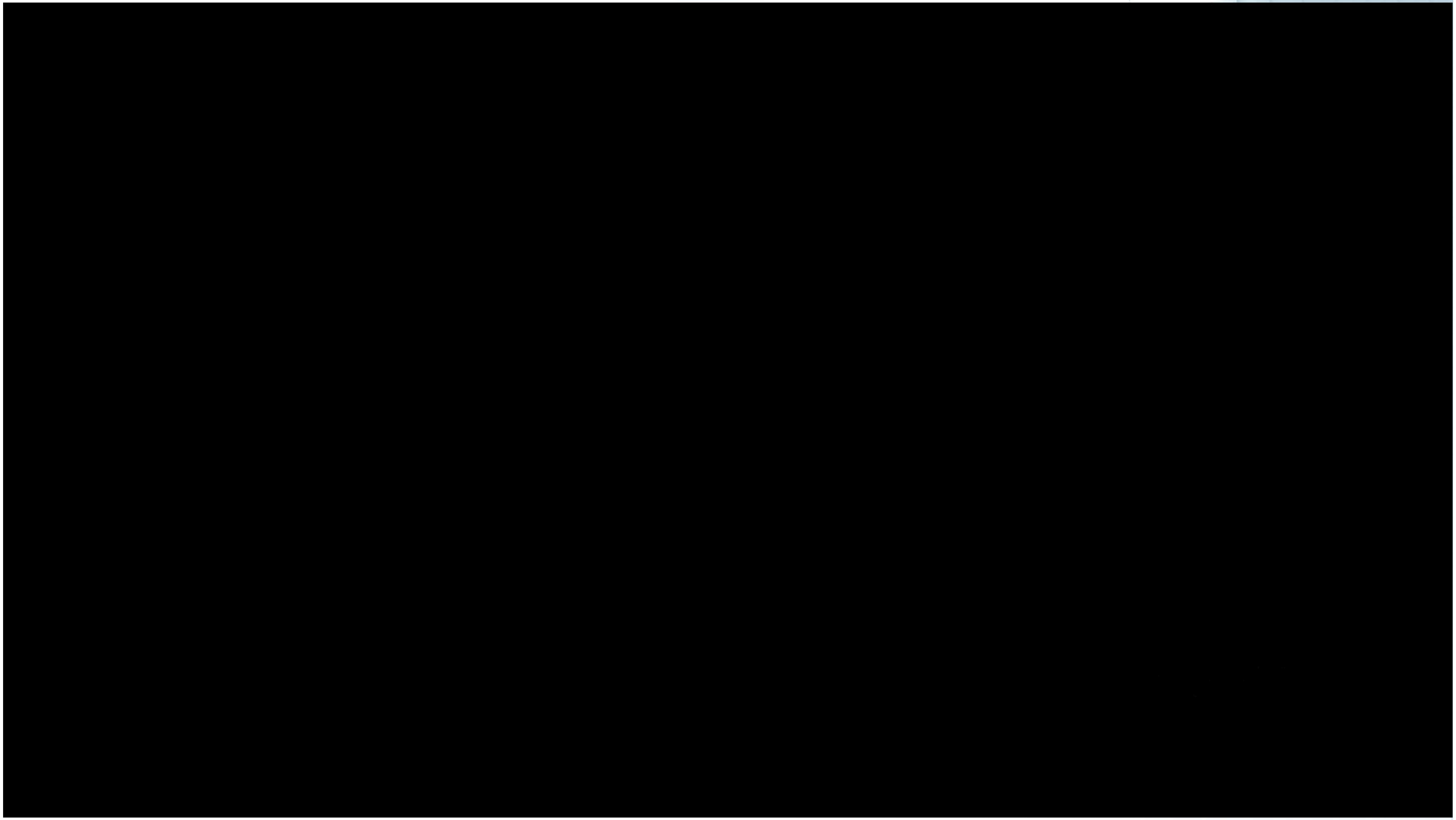
Galbally et al. IJCB 2017



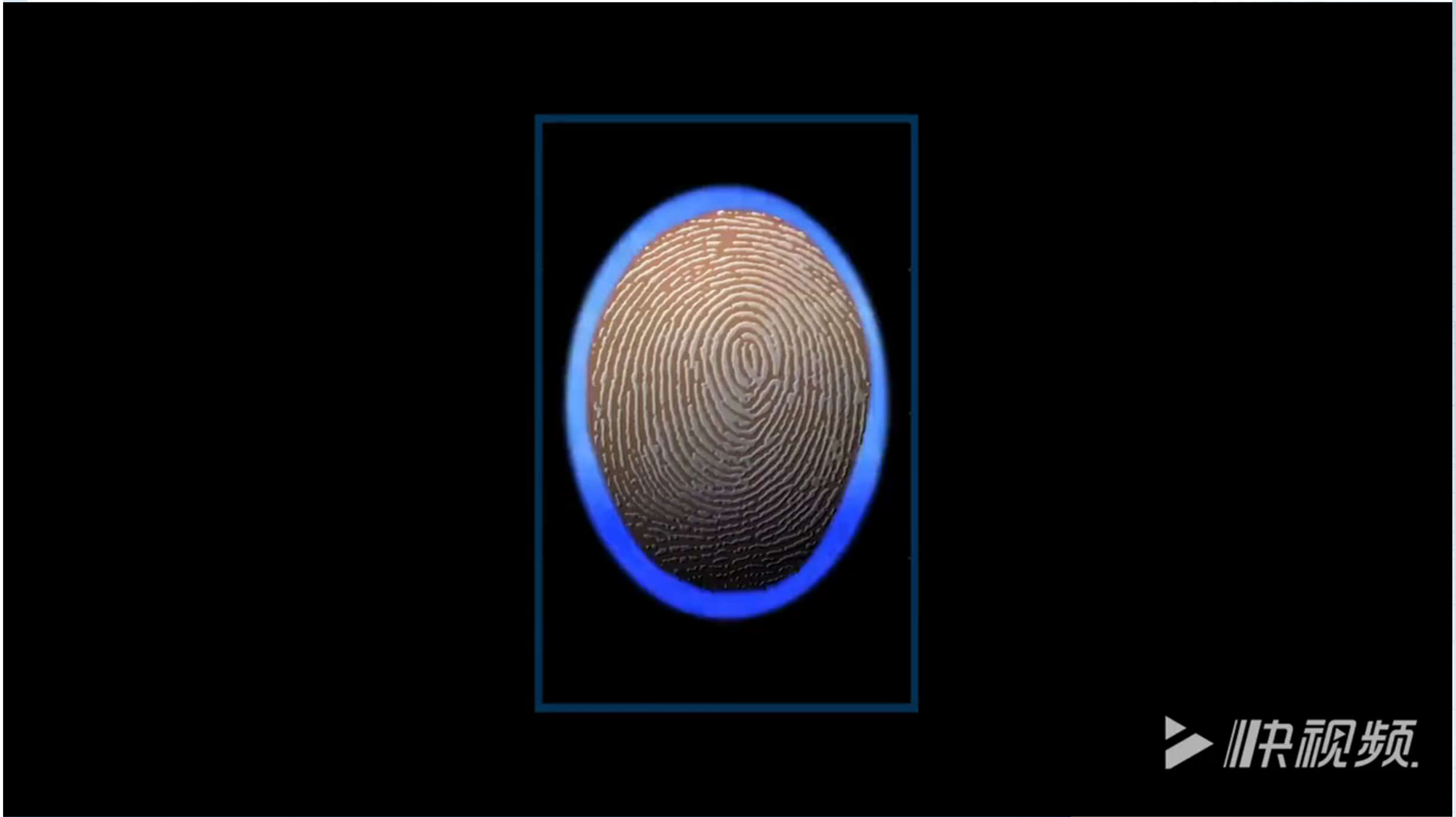
OCT

Costa et al. ICIP 2016
Nehaus et al. Boimed Opt 2017
Anksorius et al. Biomed Opt 2017

Touchless 3D Fingerprint Recognition (SAFRAN Morph)

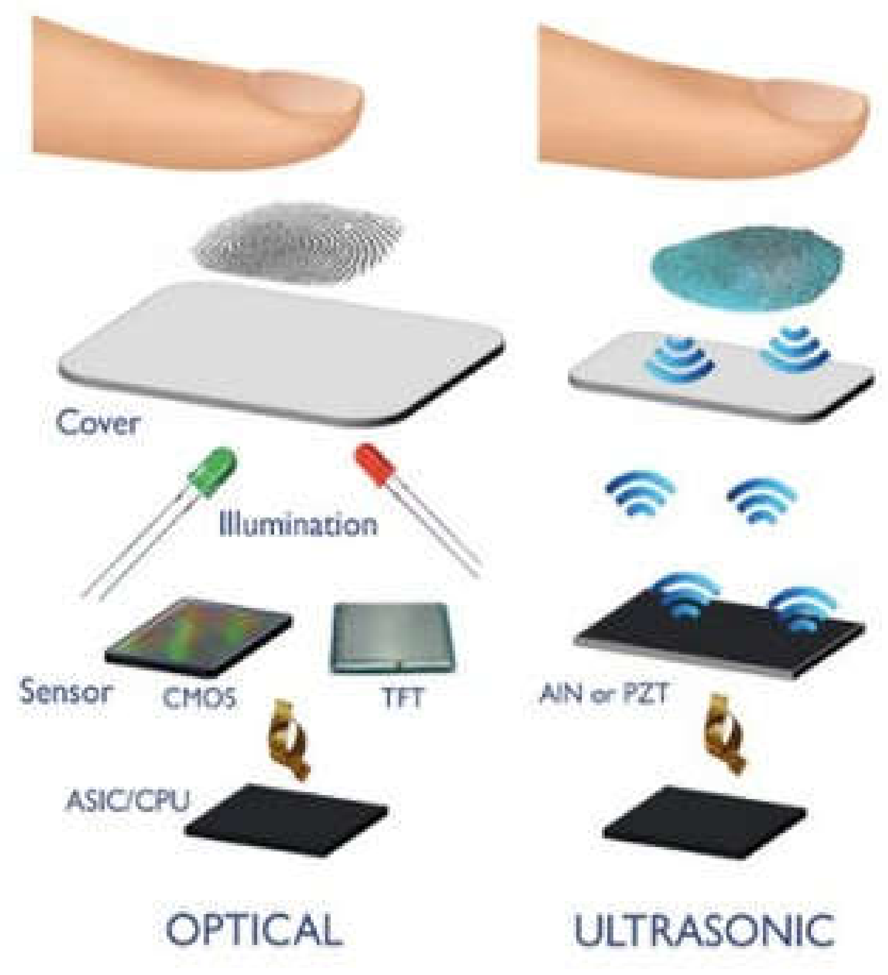


Multispectral imaging for anti-spoofing (Lumidigm)



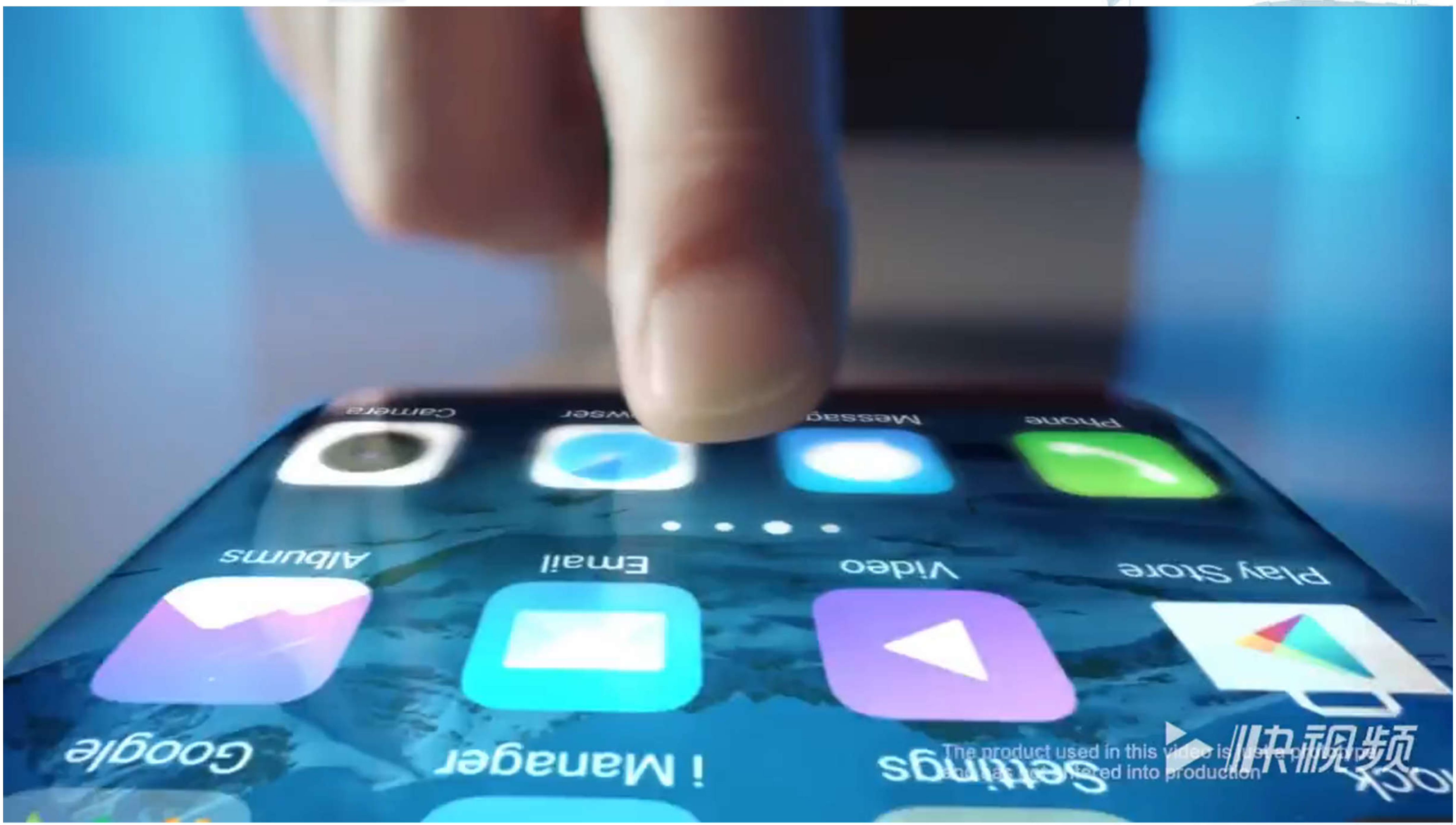
Fingerprint sensing

- Under-screen fingerprint
  - Optical based
    - Lighting required
    - Sensitive to skin conditions
    - Vendors: Synaptics, Goodix
  - Ultrasonic based
    - High quality
    - High cost
    - Vendor: Qualcomm



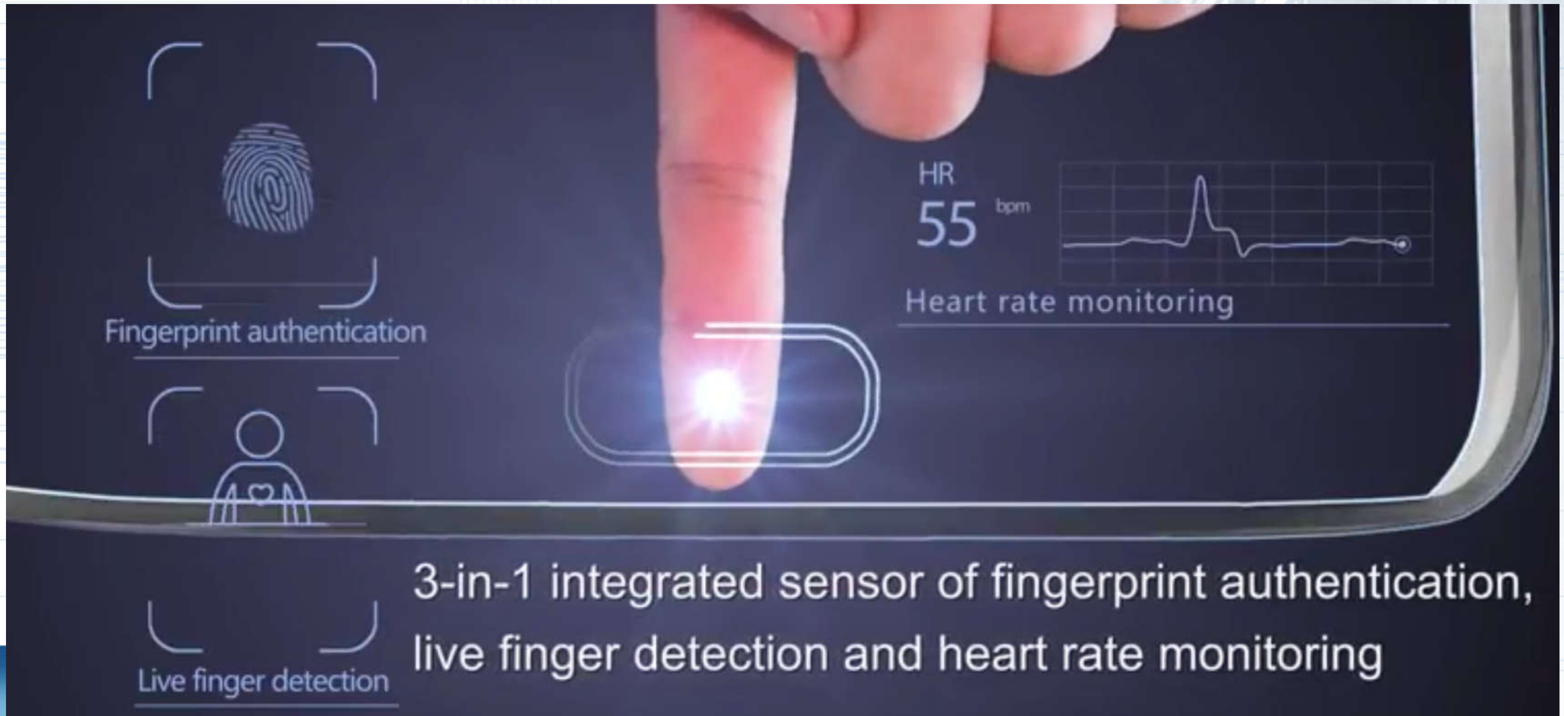
Under Display Fingerprint Scanning

(Qualcomm-Vivo, ultrasonic fingerprint solution, MWC2017)



IC Solution for Live Finger Detection

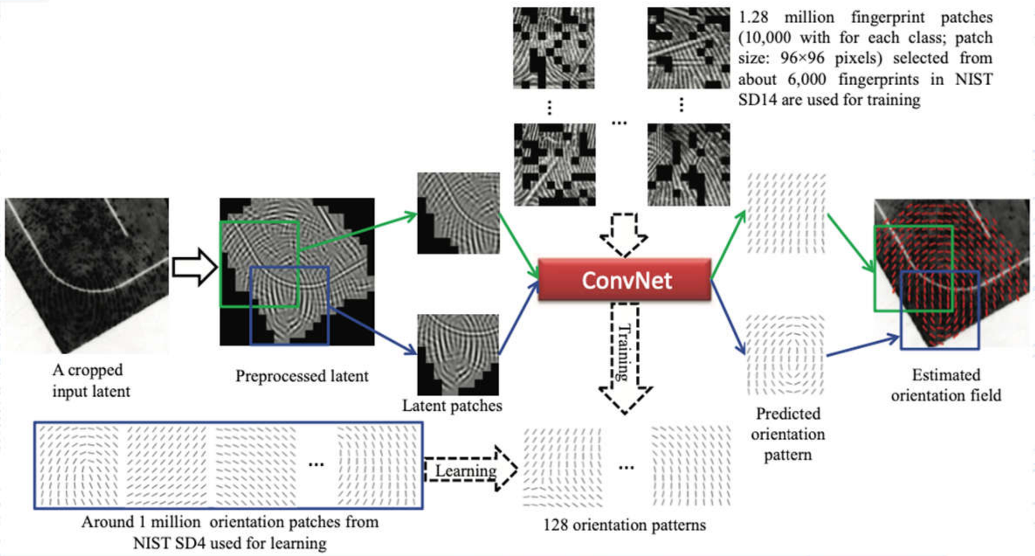
IC designer Goodix developed Live Finger Detection™ technology on mobile devices, which allows a capacitive sensor and an optical sensor to be seamlessly combined into one. Through the detection of fingerprint, blood flow and infrared signals, this cutting-edge technology embedded within the sensor is able to authenticate the user's identity and reject faked fingerprints.

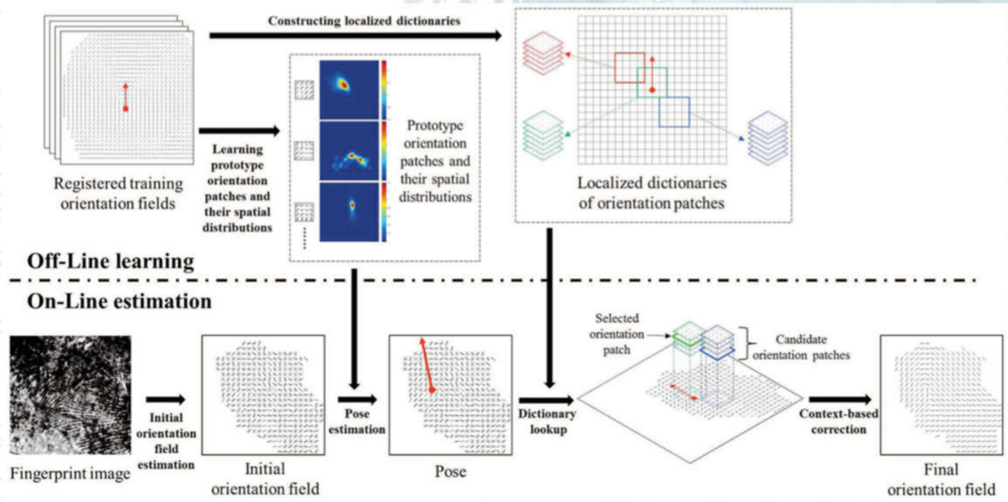


Fingerprint feature extraction

- Orientation field
  - Global dict (Feng et al. TPAMI 2012)
  - Local dict (Yang et al. TPAMI 2014)
  - Patch classification (Cao and Jain ICB 2015)

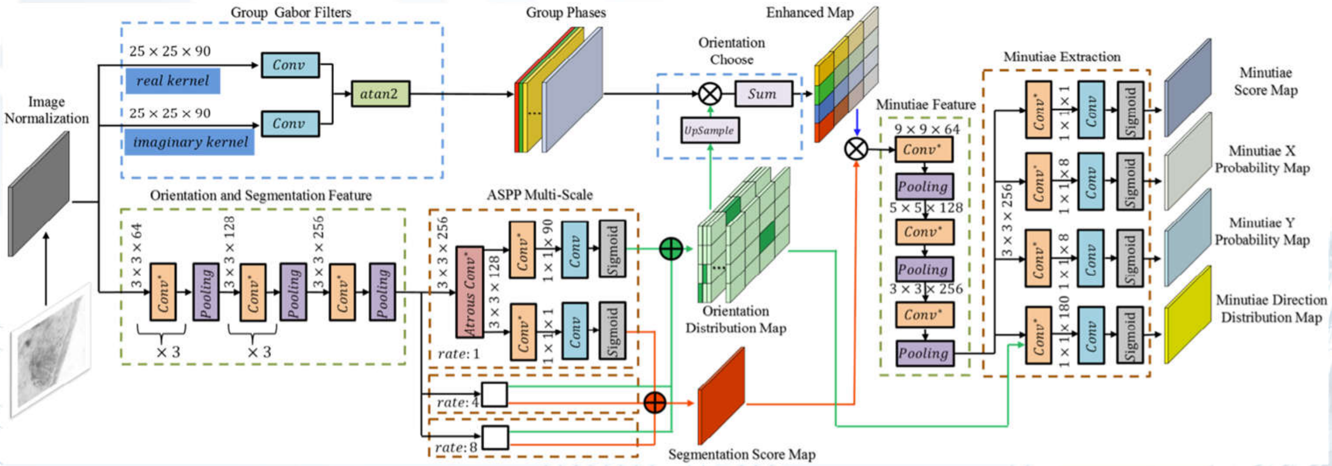
- Fingerprint pose
  - Local dict (Yang et al. TPAMI 2014)
  - Joint singular and pose (Yin et al. TIFS 2021)





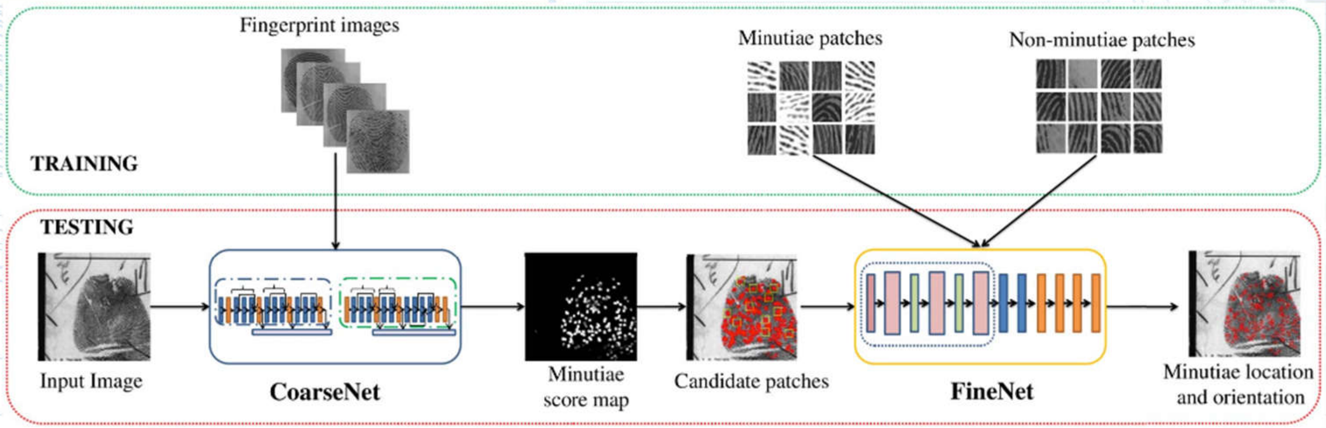
Fingerprint feature extraction

- Minutiae
  - FingerNet (Tang et al. IJCB 2017)



Traditional processing → CNN

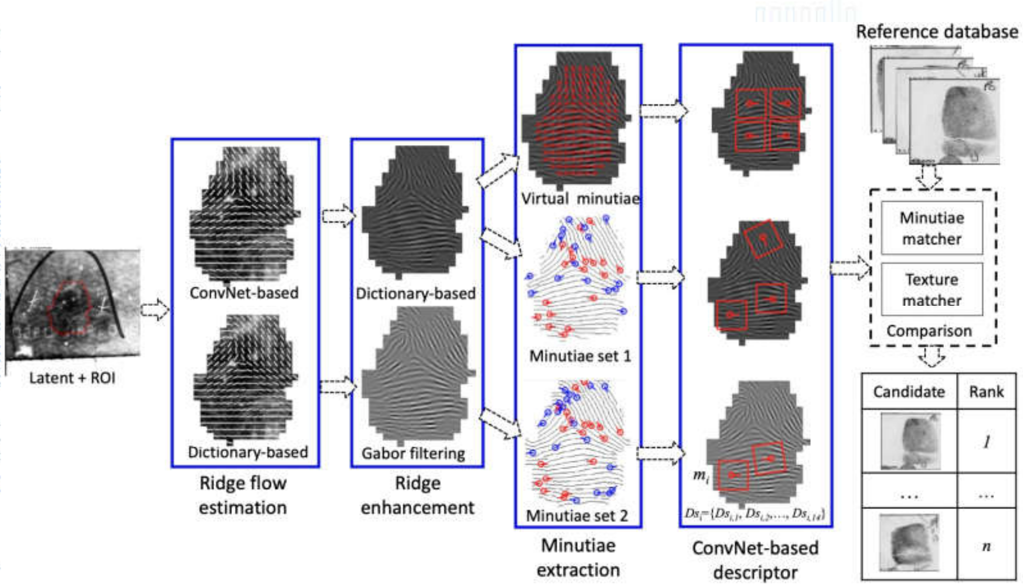
  - MinutiaeNet (Nguyen et al. ICB 2018)

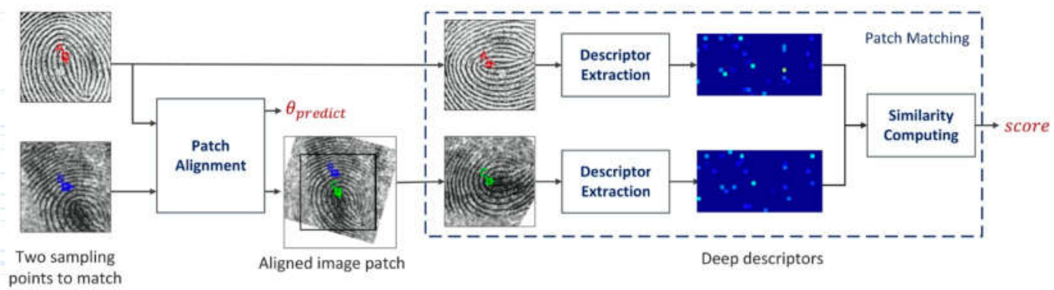


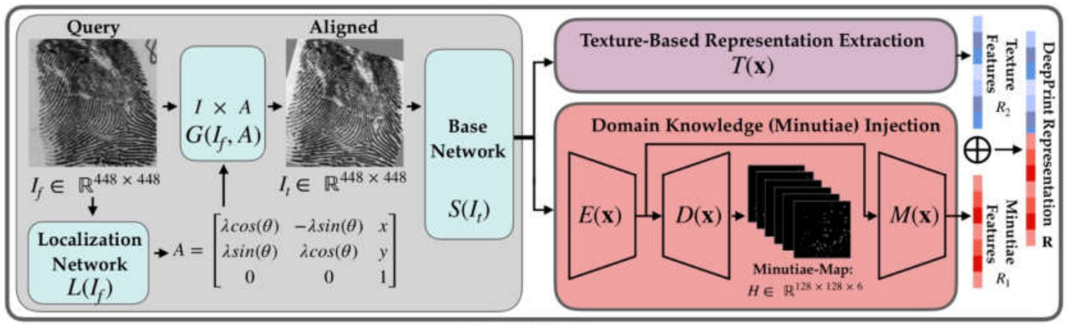
CoarseNet
Prior based minutiae distribution
↓
FineNet
Local patch fine-tuning

Fingerprint matching

- Latent fingerprint matching
  - LatentAFIS (Cao and Jain TPAMI 2019)
  - Densely sampled points (Gu et al. TIFS 2020)
  - Fixed-Length representation by DeepPrint (Engelsma et al. TPAMI 2019)



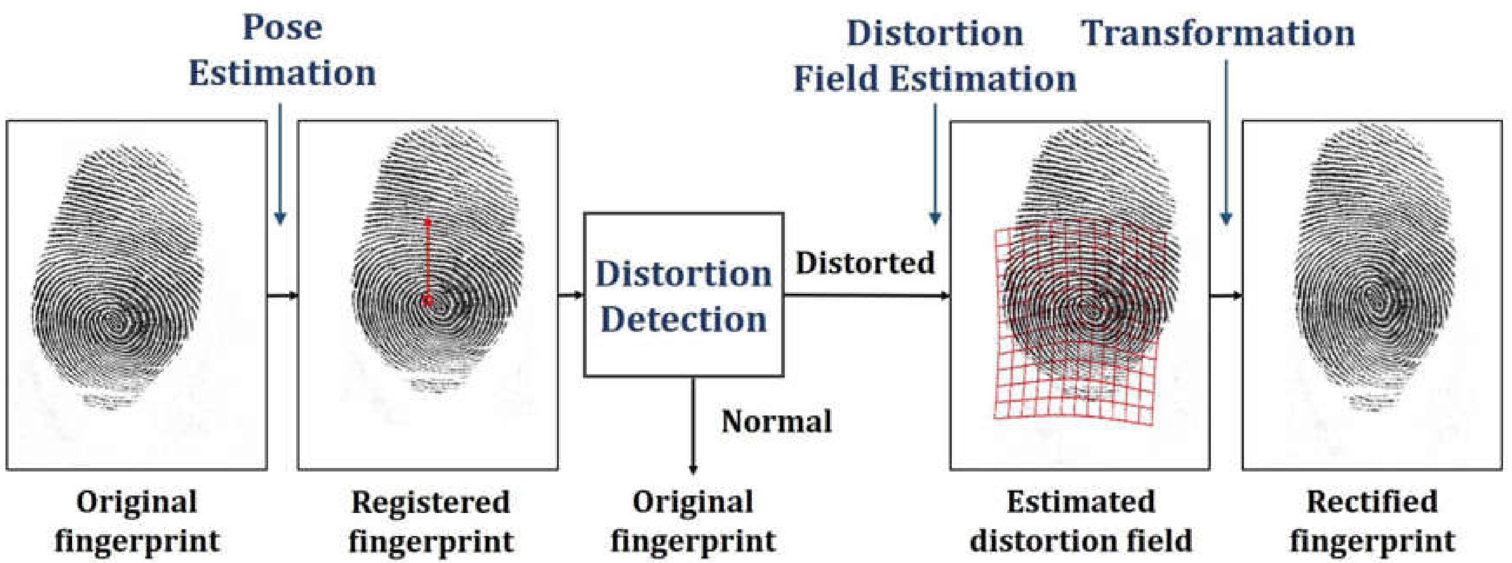




DeepPrint

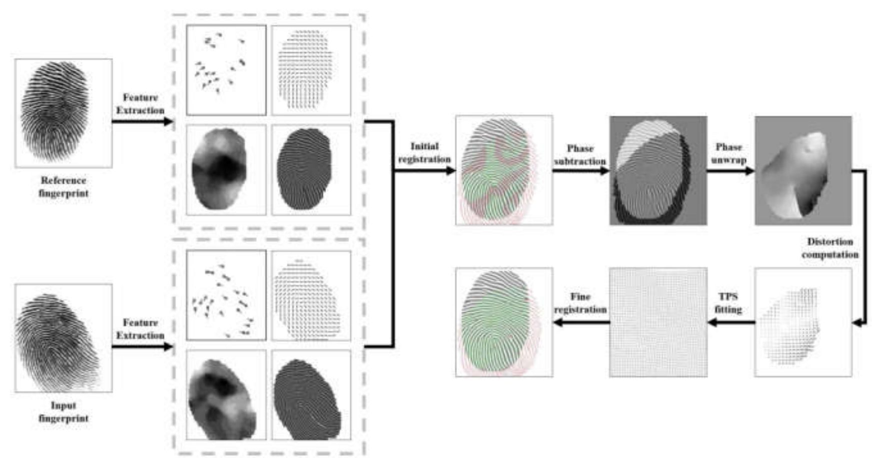
Fingerprint distortion rectification

- Fingerprint distortion rectification
  - Nearest neighbor search (Si et al. TPAMI 2015)
  - Regression (Gu et al. TIFS 2018)
  - DCNN (Dabouei et al. ICB 2018)

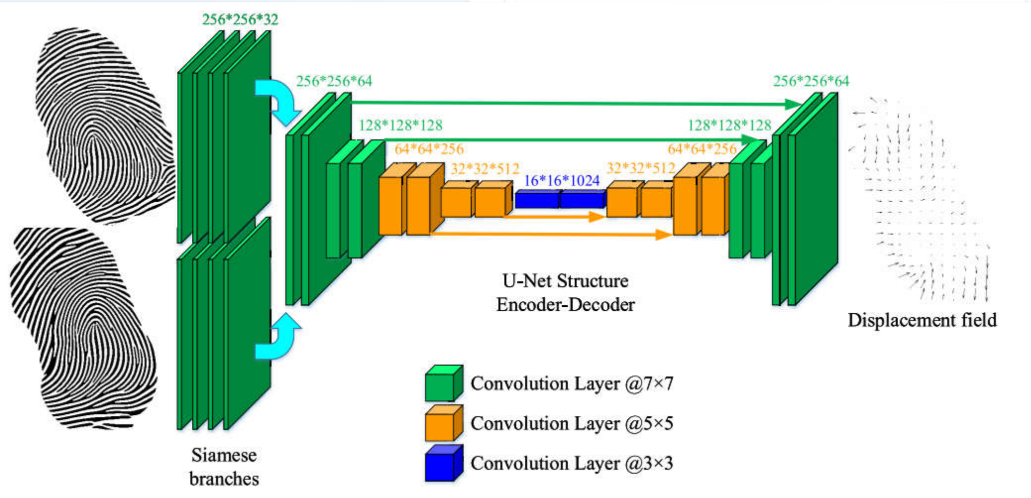


Fingerprint dense registration

- Fingerprint dense registration
  - Phase demodulation (Cui et al. TIFS 2018)

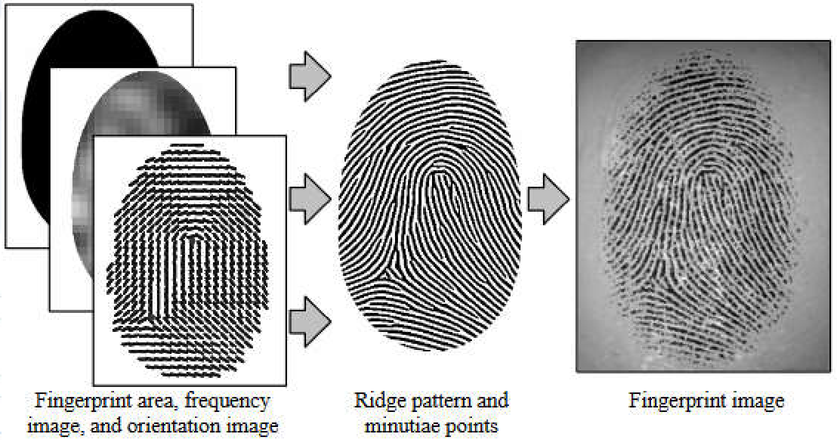


  - Deep learning based (Cui et al. TIFS 2021)

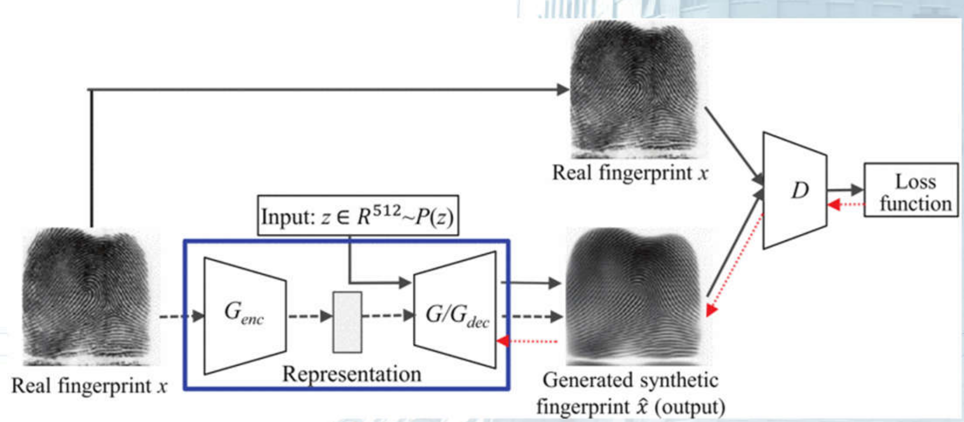


Fingerprint synthesis and spoof detection

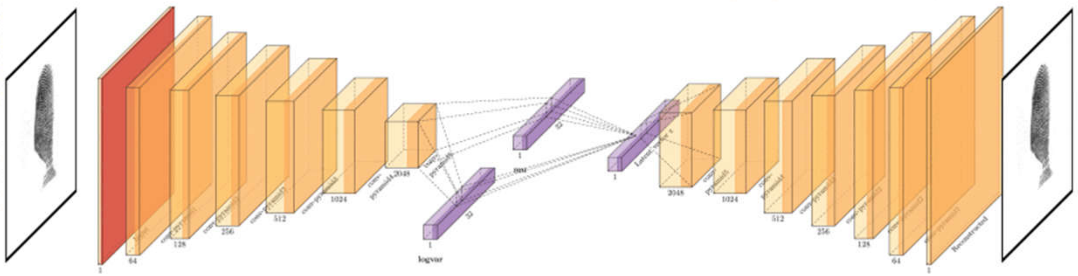
- Fingerprint synthesis
  - SFinGe (Cappelli et al. ICPR 2000)



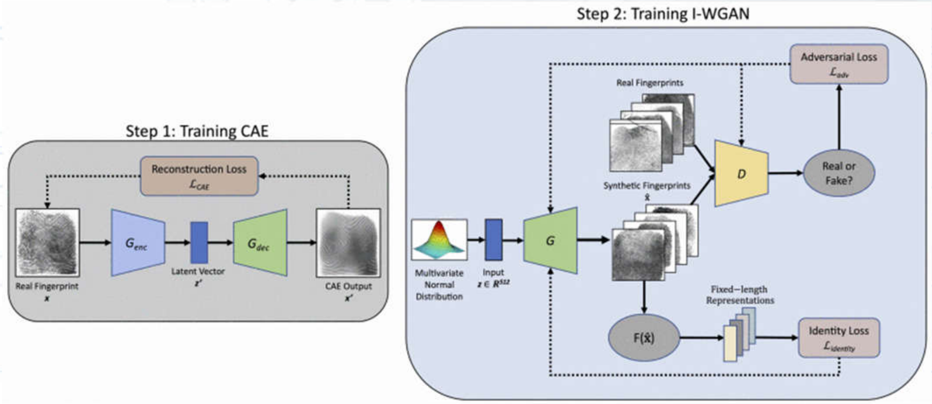
  - IWGAN and Autoencoder (Cao et al. ICB 2018)



  - Variational Autoencoder (Attia et al. SMC 2019)

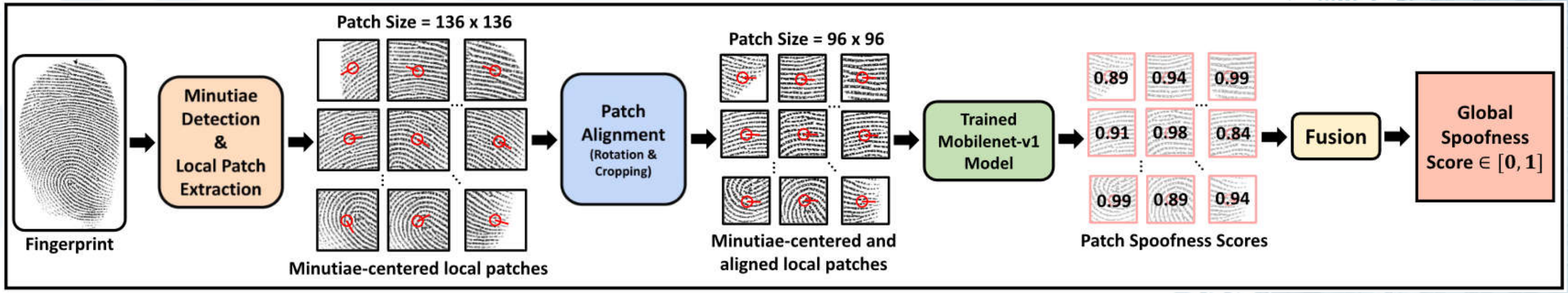


  - IWGAN and Autoencoder with Identity Loss (Mistry et al. IJCB 2020)

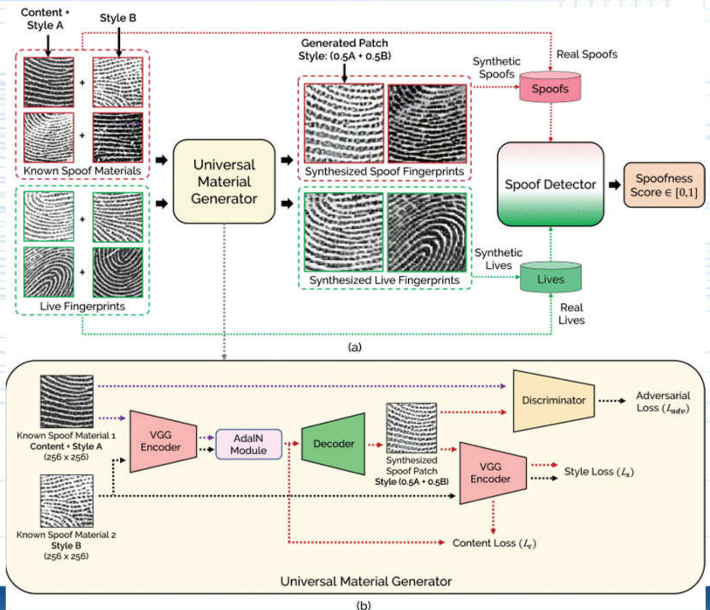


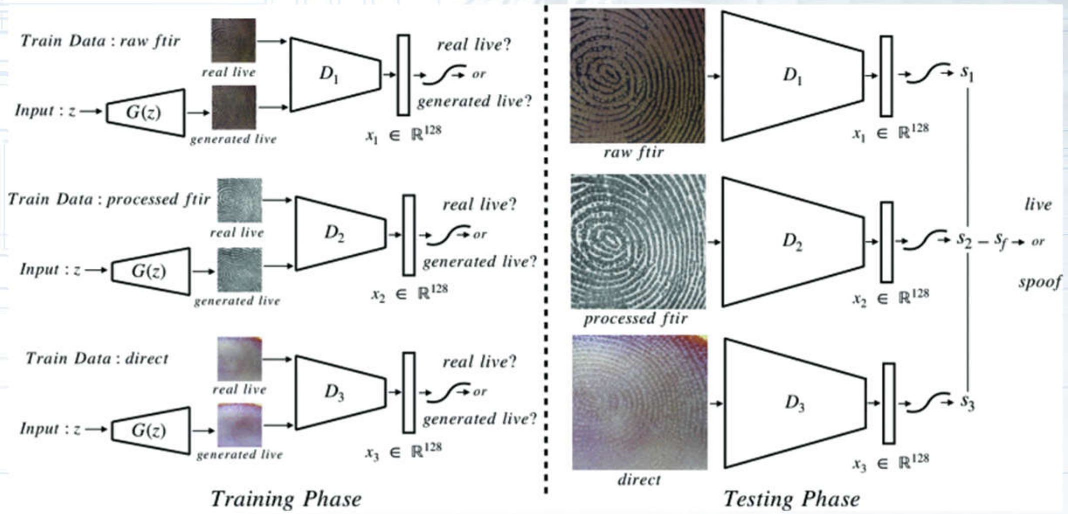
Fingerprint synthesis and spoof detection

- Fingerprint spoof detection
  - Fingerprint Spoof Buster (Chugh et al. TIFS 2018)

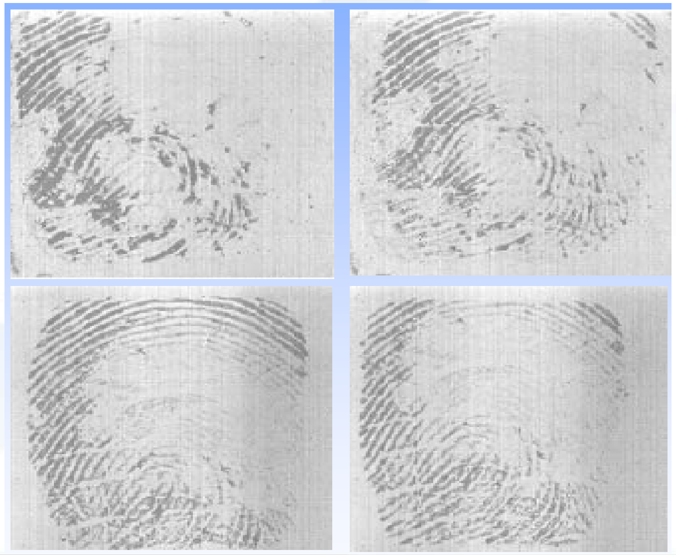


  - Universal Material Generator (Chugh and Jain TIFS 2020)
  - GAN-based spoof detection (Engelsma and Jain ICB 2020)

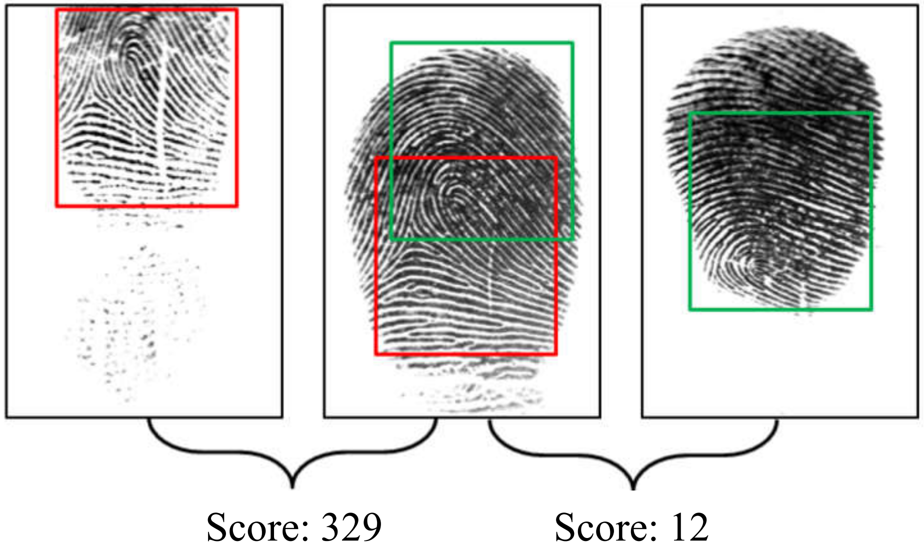




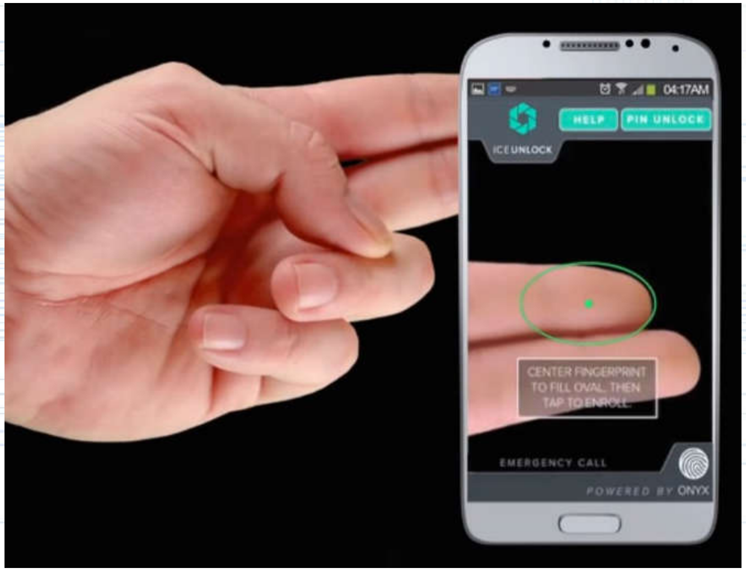
Open Problems of Fingerprint Recognition



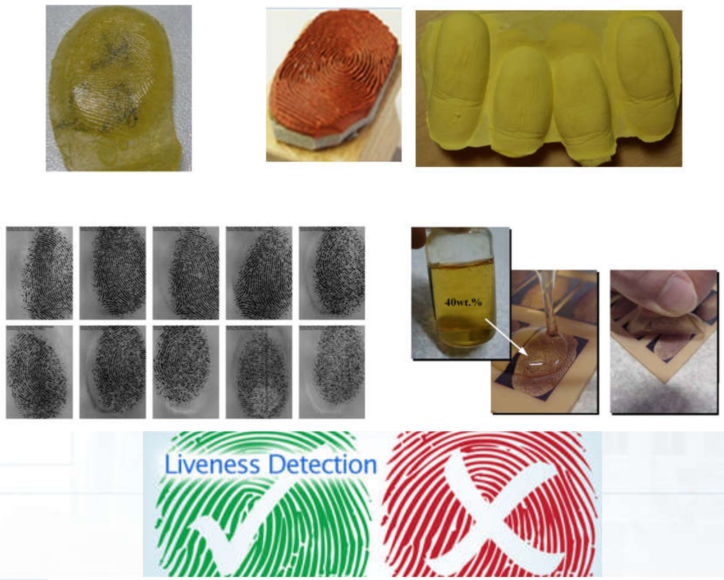
Latent fingerprint images



Distorted fingerprint images



Touchless fingerprint recognition

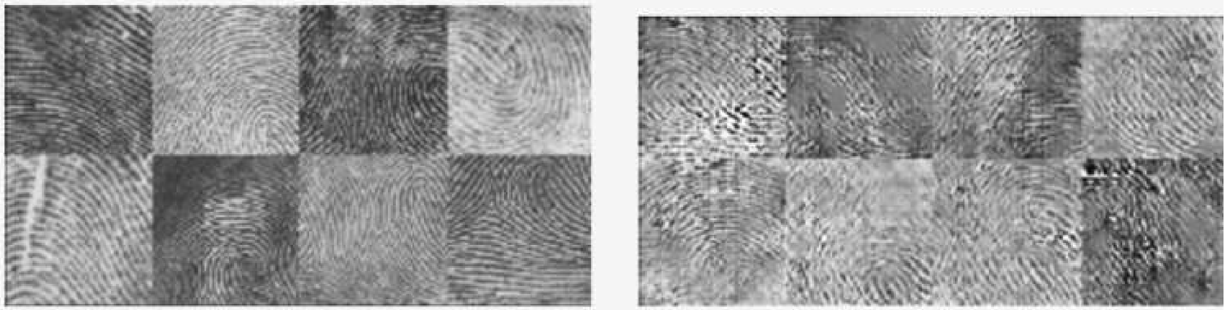


Fingerprint liveness detection

Open Problems of Fingerprint Recognition

By AstroJane

Posted on November 19, 2018



(a) Real (left) and generated (right) samples for the NIST dataset.

(b) Real (left) and generated (right) samples for the FingerPass capacitive dataset.

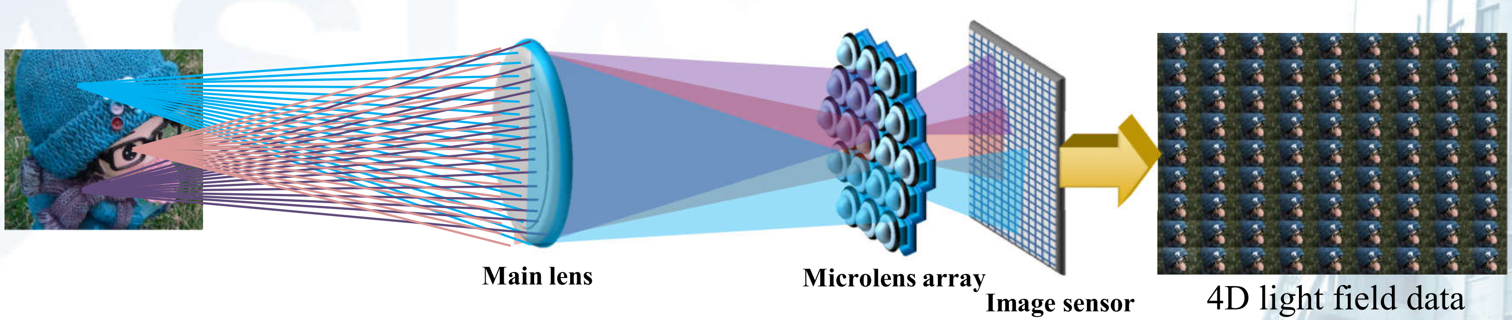
Computer scientists at New York University and Michigan State University have trained an artificial neural network to create fake digital fingerprints that can bypass locks on cell phones. The fakes are called "DeepMasterPrints", and they present a significant security flaw for any device relying on this type of biometric data authentication. After exploiting the weaknesses inherent in the ergonomic needs of cellular devices, DeepMasterPrints were able to imitate over 70% of the fingerprints in a testing database.

Philip Bontrager, Aditi Roy, Julian Togelius, Nasir Memon, Arun Ross, DeepMasterPrints: Generating MasterPrints for Dictionary Attacks via Latent Variable Evolution，IEEE BTAS 2018.

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

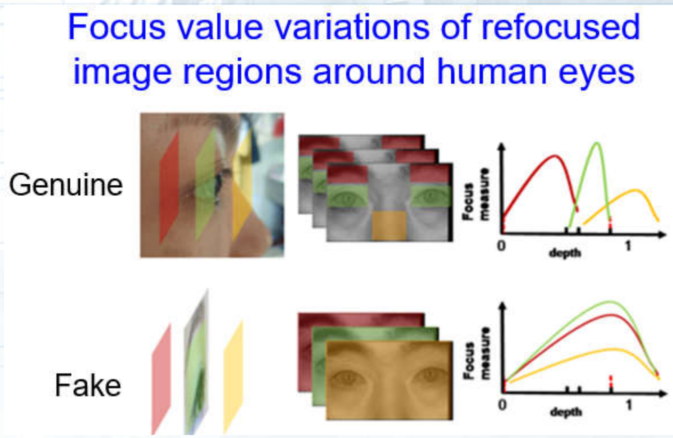
Iris Recognition Based on Light Field Imaging



Light-field Camera (Plenoptic Camera)

Extending depth of field

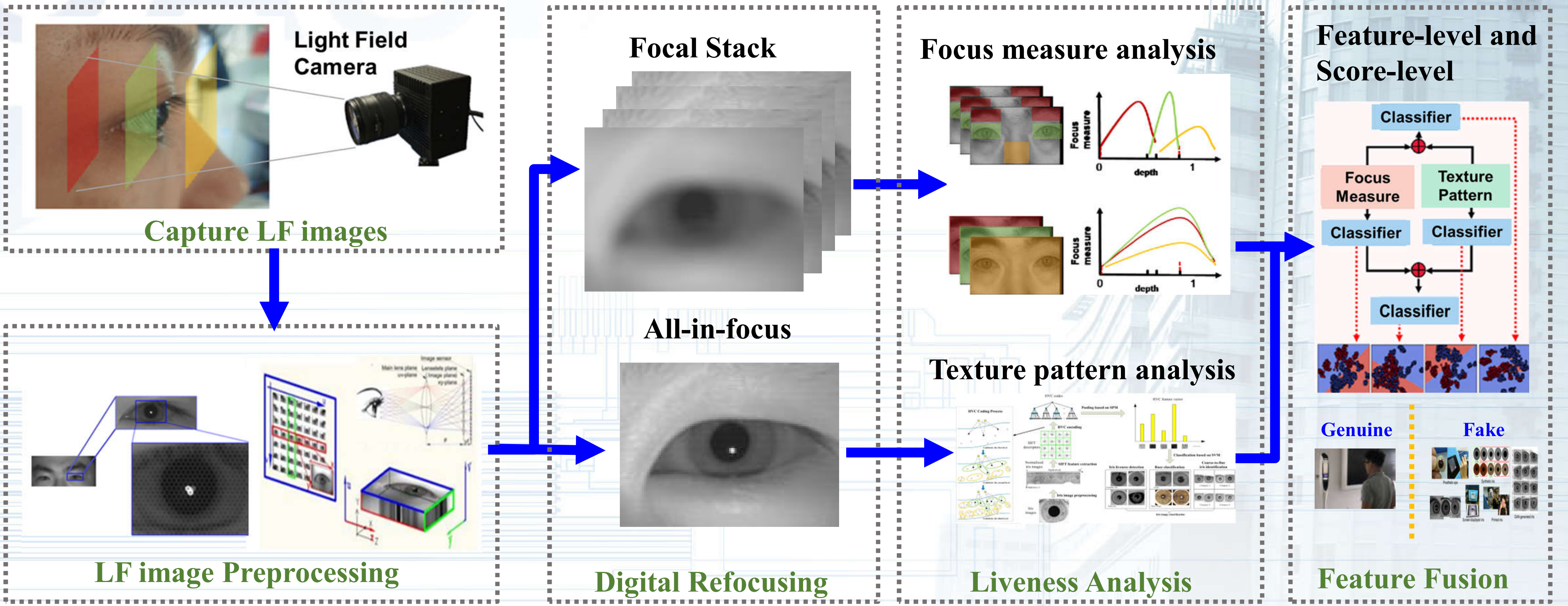
Depth perception



Liveness detection

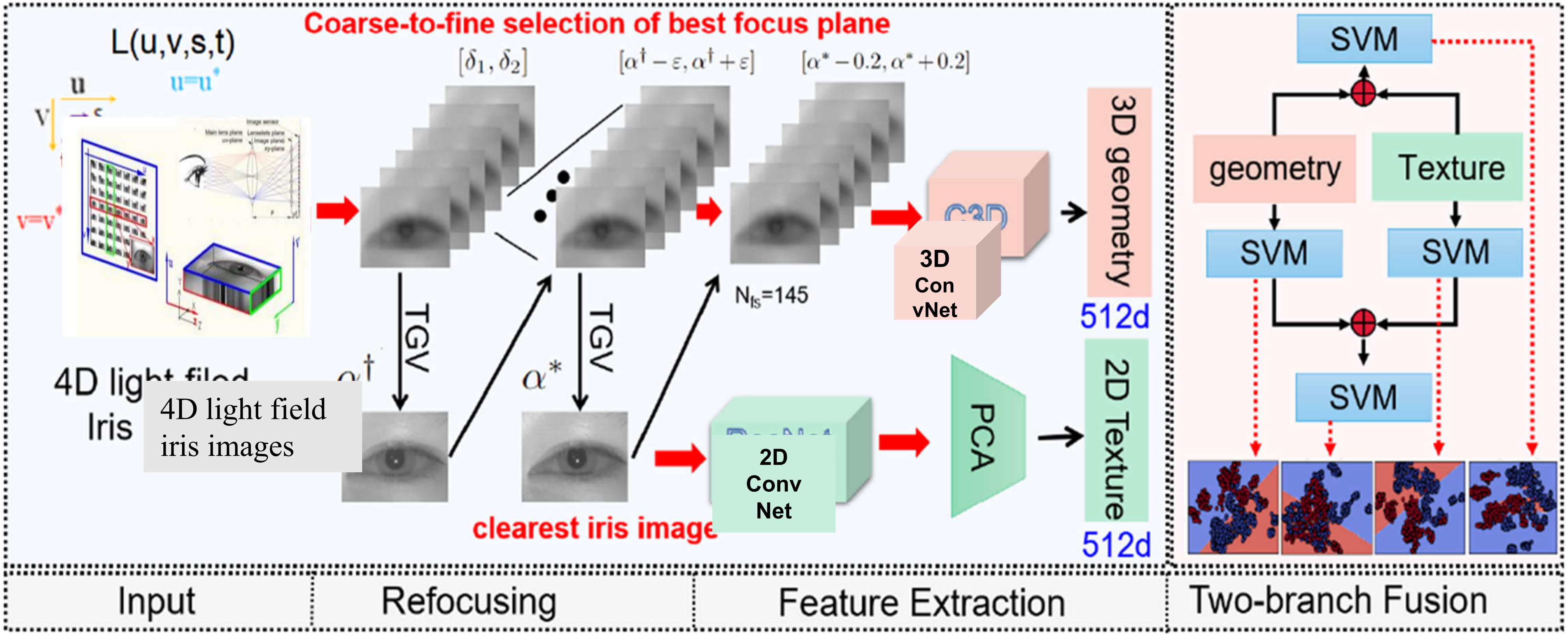
Iris Liveness Detection Using Light Field Cameras

Analysis on light field focal stack and the all-in-focus image for iris liveness detection



Ping Song, Ling Huang, Yunlong Wang, Fei Liu, Zhenan Sun. Iris Liveness Detection Based on Light Field Imaging, IEEE/CAA Journal of Automatica Sinica (**JAS**), vol.45, no.9, pp.1701-1712, 2019.

Iris Liveness Detection using Light Field Cameras



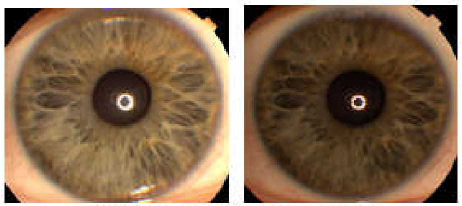
Fusion of 3D geometric structure and 2D spatial texture in light field focal stack for iris liveness detection

Zhengquan Luo, Yunlong Wang, Nianfeng Liu, Zilei Wang. "Combining 2D texture and 3D geometry features for Reliable iris presentation attack detection using light field focal stack", IET Biometrics, 2022.

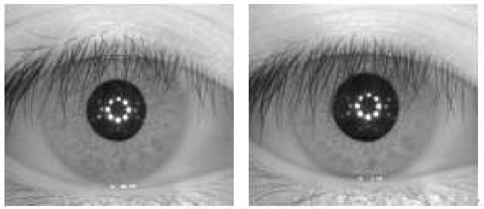
Open Problems of Iris Recognition

Less or unconstrained iris image acquisition

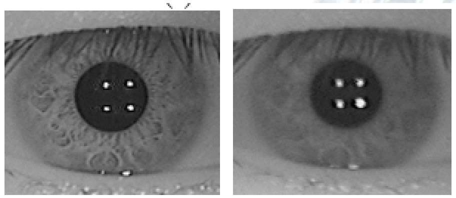
Forensic applications



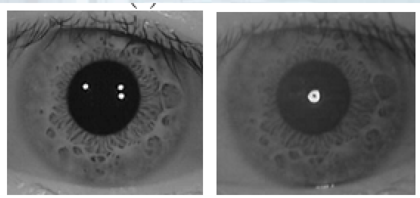
(a) Illumination changes



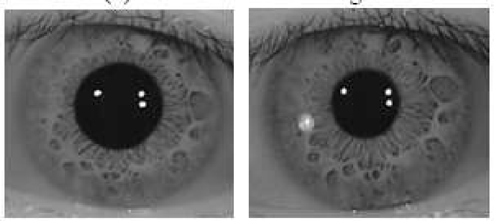
(b) Occlusions



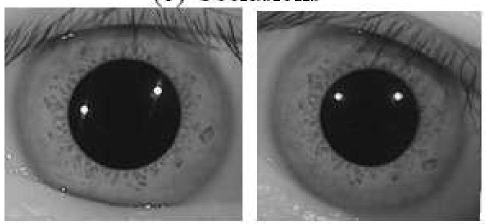
(e) Defocus



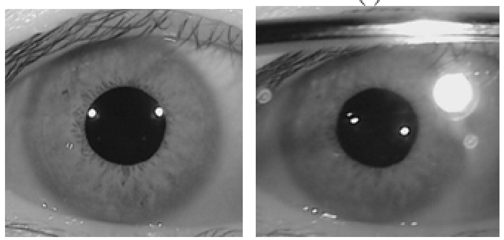
(f) Inter-sensor interoperability



(c) Deformation



(d) Rotation



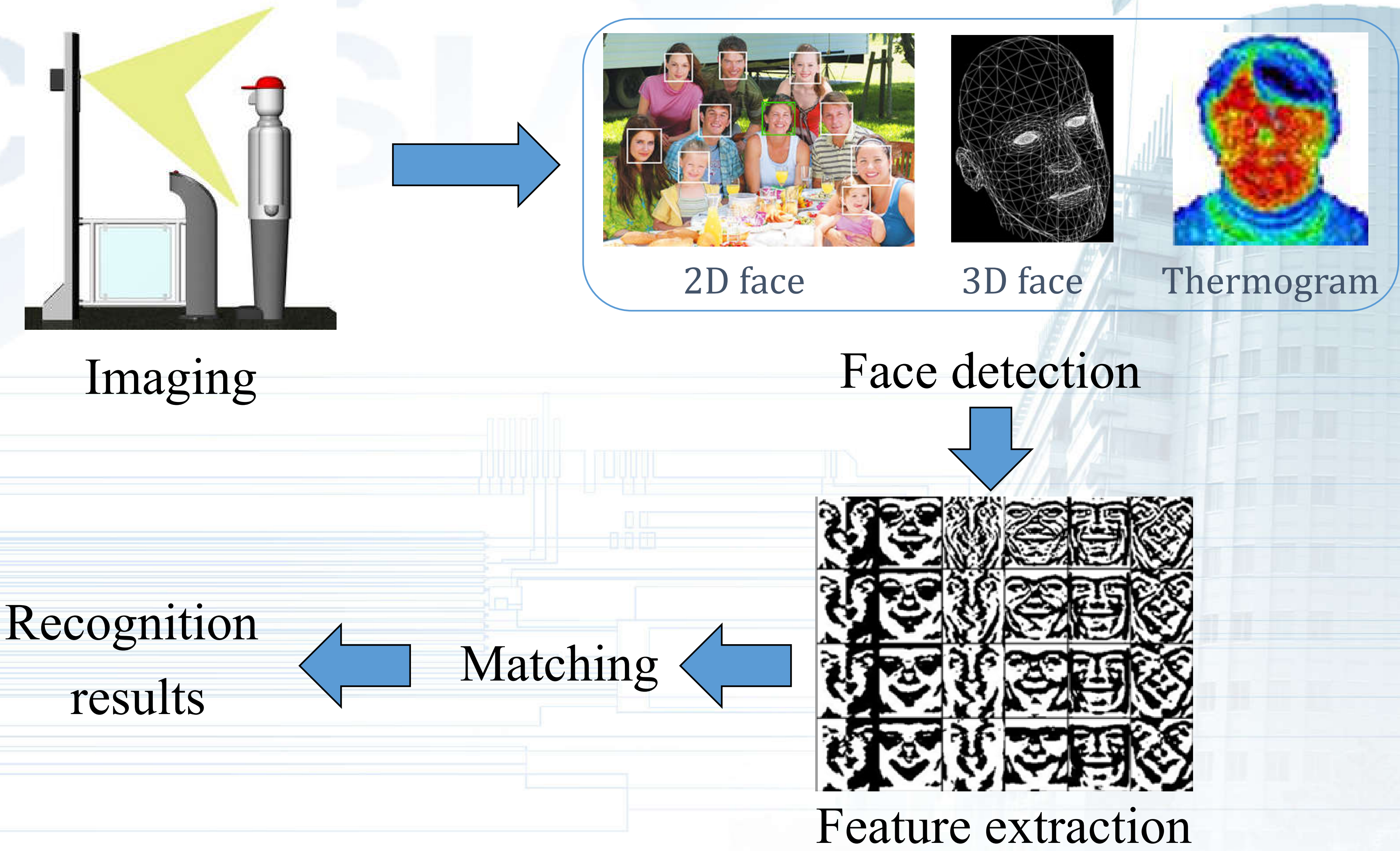
(g) Eyeglasses

Poor quality iris images

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

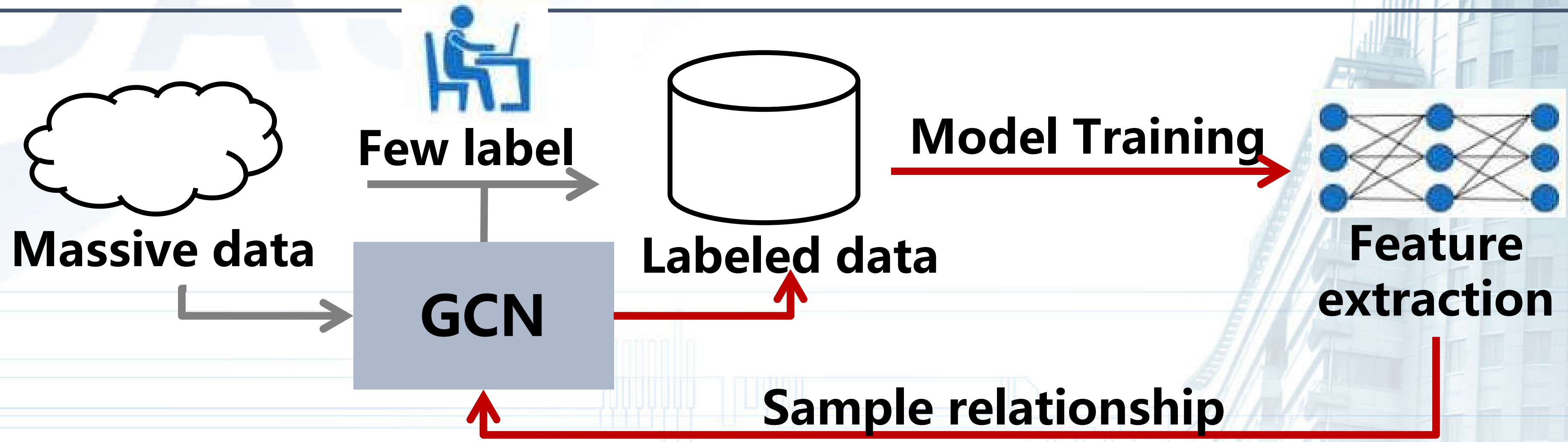
Face Recognition

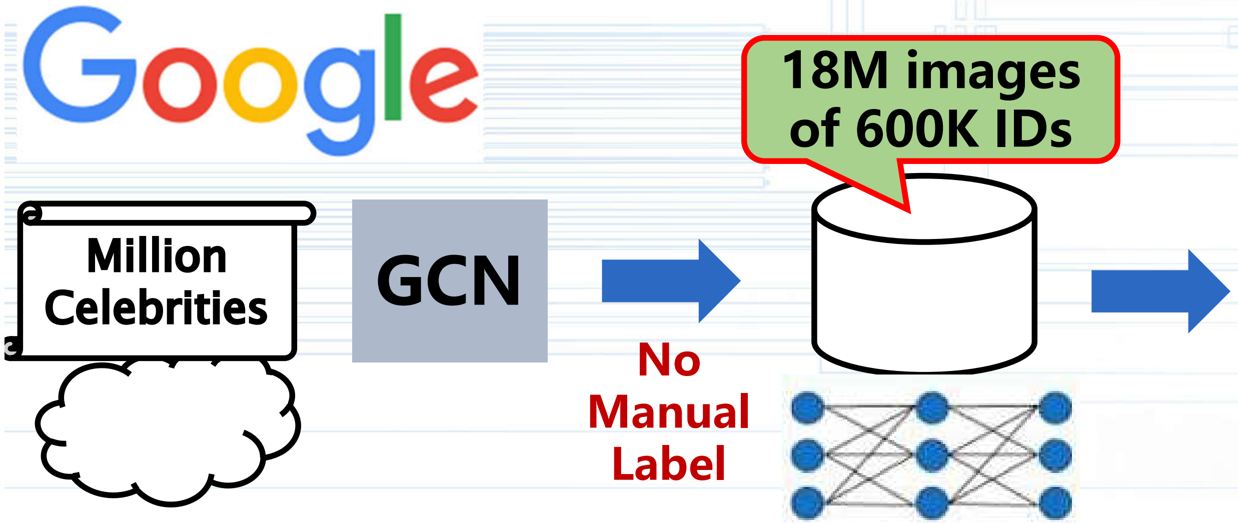


Popular methods: Gabor/LBP/Ordinal measures/Sparse representations/Deep learning

Accuracy: GCN Based Label Noise Cleansing

This work proposes a graph convolutional network (GCN) method to cleanse the results of google face search, automatically collected and labelled 18M images, and achieve SOTA performance on IJB-C by training models on the large-scale cleaned dataset.





Video Surveillance face recognition IJB-C

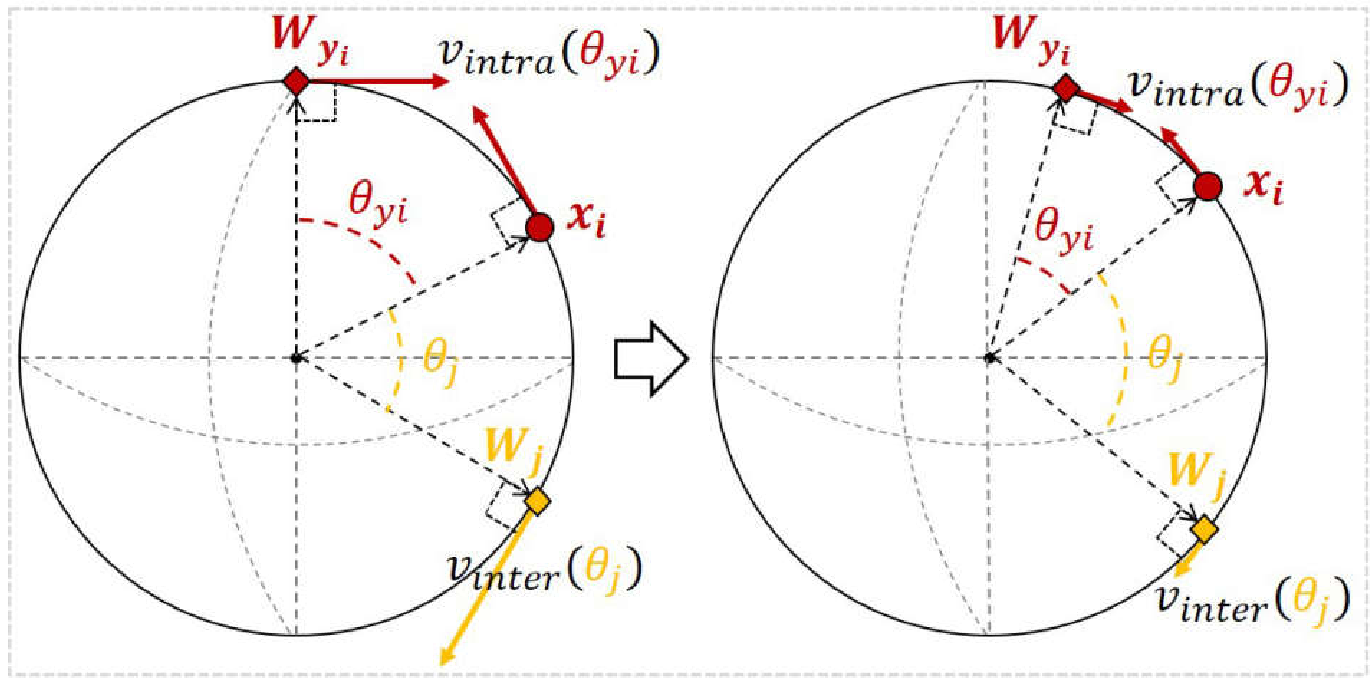
IJB-C	SOTA	Close Loop
1:N accuracy	94.44 %	96.93%

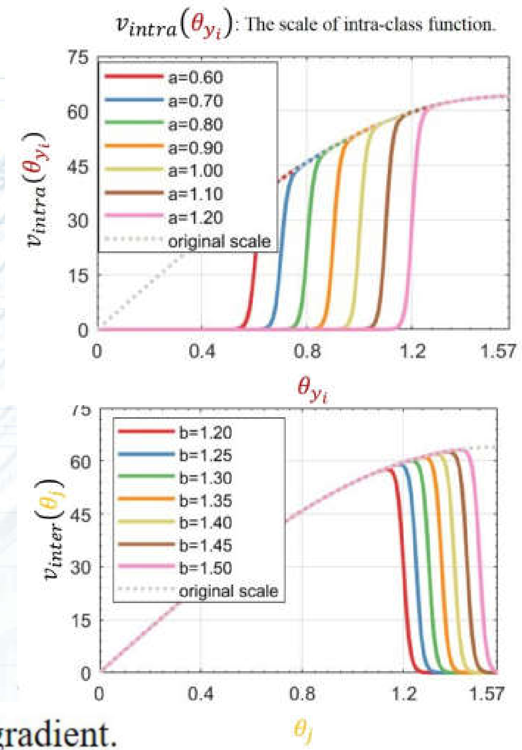
Yaobing Zhang, Weihong Deng, et al., Global-Local GCN: Large-Scale Label Noise Cleansing for Face Recognition, CVPR 2020

Accuracy: Noise-Aware Loss Function

This work proposes a novel loss function, named sigmoid constrained hypersphere loss (SFace), which imposes intra-class and inter-class constraints on a hypersphere manifold controlled by two sigmoid curves respectively.

- The optimizing directions are always along the tangent of the hypersphere while the moving speed is controlled precisely.
- The moving speed of $\boldsymbol{x_i}$ and $\boldsymbol{W_{y_i}}$ decreases gradually as they approaching to each other, while the moving speed of $\boldsymbol{x_i}$ and $\boldsymbol{W_j}$ increases rapidly as they start approaching to each other.



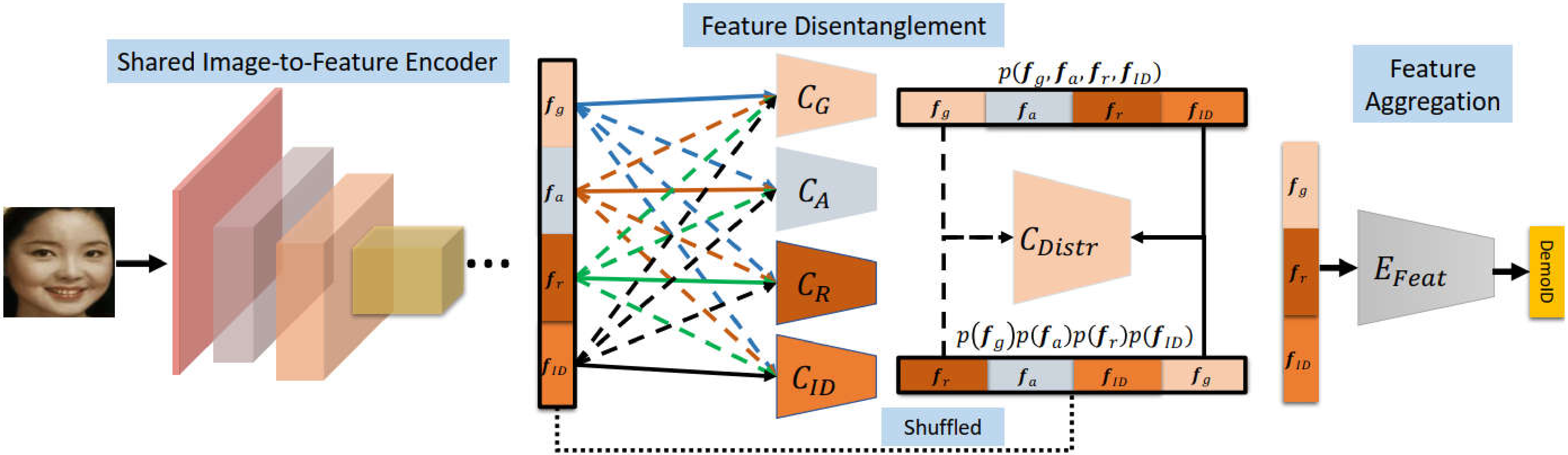


$v_{intra}(\theta_{yi})$: The scale of intra-class gradient. $v_{inter}(\theta_j)$: The scale of inter-class gradient.

Yaoyao Zhong, Weihong Deng, and et. al., "SFace: Sigmoid-constrained hypersphere loss for robust face recognition," IEEE Transactions on Image Processing, vol. 30, pp. 2587–2598, 2021.

Fairness: Feature Disentanglement

- Debface **adversarially disentangles** identity-related features from demographic information to mitigate bias.



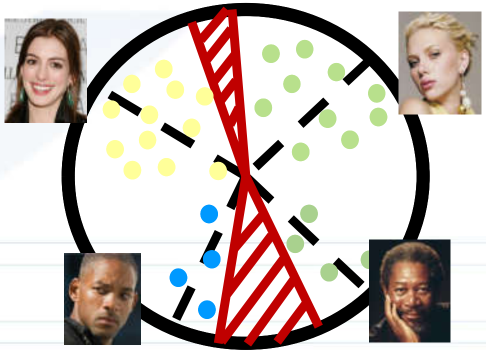
Trained on BUPT-Balanced and tested on RFW					
method	**Caucasian**	**Indian**	**Asian**	**African**	**Bias**
arcface	96.18	94.67	93.72	93.98	1.11
Debface	95.95	94.78	94.33	93.67	**0.83↓**

Sixue Gong, Xiaoming Liu, Anil K. Jain. Jointly de-biasing face recognition and demographic attribute estimation. ECCV 2020.

Fairness: Reinforcement margin learning

- RL-RBN adopts deep reinforcement learning to **adaptively learn margins** for different demographic groups.

① **Adaptive margin loss:**

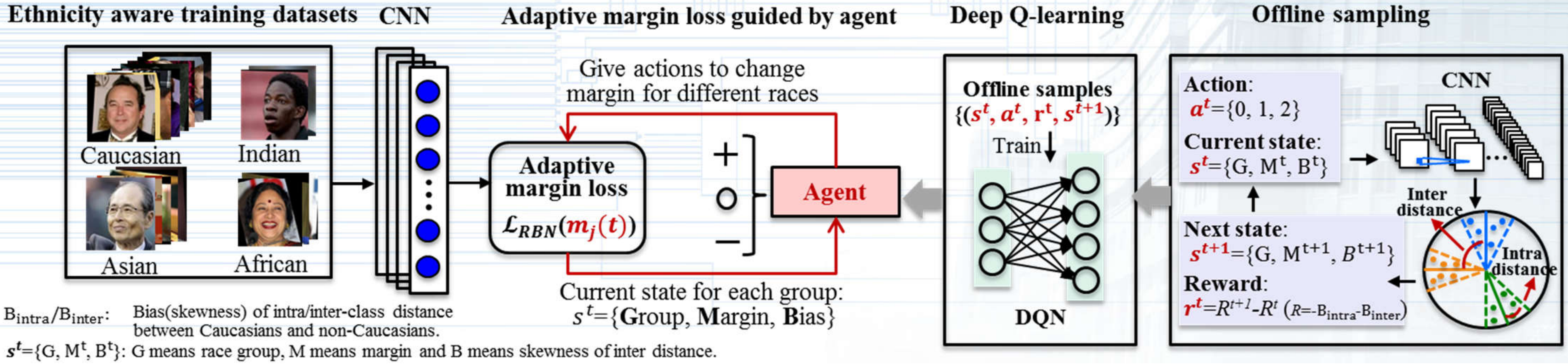


$$L_{RBN} = -\frac{1}{N}\sum_{i=1}^{N} log \frac{e^{s\cos\left(\theta_{y_i}+\boxed{\alpha_i(t)}\right)}}{e^{s\cos(\theta_{y_i}+\alpha_i(t))} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos(\theta_j)}}$$

Where, $\alpha_i(t) = \begin{cases} m, & if\ i \in Caucasian \\ m_i(t), & else \end{cases}$

Margins are learned adaptively for demographic groups by deep Q-learning.

② **Deep Q-learning for adaptive learning:**



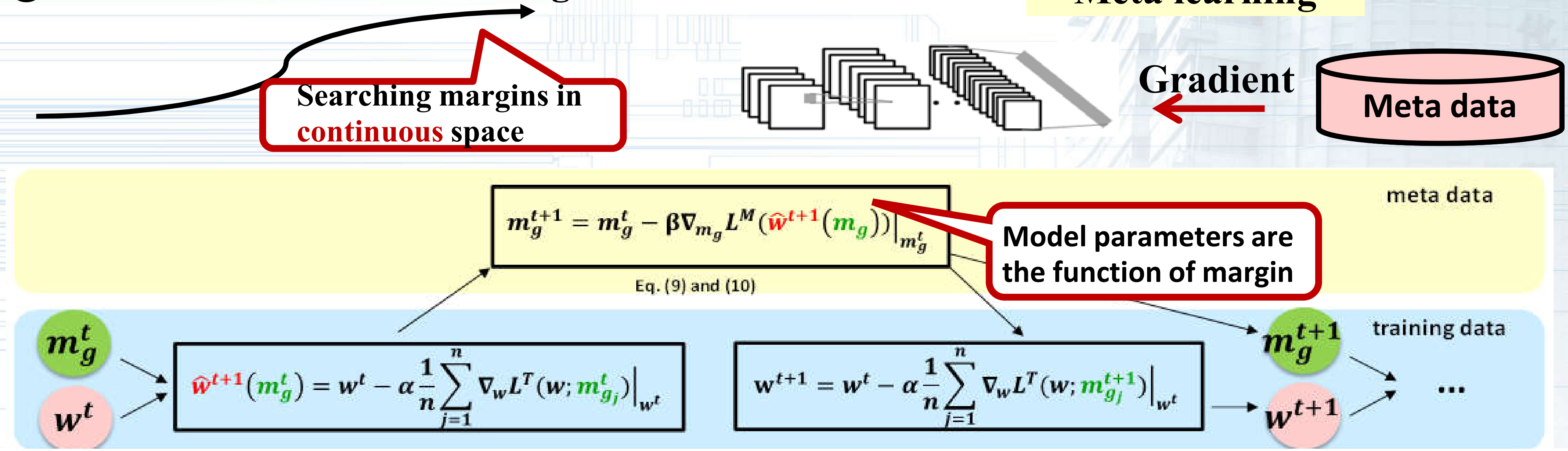
Mei Wang, Weihong Deng. "Mitigating bias in face recognition using skewness-aware reinforcement learning." CVPR 2020

Fairness: Meta Balanced Network

- Meta learning enables adaptive margin learning to search margin parameters **continuously** leading to fairer performance.

① RL-RBN based on Q-learning:

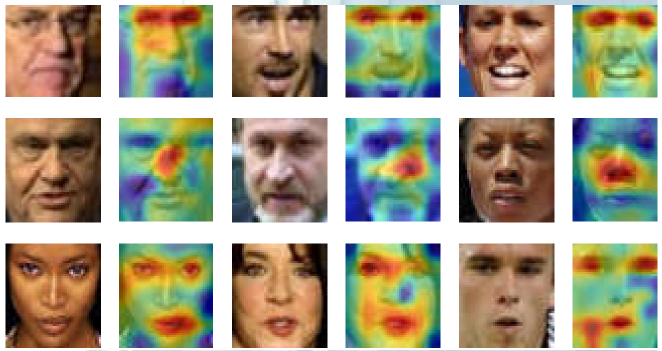
② MBN based on meta learning:



Mei Wang, Yaobin Zhang, Weihong Deng. Meta Balanced Network for Fair Face Recognition. TPAMI 2021.

Adaptiveness: Personalized Convolution

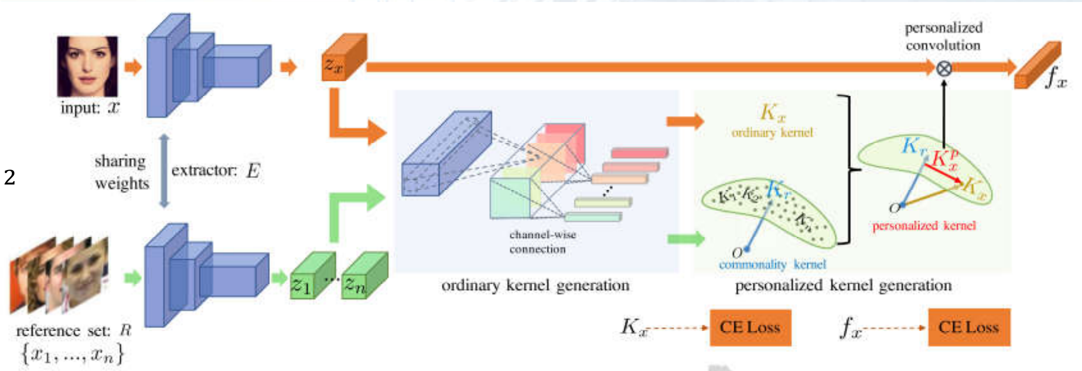
- Motivation
  - Vanilla CNN: Fixed kernel, same attention for all faces
  - Human: impressed by distinct characteristics of different faces



Personalized Kernel adaptive to each person can extract special distinguishing characteristics of each person for more accurate face recognition

- Method
  - Obtain personalized kernel by filtering out commonality with a reference set: Personalized kernel= Ordinary kernel- Commonality kernel

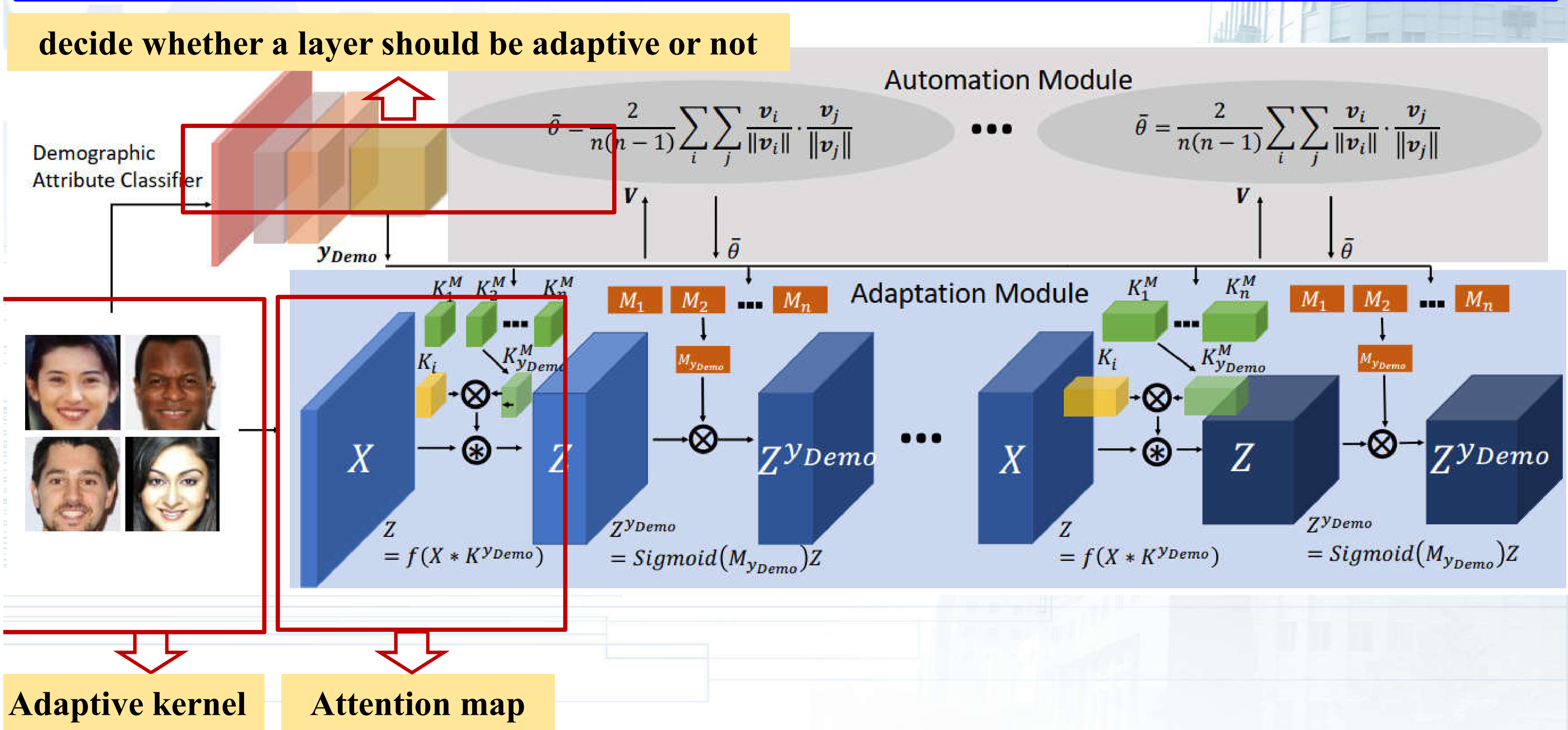
    - Ordinary kernel: $K_x = G(z_x)$
    - Commonality kernel:
      $$K_r = \underset{K_r}{\operatorname{argmin}} \frac{\alpha}{n}\|K^T K_r - \mathbb{1}\bar{K}^T K_r\|^2 + \frac{1-\alpha}{n}\|K^T K_r\|^2$$
    - Personalized kernel:
      $$K_x^p = K_x - K_r(K_x)^T K_r$$



Chunrui Han, Meina Kan, Shiguang Shan, Xilin Chen. Personalized Convolution for Face Recognition. International Journal of Computer Vision (IJCV), 2021 (Accepted).

Adaptiveness: Race-Aware Attention

- **Adaptive kernels and attention maps** are learned for different demographic groups to meet specific requirements.



Sixue Gong, Xiaoming Liu, Anil K. Jain.. Mitigating face recognition bias via group adaptive classifier. CVPR 2021.

Uncertainty: Probabilistic Face Embeddings

Model each face image as a Gaussian distribution:

$$p(\mathbf{z}|\mathbf{x}_i) = \mathcal{N}(\mathbf{z}; \boxed{\boldsymbol{\mu}_i}, \boxed{\boldsymbol{\sigma}_i^2}\mathbf{I})$$

fix learn

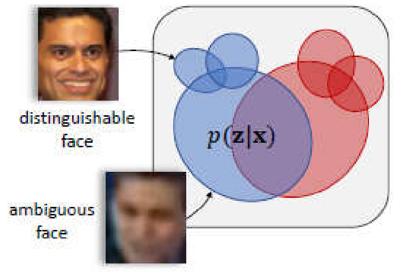


Figure1: Probabilistic embedding

The representation should have the following properties:

1. The center μ should encode the most likely facial features of the input image.
2. The uncertainty σ should encode the model's **confidence** along each feature dimension.

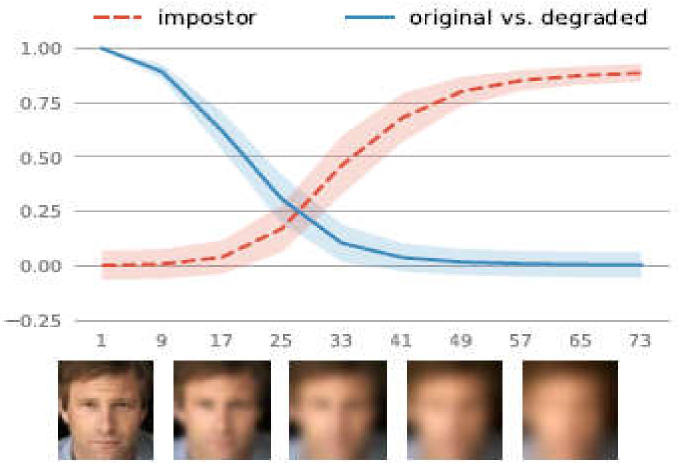


Figure2: Illustration of feature ambiguity dilemma

solved

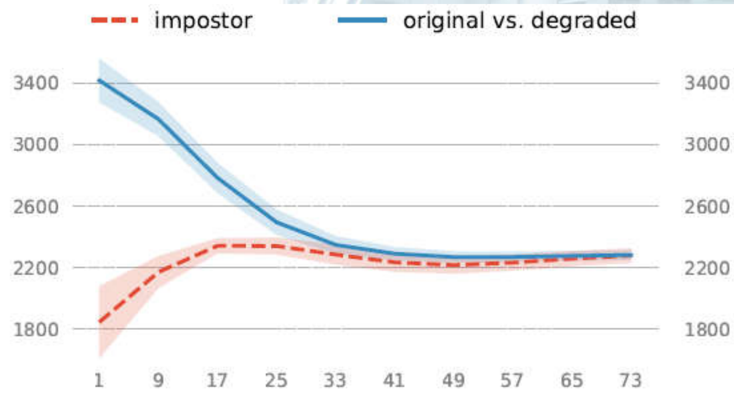


Figure3: Repeated experiments on feature ambiguity dilemma with PFE

Feature Ambiguity Dilemma:

- false accept of impostor low-quality pairs
- false reject of genuine cross-quality pairs

The scores of **cross-quality genuine pairs** converge to a point that is higher than the majority of **impostor scores.**

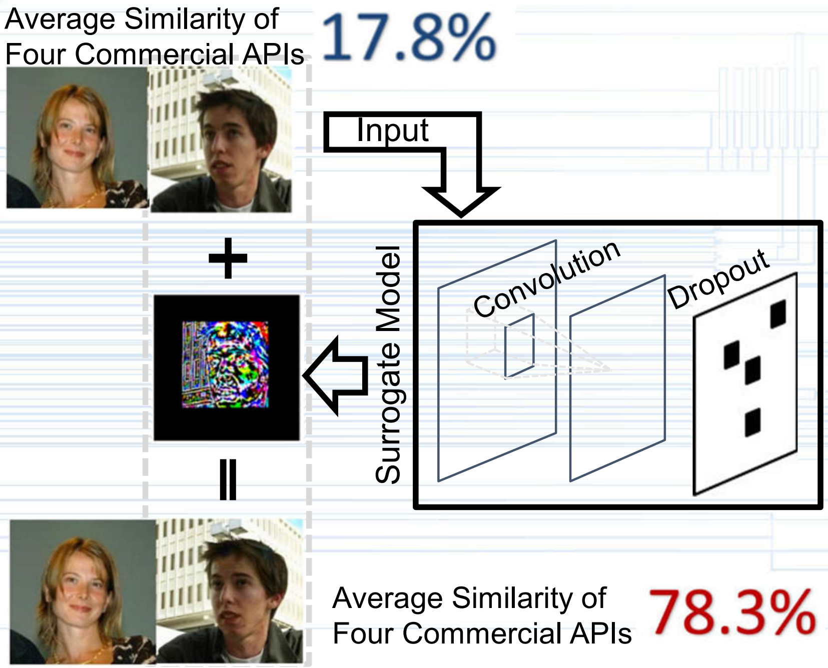
Shi Y, Jain A K. Probabilistic face embeddings[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6902-6911.

Uncertainty: Transferable Adversarial Attacks

Face recognition has achieved great success. However, the existence of transferable adversarial examples could severely hinder the robustness, since this type of attacks could be applied in a fully black-box manner without queries on the target system.

1. We investigate the transferable adversarial attacks and propose DFANet, which could increase the diversity of surrogate models and obtain ensemble-like effects.

2. Based on the proposed DFANet [1], we generate the adversarial images from the well-known LFW database with visually imperceptible noise, which provides a new database, TALFW, to serve as a benchmark to evaluate the robustness of deep face models.



	LFW Similarity	TALFW Similarity
Amazon	99%	57%
Microsoft	63%	25%
Baidu	92%	36%
Face++	84%	49%

	LFW Similarity	TALFW Similarity
Amazon	100%	28%
Microsoft	69%	28%
Baidu	96%	61%
Face++	90%	62%

	LFW Similarity	TALFW Similarity
Amazon	13%	84%
Microsoft	9%	43%
Baidu	51%	90%
Face++	43%	75%

	Model	LFW	TALFW
SOTA Algorithms	Center-loss [3]	98.78	70.65
	SphereFace [4]	99.27	62.47
	VGGFace2 [32]	99.43	71.47
	ArcFace (MobileNet) [7]	99.35	50.77
	ArcFace (ResNet-100) [7]	99.82	63.45
Commercial APIs	Amazon [25]	99.47	69.28
	Microsoft [26]	98.12	70.93
	Baidu [27]	97.72	72.07
	Face++ [28]	96.95	73.90
	Fusion of four APIs	99.65	72.33
Defensive Methods	No Defense	99.78	54.15
	JPEG Encoding [41]	99.55	73.93
	Gaussian Blur [41]	99.57	77.95
	Adversarial Training [38]	99.62	82.17

Yaoyao Zhong and Weihong Deng, Towards transferable adversarial attack against deep face recognition," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1452–1466, 2020.

Recent Work on Face Image Generation

Photo-realistic Face Image Generation

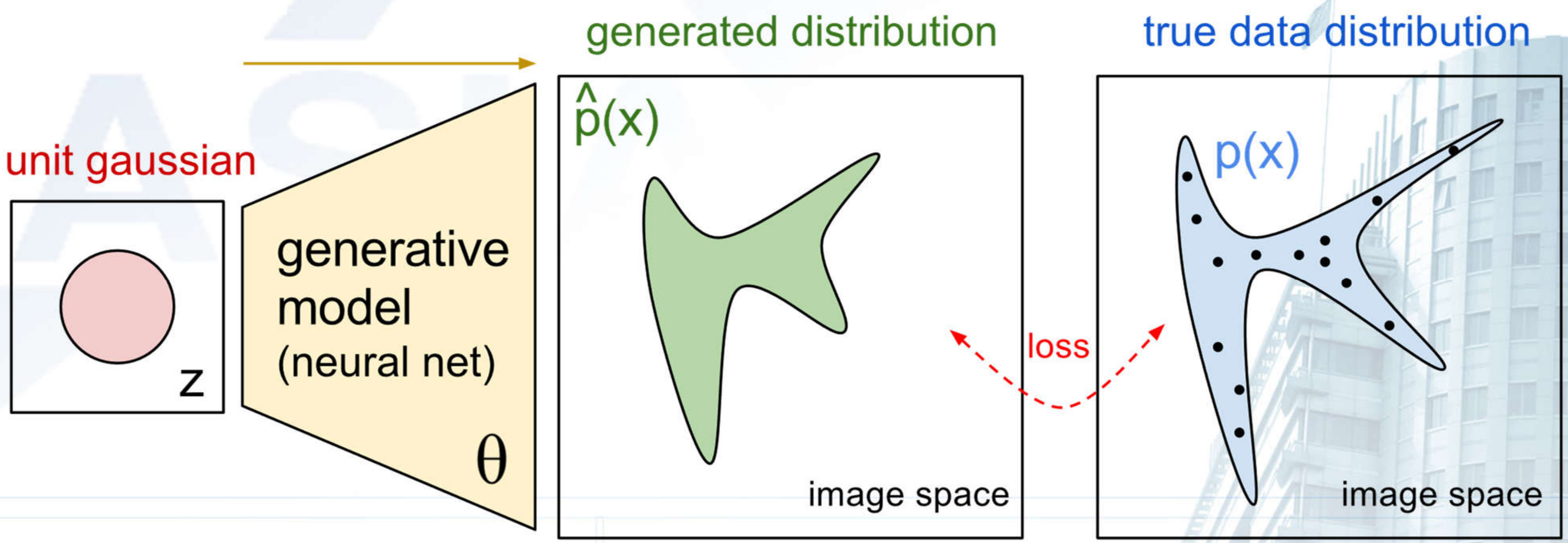
Unconditional Generation	• IntroVAE [PAMI 2021, NIPS 2018]
Rotation	• TP-GAN [ICCV 2017] • CAPG-GAN [CVPR2018] • HF-PIM [IJCV 2019, NIPS 2018]
Super-resolution	• Wavelet-SRNet [IJCV 2019, ICCV 2017]
Make-up	• BLAN [AAAI 2018]
Cross-spectral	• AD-HFR [AAAI 2018][PAMI 2020]
Completion	• FCENet [AAAI 2019][ACM MM2020]
Expression synthesis	• G2-GAN [ACM MM 2018] • CAFP-GAN [ACM MM 2018]
Aging	• Hierarchical Face Aging [ECCV 2020] • Attribute-aware Face Aging [CVPR 2019]

Recent Work on Face Image Generation

Photo-realistic Face Image Generation

Aging	• Age Progression and Regression [AAAI 2020] • Biphasic Facial Age Translation [TIFS 2022]
Facial Attribute Editing	• Controllable Multi-Attribute Editing [TIFS 2019] • Reference-guided Face Component Editing [IJCAI 2020]
Face Swapping	• MegaFS [CVPR 2021] • Information Bottleneck Disentanglement [CVPR 2021] • FaceInpainter [CVPR 2021] • AOT [NeurIPS 2020]
Face Synthesis	• Spatially Disentangled [TCSVT 2022] • Semantic-aware Noise Driven [TMM 2022]
Face Reenactment	• Semantic-aware One-shot Face Re-enactment [MIR 2022]
Text-to-Face	• SEA-T2F [ACM MM 2021] • AnyFace[CVPR 2022]

Generative Models: An Overview



An illustration of the main idea of generative models[1]

- **Motivation:** Learn a parameterized mapping function g_θ, such that

$$g_\theta(z) = \hat{p}(x) \rightarrow p(x)$$

where z is a latent variable sampled from a generic distribution.

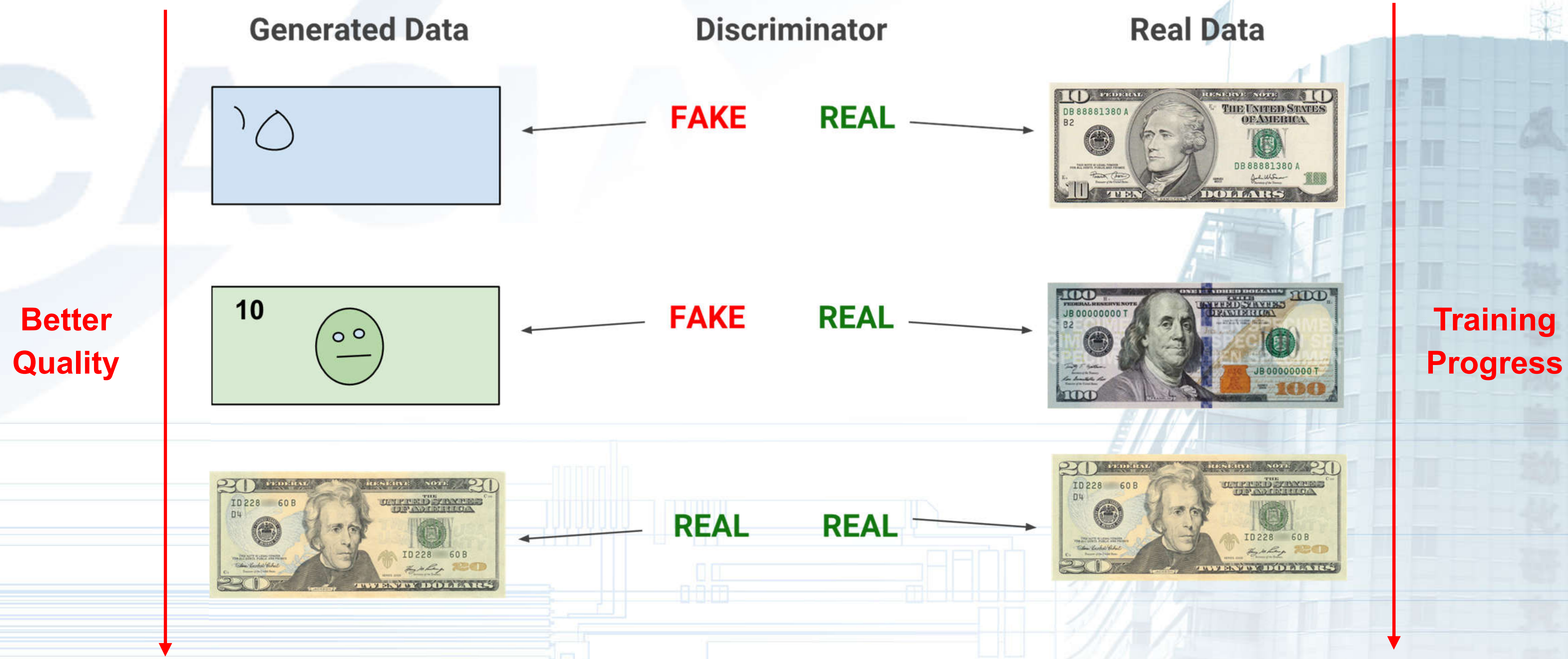
- **Problem:** How to measure the similarity between the distribution between generated samples $\hat{p}(x)$ and true data $p(x)$?

[1] https://openai.com/blog/generative-models/

Typical Generative Models

- Generative Adversarial Network (GAN)
- Variational Auto-encoder (VAE)
- Flow-based Model
- Diffusion Model

GAN: Main Idea

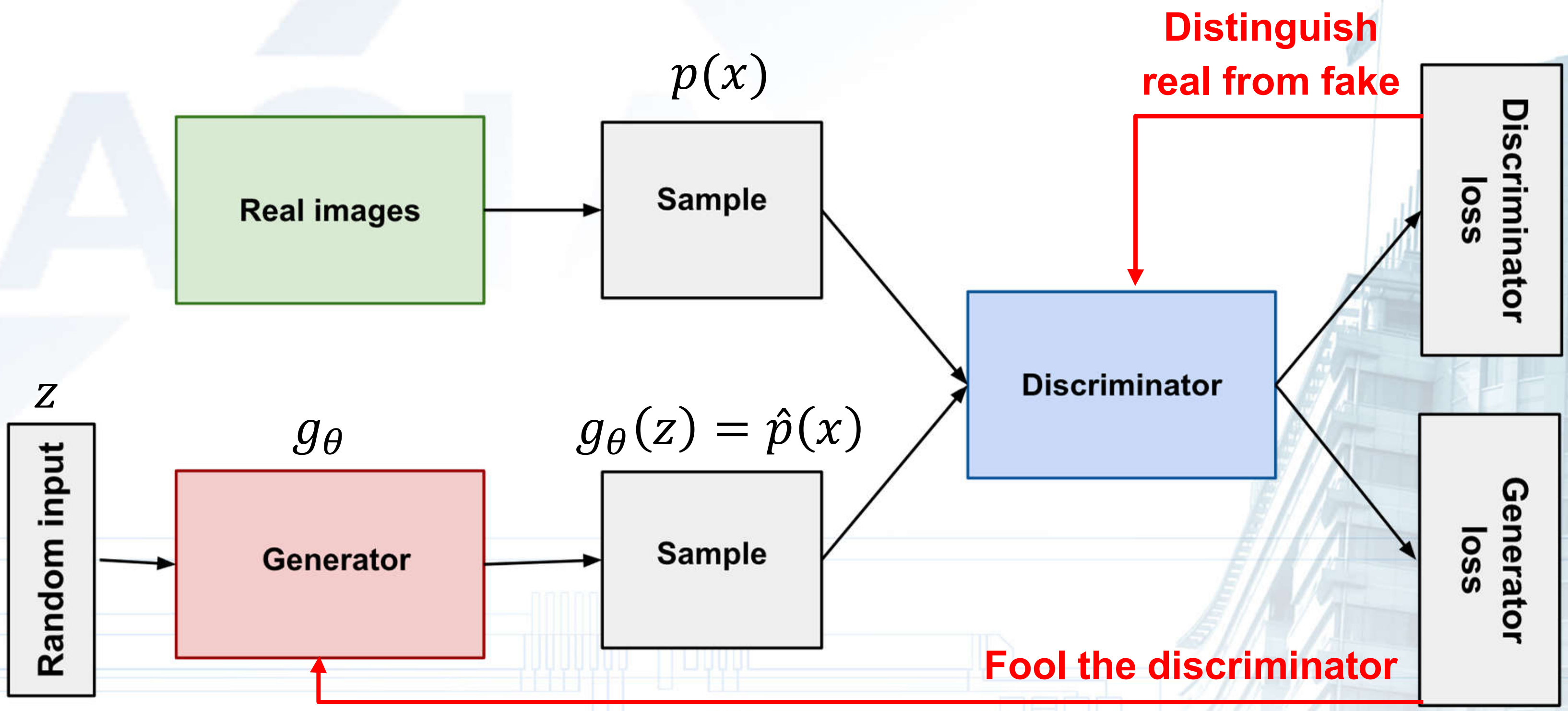


An illustration of the main idea of adversarial training[1]

- **Motivation:** Train an additional network (the discriminator) to distinguish generated sample from real ones.

[1] https://developers.google.com/machine-learning/gan/gan_structure

GAN: Adversarial Training



A diagram of the GAN model[1]

- **Adversarial Training:** the discriminator aims to distinguish $g_\theta(z) = \hat{p}(x)$ from $p(x)$, while the generator attempts to fool the discriminator

[1] https://developers.google.com/machine-learning/gan/gan_structure

The Success of GAN

Generative adversarial networks (GANs) have been successfully applied in image/video/music/art generation, computer vision and pattern recognition.

Dueling Neural Networks

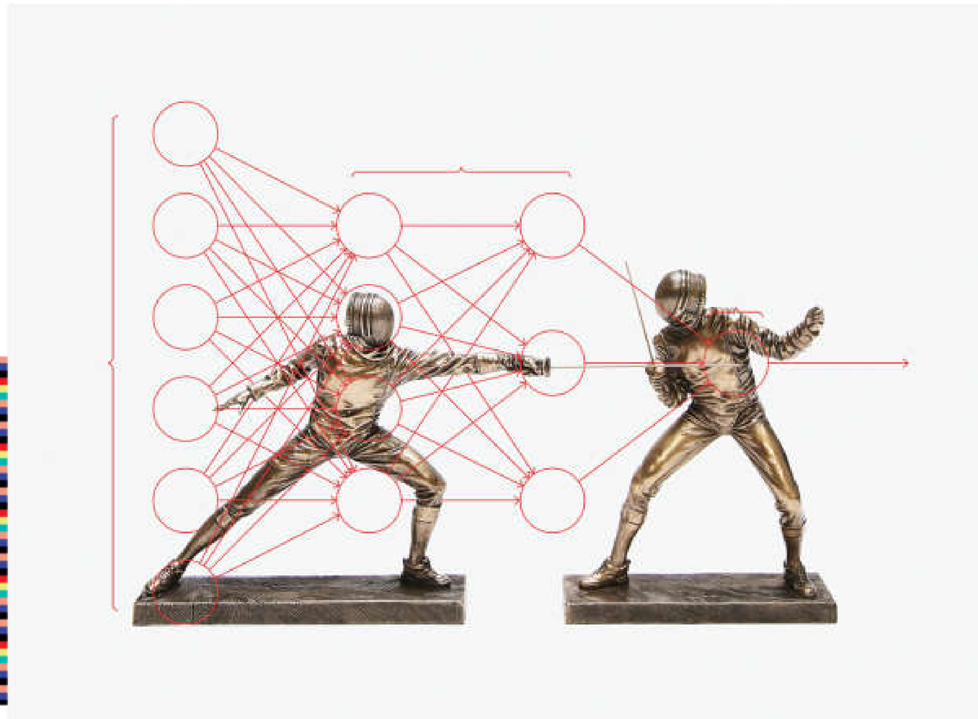


ILLUSTRATION BY DEREK BRAHNEY | DIAGRAM COURTESY OF MICHAEL NIELSEN, "NEURAL NETWORKS AND DEEP LEARNING", DETERMINATION PRESS, 2015

Dueling Neural Networks

Breakthrough

Two AI systems can spar with each other to create ultra-realistic original images or sounds, something machines have never been able to do before.

Why It Matters

This gives machines something akin to a sense of imagination, which may help them become less reliant on humans—but also turns them into alarmingly powerful tools for digital fakery.

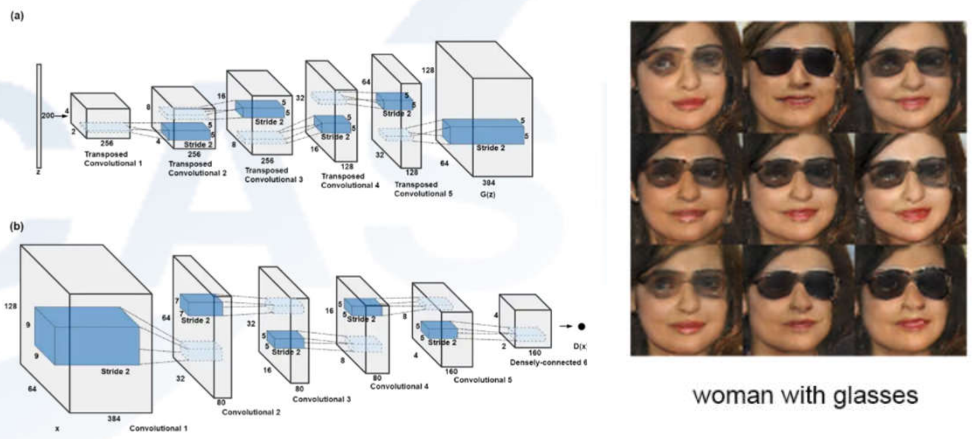
Key Players

Google Brain, DeepMind, Nvidia

Availability

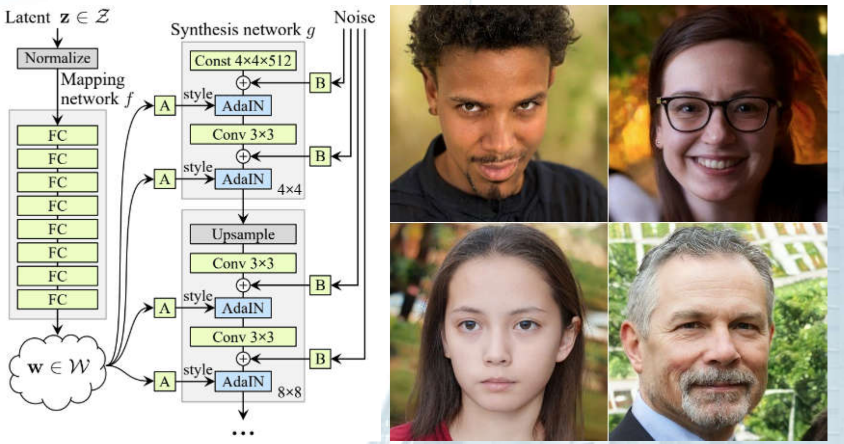
Now

GAN: Progress and Achievements



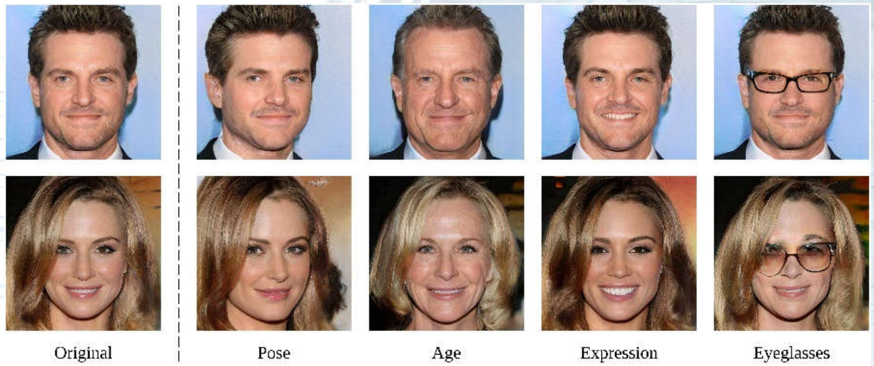
woman with glasses

DCGAN, 2015[1]



StyleGAN, 2019[2]

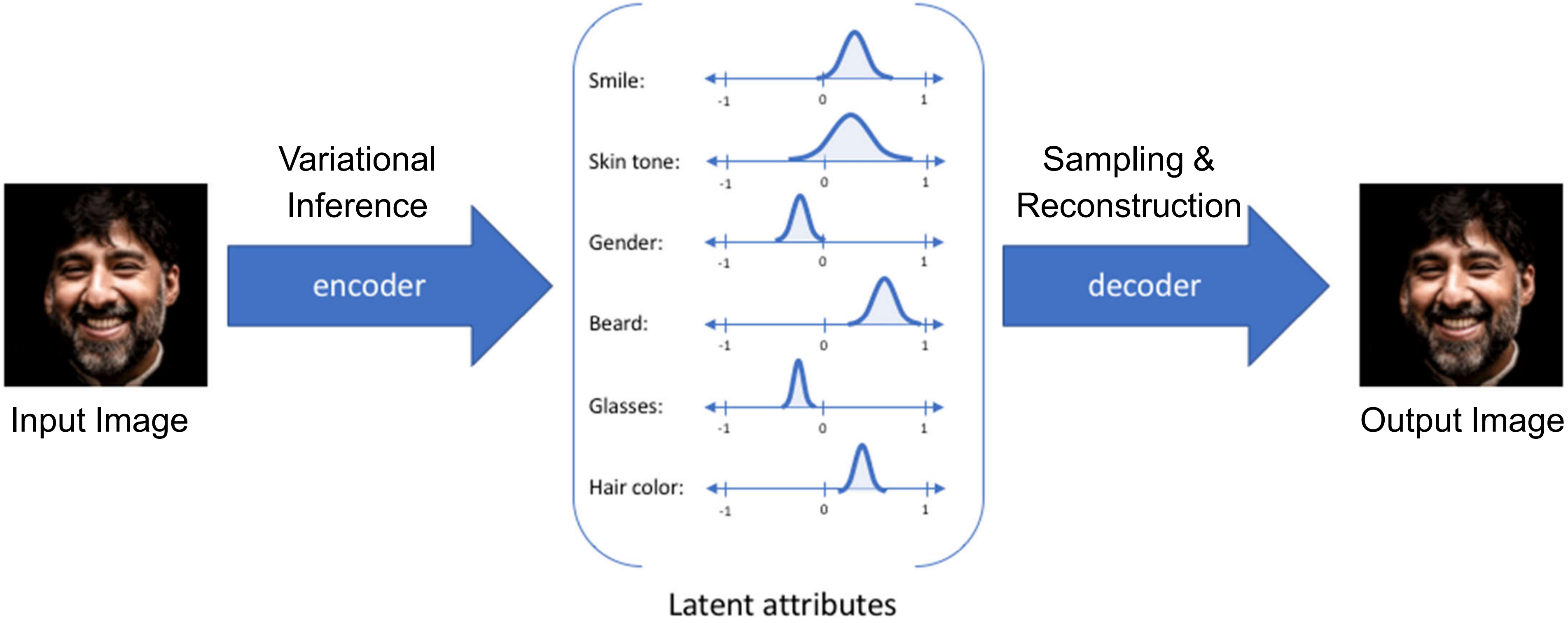
CycleGAN, 2017[3]



InterFaceGAN, 2022[4]

[1] Radford, Alec, Luke Metz, and Soumith Chintala. “Unsupervised representation learning with deep convolutional generative adversarial networks.” *arXiv preprint arXiv:1511.06434.* 2015.
[2] Karras, Tero, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." In *CVPR*, pp. 4401-4410. 2019.
[3] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In *ICCV*, pp. 2223-2232. 2017.
[4] Y. Shen, C. Yang, X. Tang and B. Zhou, "InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs," in TPAMI, vol. 44, no. 4, pp. 2004-2018, 2022.

VAE: Main Idea

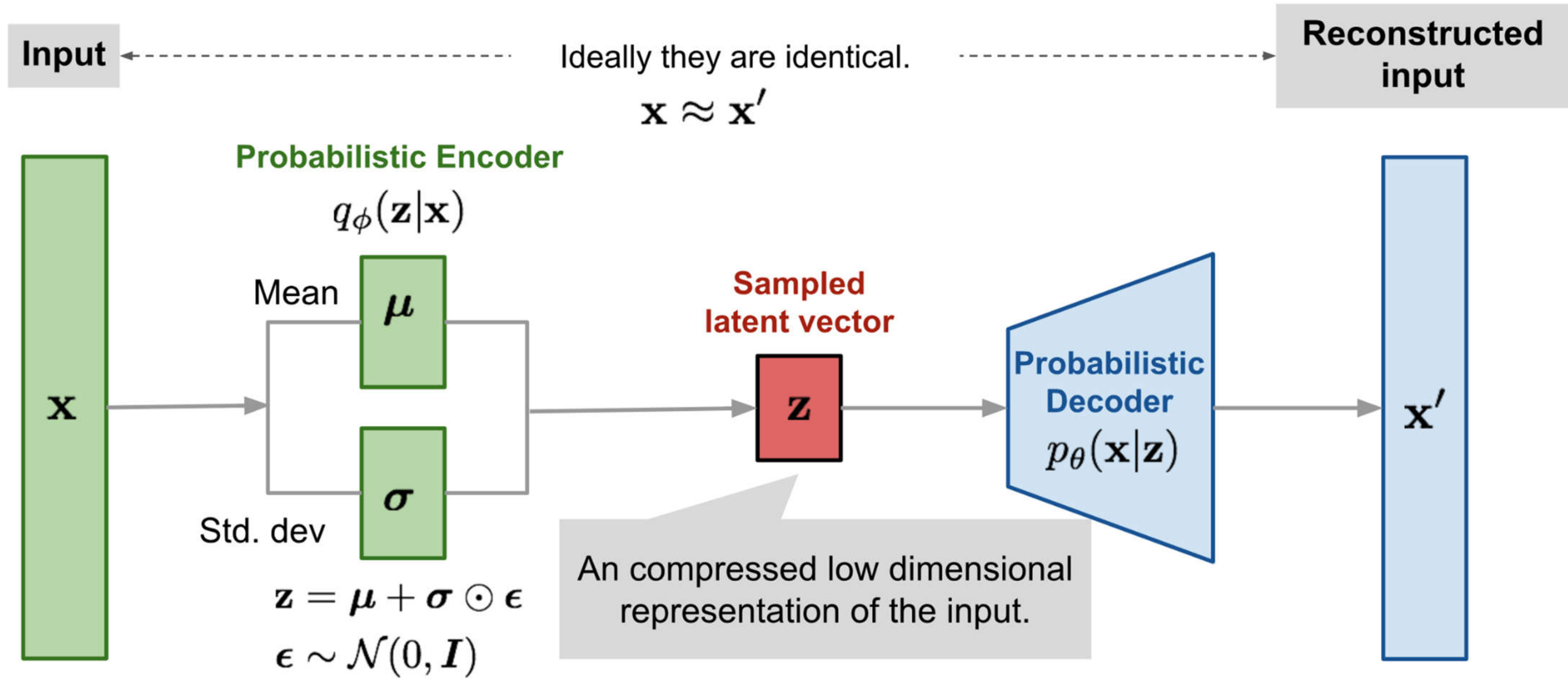


A illustration of the main idea of VAE[1]

- **Motivation:** Solve for variational latent components for point-wisely reconstructing the input data from a probabilistic perspective.

[1] https://www.jeremyjordan.me/variational-autoencoders/

VAE: Main Idea

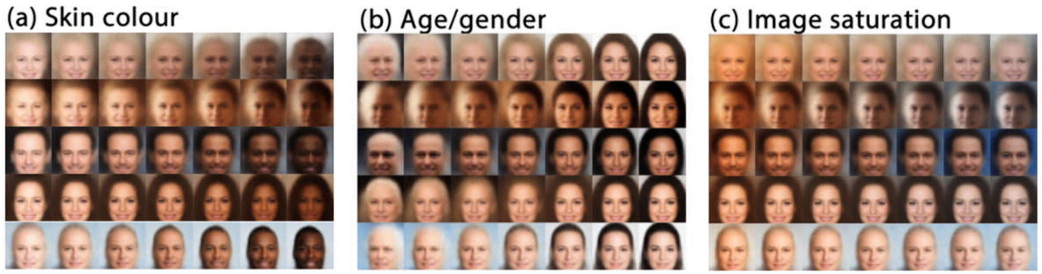


An overview of the framework and training object of VAE[1]

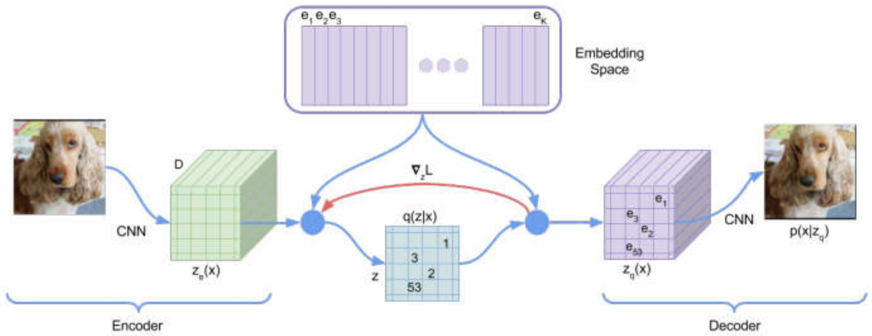
- **Implementation:** Predict the mean and var. from input, sample the latent code, obtain the output image, and optimize w.r.t the loss

[1] https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

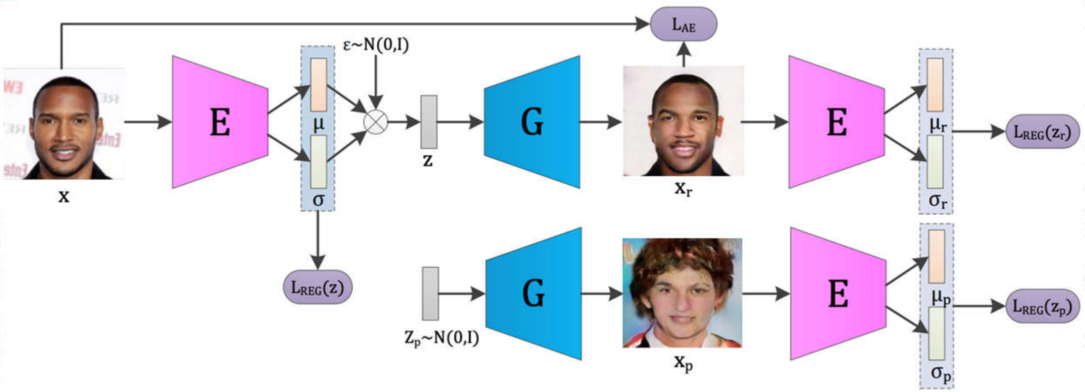
VAE: Progress and Achievements



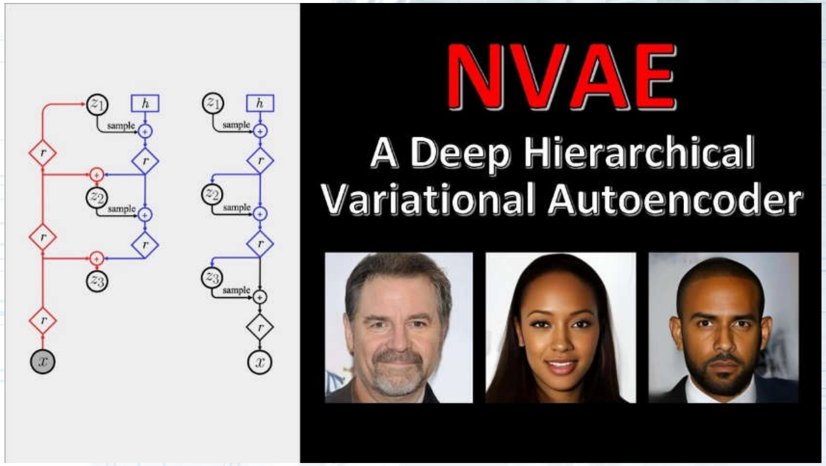
Beta-VAE, 2016[1]



VQVAE, 2017[2]



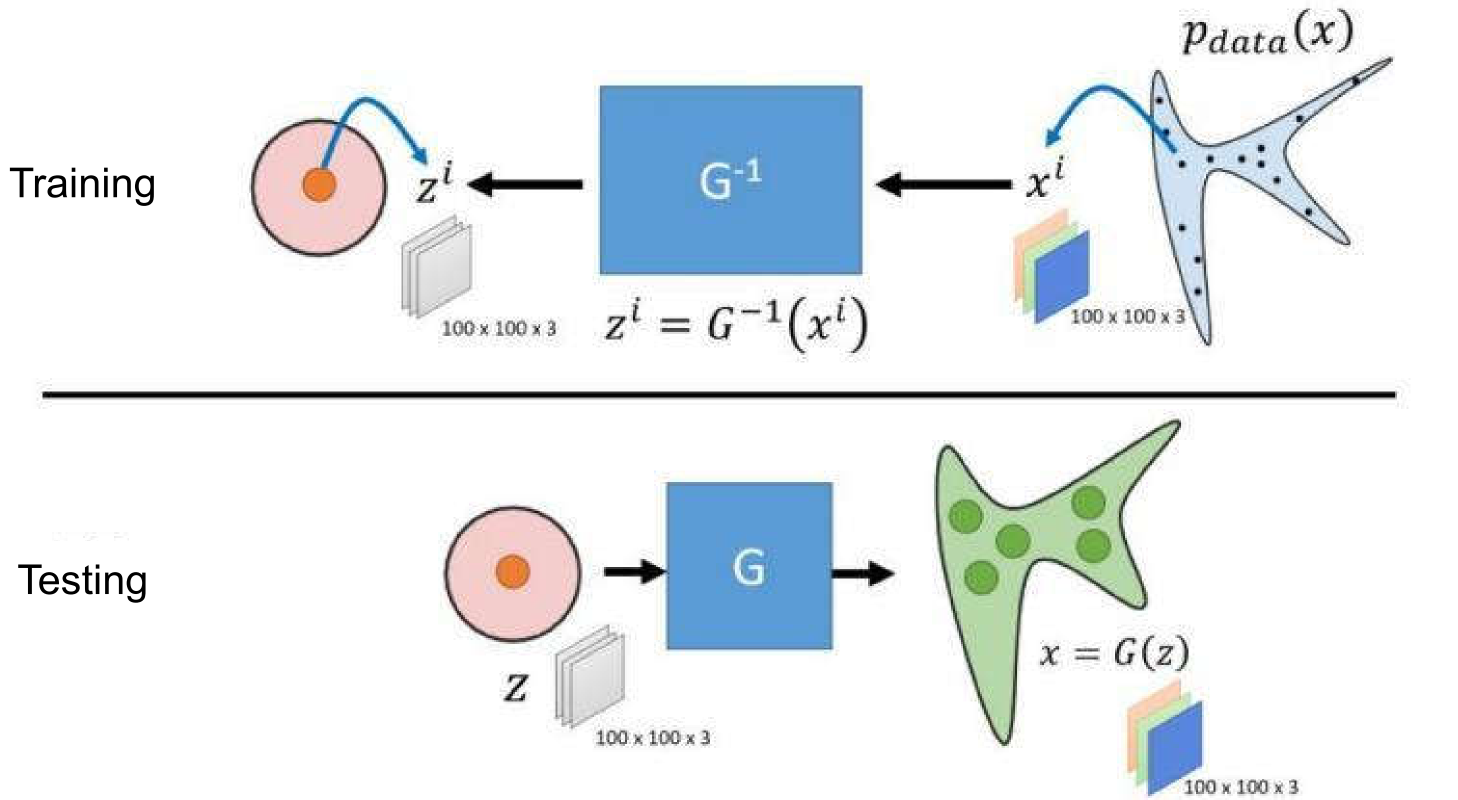
IntroVAE, 2018[3]



NVAE, 2020[4]

[1] Higgins, Irina, et al. “beta-vae: Learning basic visual concepts with a constrained variational framework.”. 2016.
[2] Van Den Oord, Aaron, and Oriol Vinyals. “Neural discrete representation learning.” *NeurIPS*. 2017.
[3] Huang, Huaibo, Ran He, Zhenan Sun, and Tieniu Tan. “Introvae: Introspective variational autoencoders for photographic image synthesis.” *NeurIPS. 2018.*
[4] Vahdat, Arash, and Jan Kautz. “NVAE: A deep hierarchical variational autoencoder.” *NeurIPS*. 2020.

Flow-based Model: Main Idea

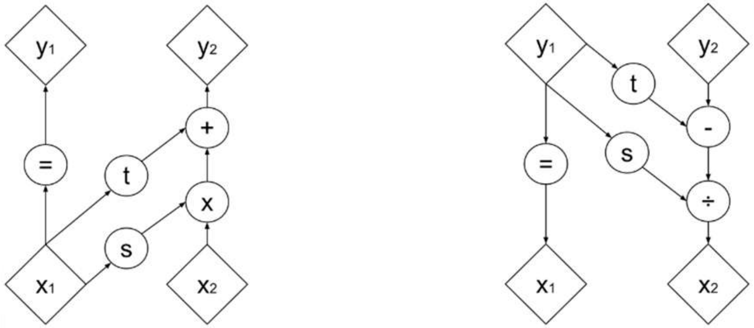


A illustration of the training and testing process for flow-based methods[1]

- **Motivation:** Explicitly learns the probability density function of real data with normalizing flows, a powerful statistics tool for density estimation.

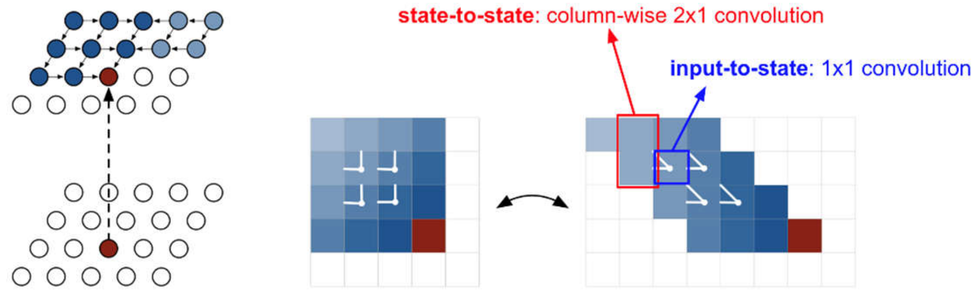
[1] https://zhuanlan.zhihu.com/p/267305869

Representative Flow-based Models



(a) Forward propagation (b) Inverse propagation

RealNVP, 2016[1]



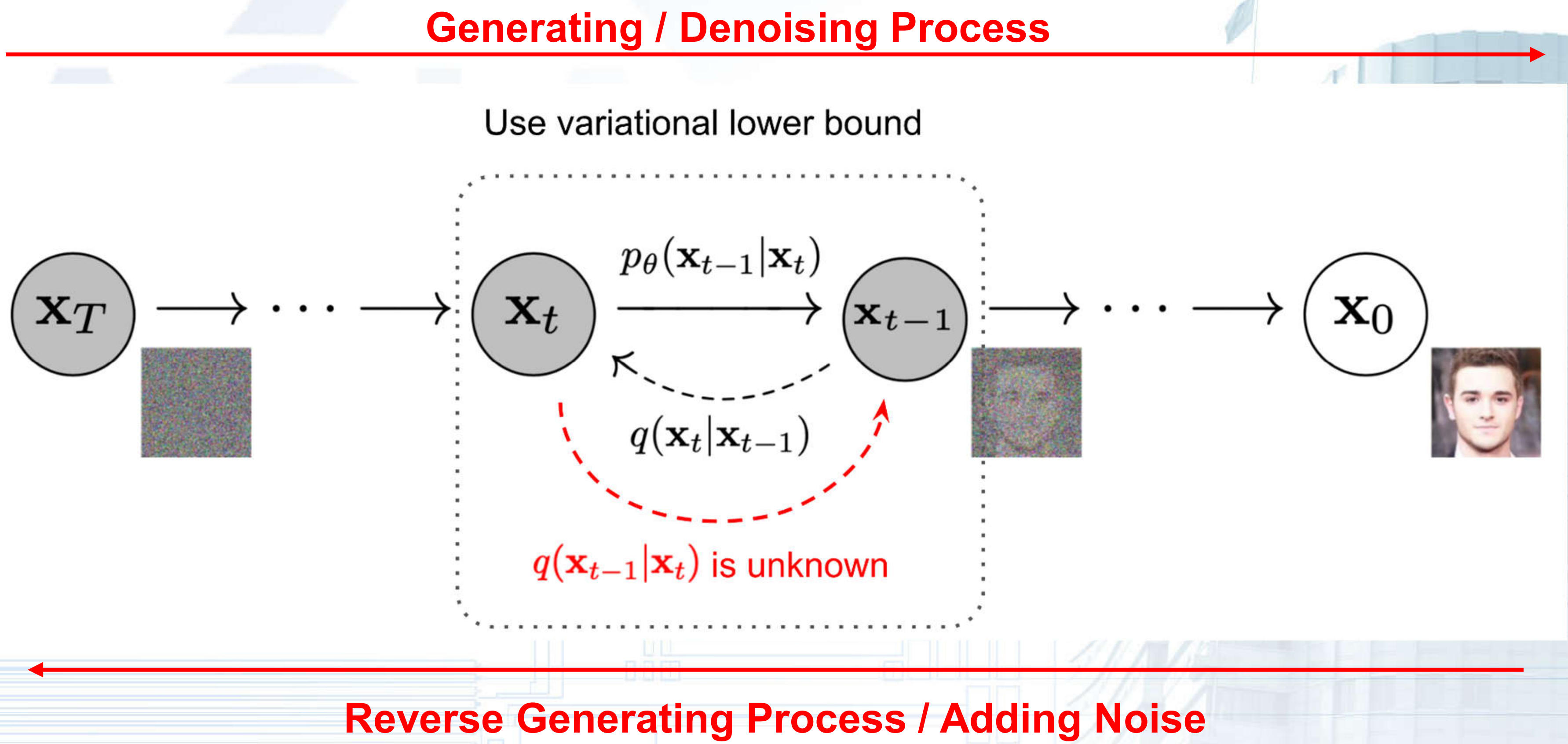
(a) Diagonal BiLSTM (b) Skewing operation

PixelRNN, 2016[2]

GLOW, 2018[3]

[1] Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. “Density estimation using real nvp.” *arXiv preprint arXiv:1605.08803*. 2016.
[2] Van Den Oord, Aäron, Nal Kalchbrenner, and Koray Kavukcuoglu. “Pixel recurrent neural networks.” In *ICML. 2016.*
[3] Kingma, Durk P., and Prafulla Dhariwal. “Glow: Generative flow with invertible 1x1 convolutions.” In *NeurIPS. 2018.*

Diffusion Model: Main Idea



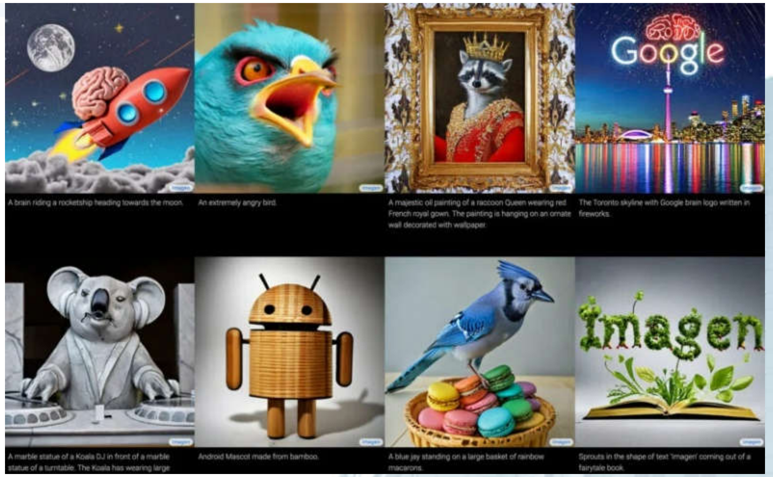
An overview of the framework of diffusion models[1]

- **Motivation:** Add random noise to data and then learn to reverse the diffusion process to construct desired data samples from the noise.

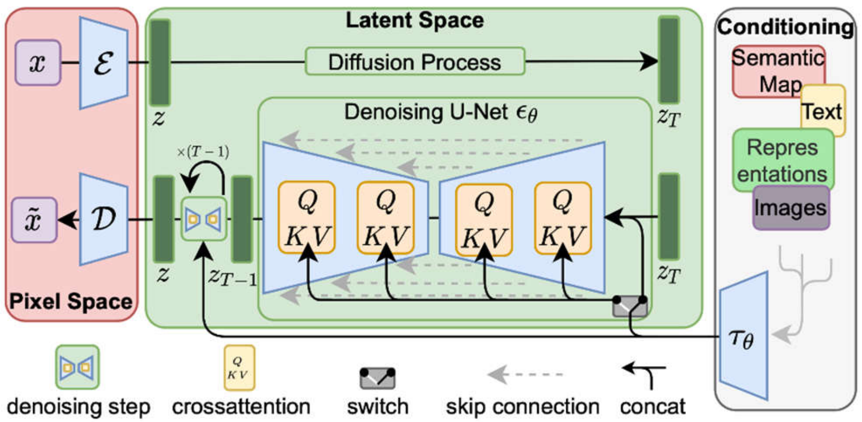
[1] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." NeurIPS. 2020.

Representative Diffusion Models

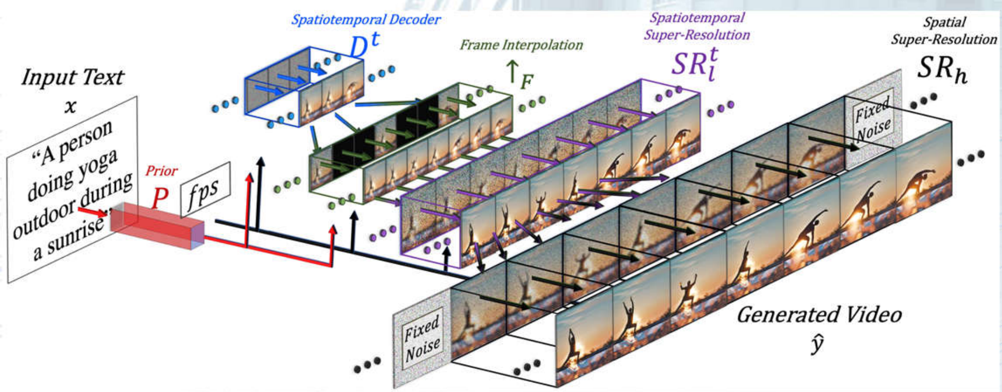
DALLE2, 2022[1]



Imagen, 2022[2]



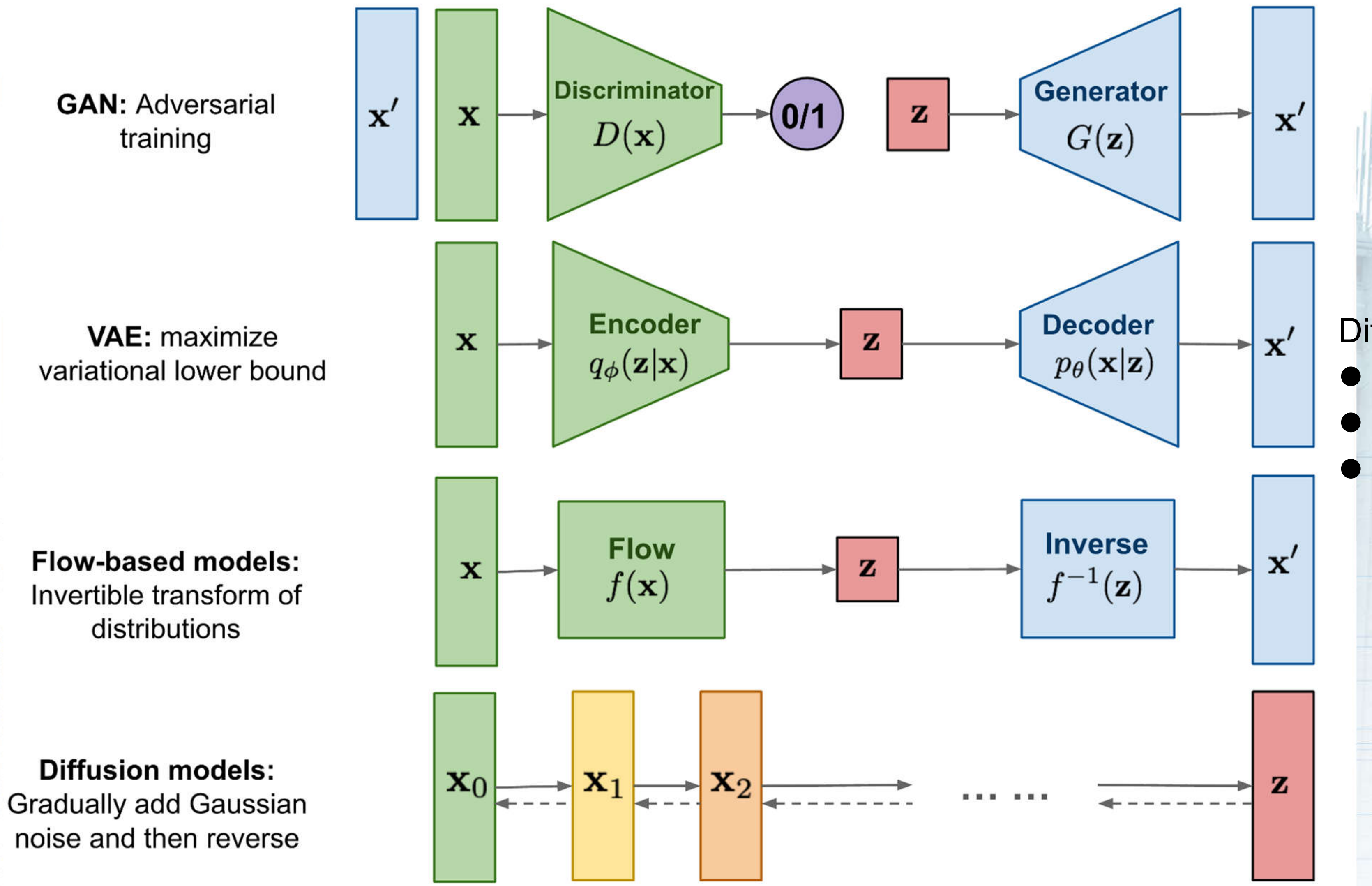
Latent Diffusion Model, 2022[3]



Make-a-video, 2020[4]

[1] Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. "Hierarchical text-conditional image generation with clip latents." *arXiv preprint arXiv:2204.06125*. 2022.
[2] Saharia, Chitwan, et al. "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding." arXiv preprint arXiv:2205.11487. 2022.
[3] Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models." In CVPR. 2022.
[4] Singer, Uriel et al. "Make-a-video: Text-to-video generation without text-video data." arXiv preprint arXiv:2209.14792. 2022.

Comparison of Generative Models



Comparison of the framework of different generative models[1]

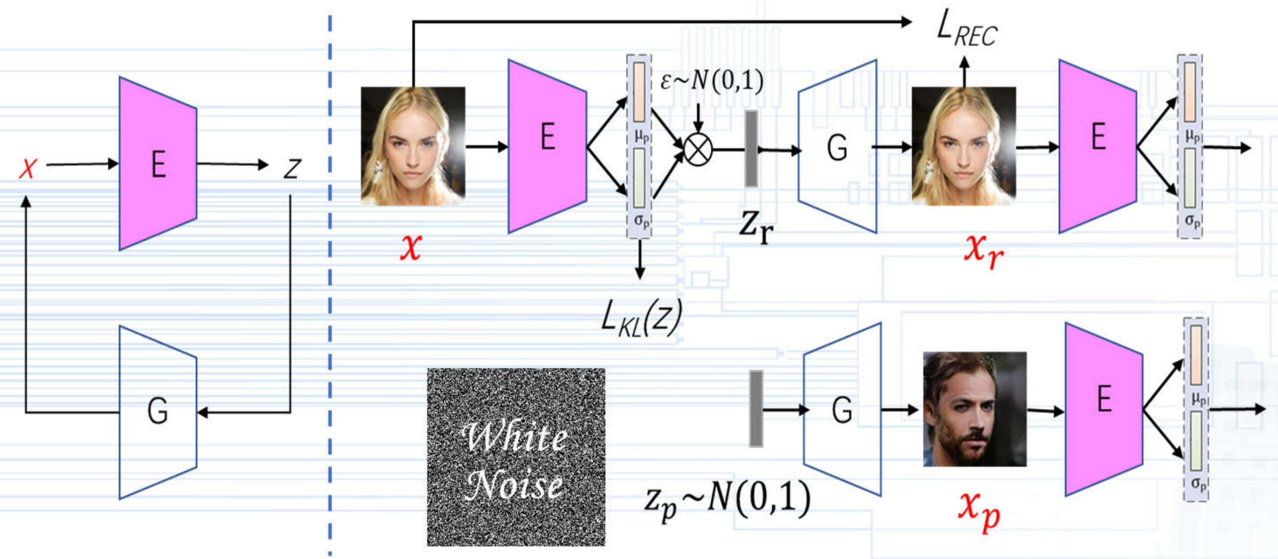
Different in:
- Motivation
- Architecture
- Training Objectives

[1] https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Unconditional Generation

Introspective VAE (unconditional)

- **Generate virtual faces from white noise**
- **Adversarial distribution matching：** use the KL-regularization term as the adversarial training cost function
- **Introspective variational inference：** combine the adversarial object with the ELBO object of VAEs

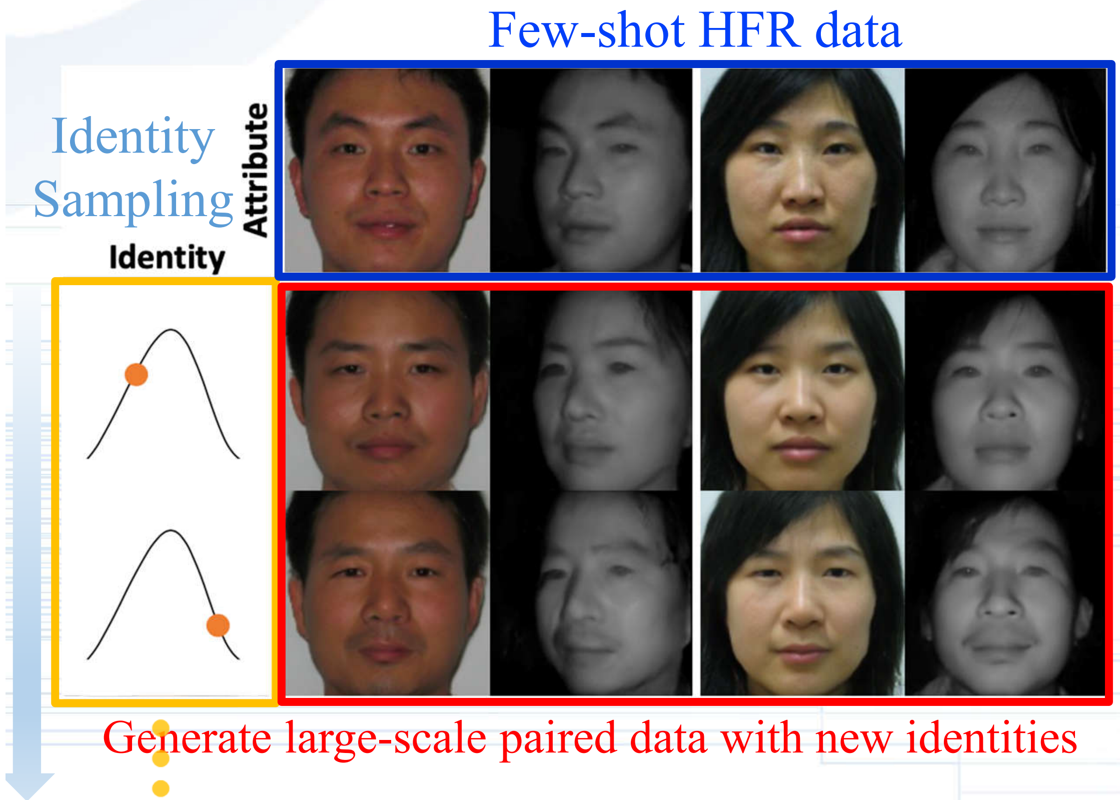


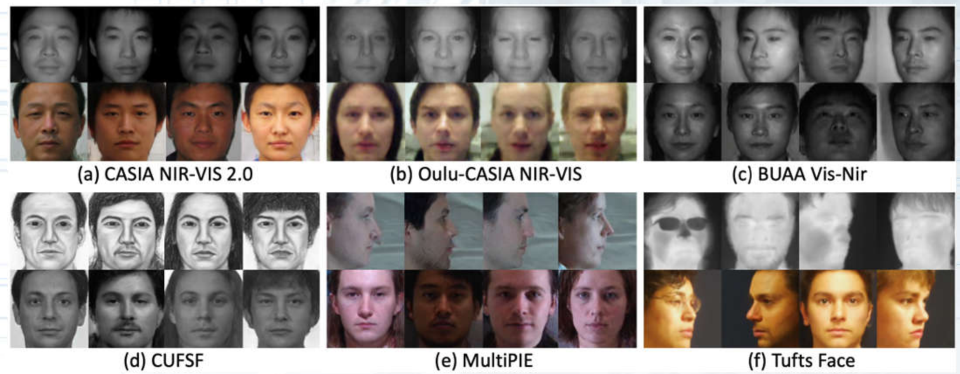
Huaibo Huang, Zhihang Li, Ran He, Zhenan Sun, Tieniu Tan. IntroVAE: Introspective Variational Autoencoders for Photographic Image Synthesis. NeurIPS 2018: 52-63.

Unconditional Generation

Dual Variational Generation (unconditional)

- Generate paired images of one identity from noise
- Data Augmentation: Integrate virtual identities into few-shot HFR data





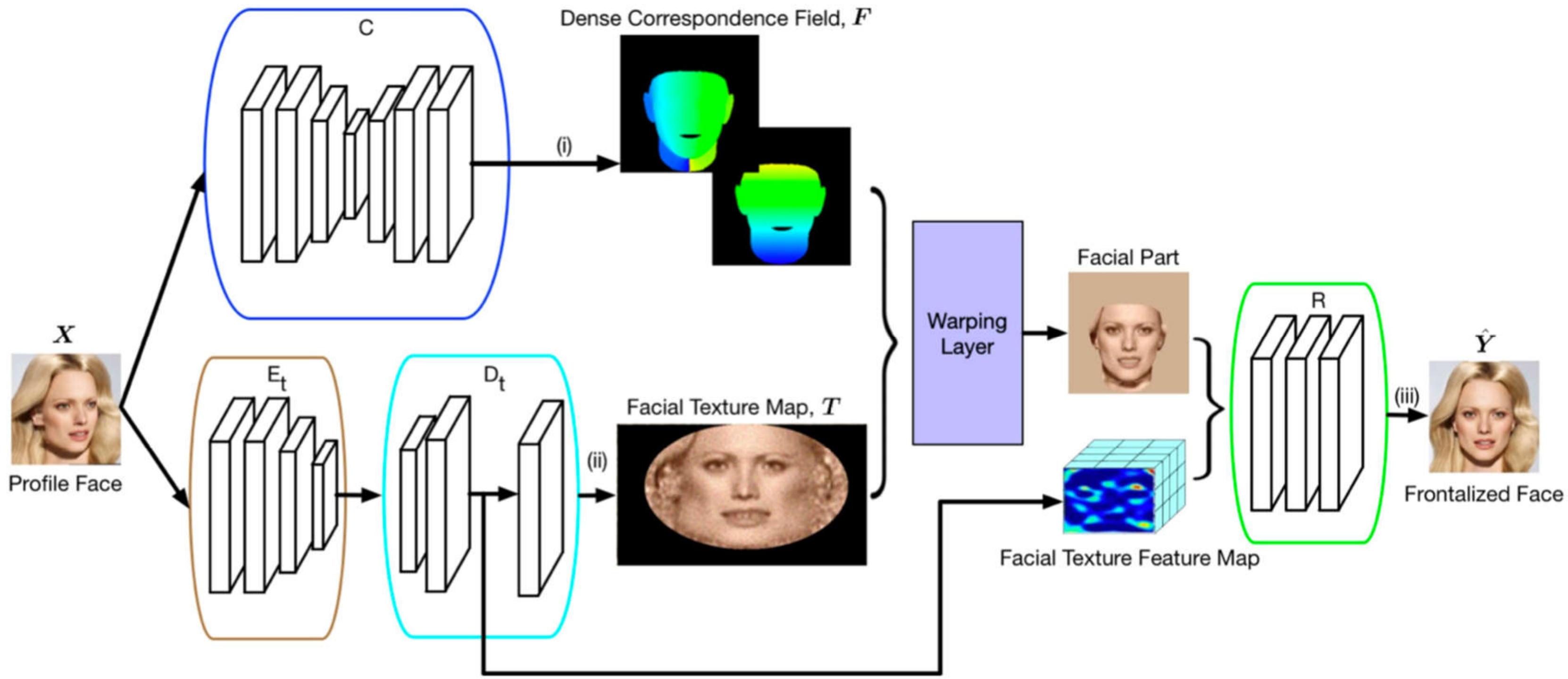
Achieve the best results on seven datasets

[1] Chaoyou Fu, et al. DVG-Face: Dual Variational Generation for Heterogeneous Face Recognition. IEEE TPAMI, 2021
[2] Chaoyou Fu, et al. Dual Variational Generation for Low Shot Heterogeneous Face Recognition. NeurIPS, 2019
[3] Peipei Li, et al. Dual-structure Disentangling Variational Generation for Data-limited Face Parsing. ACM MM 2020

Rotation

Towards High Fidelity Face Frontalization in the Wild

- High Fidelity Pose Invariant Model (HF-PIM) is proposed to produce realistic and identity-preserving frontalized face images with the highest resolution (256*256) in the literature.



Jie Cao, Yibo Hu, Hongwen Zhang, Ran He, Zhenan Sun. Towards High Fidelity Face Frontalization in the Wild, IJCV, 2019.

Rotation

中国科学院

Towards High Fidelity Face Frontalization in the Wild

Table 4: Face recognition/verification performance (%) comparisons on IJB-A. The results are averaged over 10 testing splits. "-" means the result is not reported.

Method	Verification		Recognition	
	FAR=0.01	FAR=0.001	Rank-1	Rank-5
DR-GAN [53]	77.4±2.7	53.9±4.3	85.5±1.5	94.7±1.1
FF-GAN [60]	85.2±1.0	66.3±3.3	90.2±0.6	95.4±0.5
PIM [61]	93.3±1.1	87.5±1.8	94.4±1.1	-
Light CNN [56]	91.5±1.0	84.3±2.4	93.0±1.0	-
HF-PIM(Ours)	**95.3±0.7**	**89.9±1.3**	**96.4±0.5**	**98.1±0.2**

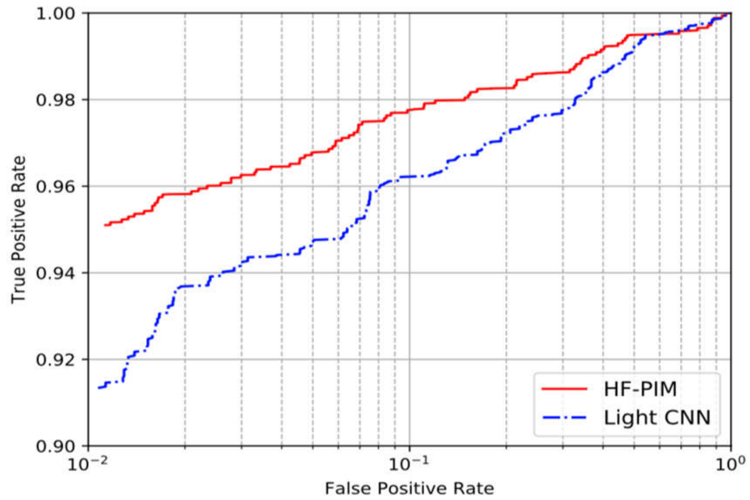
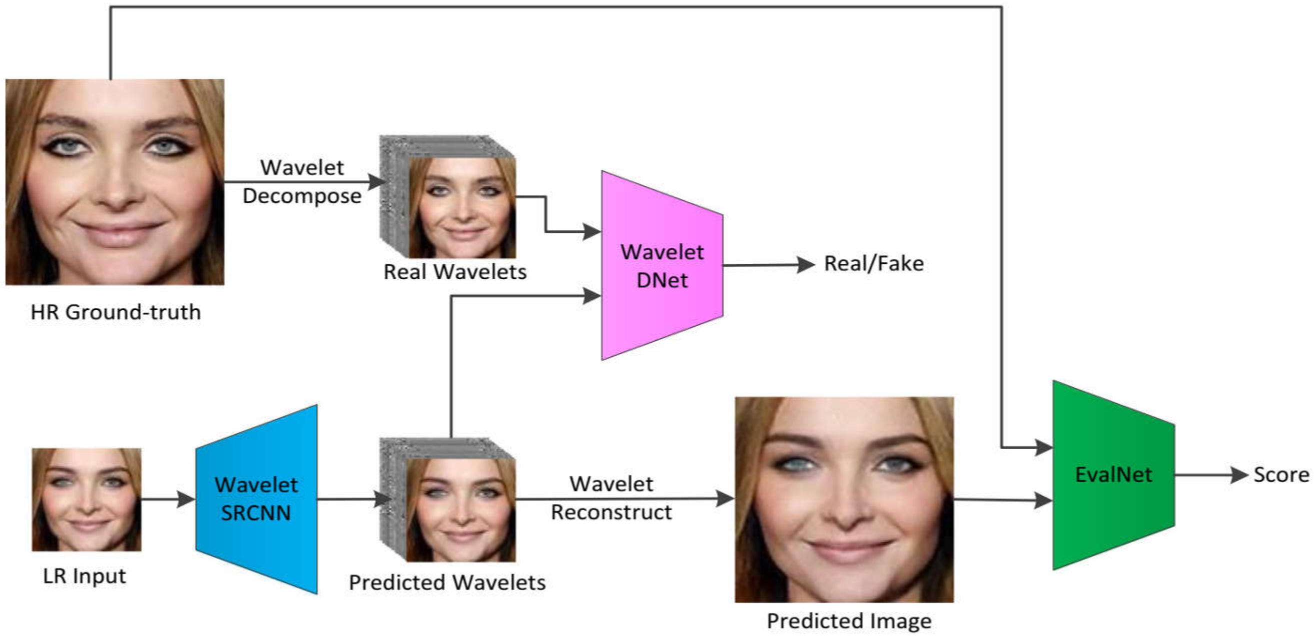


Fig. 5: ROC curves on the IJB-A verification protocol.

Face Super Resolution

Wavelet domain CNN and GAN

- Wavelet domain CNN [1] and GAN [2] solutions to face super resolution
- Special design of loss functions to capture both global topology information and local textual details



[1] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan, Wavelet-SRNet: A Wavelet-based CNN for Multi-scale Face Super Resolution, ICCV, 2017.

[2] Huaibo Huang, Ran He, Zhenan Sun, Tieniu Tan, Wavelet Domain Generative Adversarial Network for Multi-scale Face Hallucination，International Journal of Computer Vision，Volume 127, Issue 6–7, pp.763–784, 2019.

(a) LR **(b)** GT **(c)** Bicubic **(d)** WTIP **(e)** SRCNN **(f)** SRGAN **(g)** URDGN **(h)** SRDense **(i)** CBN **(j)** Our-CNN **(k)** Ours

Table 3 Face verification results on the LFW dataset

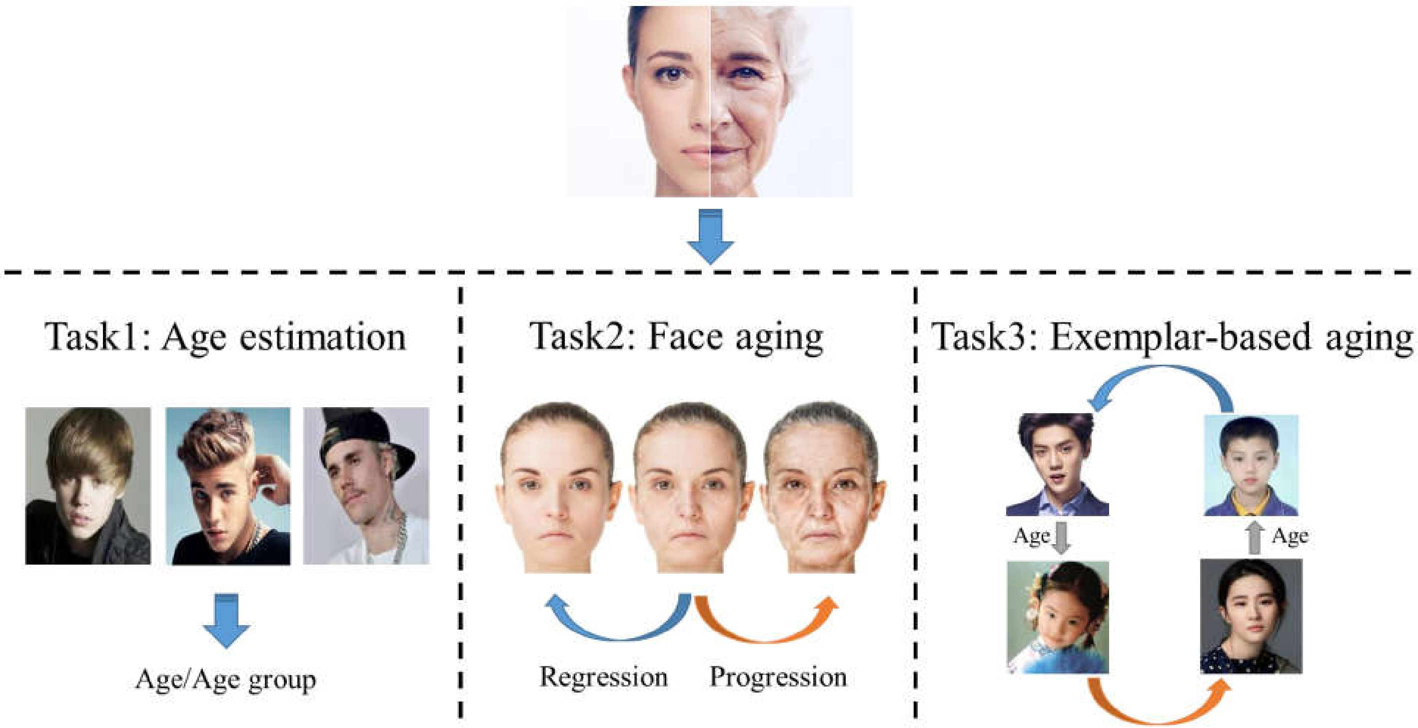
Model	Settings	Metric	Original	Bicubic	WTIP	SRCNN	SRGAN	URDGN	SRDense	CBN	Our-CNN	Ours
LightCNN	32×32, $4\times$	AUC	99.31	99.16	99.04	99.17	99.22	–	99.21	90.80	99.25	**99.28**
		FAR = 1%	97.77	96.10	95.83	96.23	96.93	–	96.90	46.77	**97.40**	97.03
		FAR = 0.1%	96.23	91.90	91.70	92.87	94.07	–	94.97	32.53	95.73	**96.10**
	16×16, $8\times$	AUC	99.31	90.68	89.97	91.42	96.77	93.60	96.35	89.98	97.92	**98.48**
		FAR = 1%	97.77	45.50	40.53	48.70	78.83	53.57	77.50	46.90	87.97	**90.86**
		FAR = 0.1%	96.23	21.17	24.47	23.50	56.60	27.10	57.03	31.13	68.33	**81.20**
	8×8, $16\times$	AUC	99.31	60.89	59.40	61.47	77.10	–	74.30	63.00	87.29	**89.40**
		FAR= = 1%	97.77	3.17	2.90	2.83	16.40	–	12.67	4.57	38.43	**42.87**
		FAR= = 0.1%	96.23	0.27	0.47	0.30	4.23	–	3.73	1.30	12.93	**22.83**

Aging

Hierarchical Face Aging through Disentangled Latent Characteristics

Disentangled Adversarial Autoencoder (DAAE)

- DAAE is the first attempt to achieve facial age analysis tasks in a universal framework.

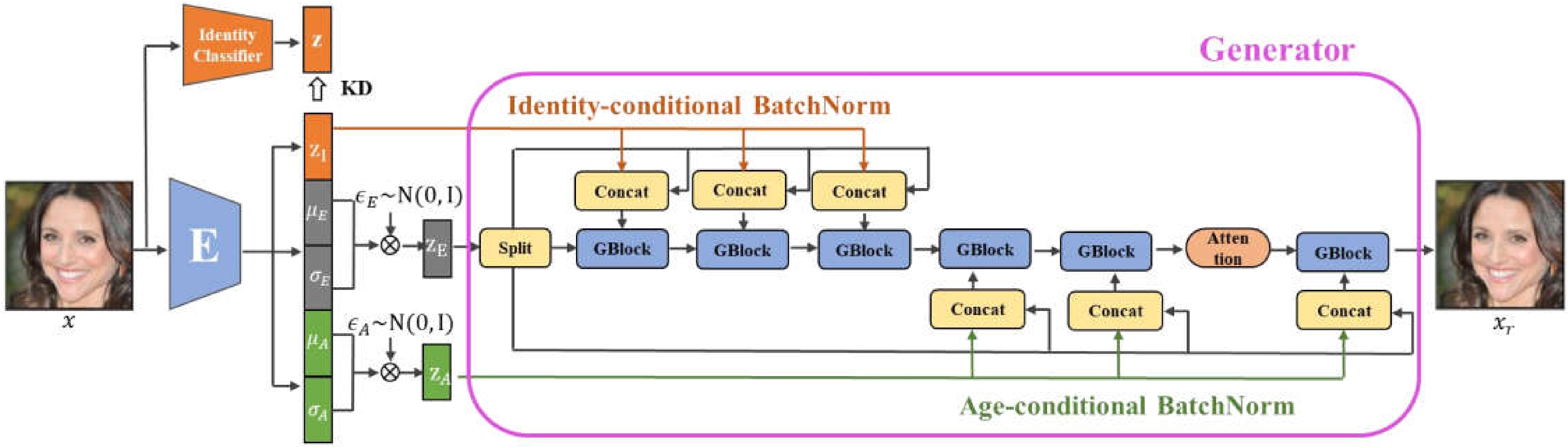


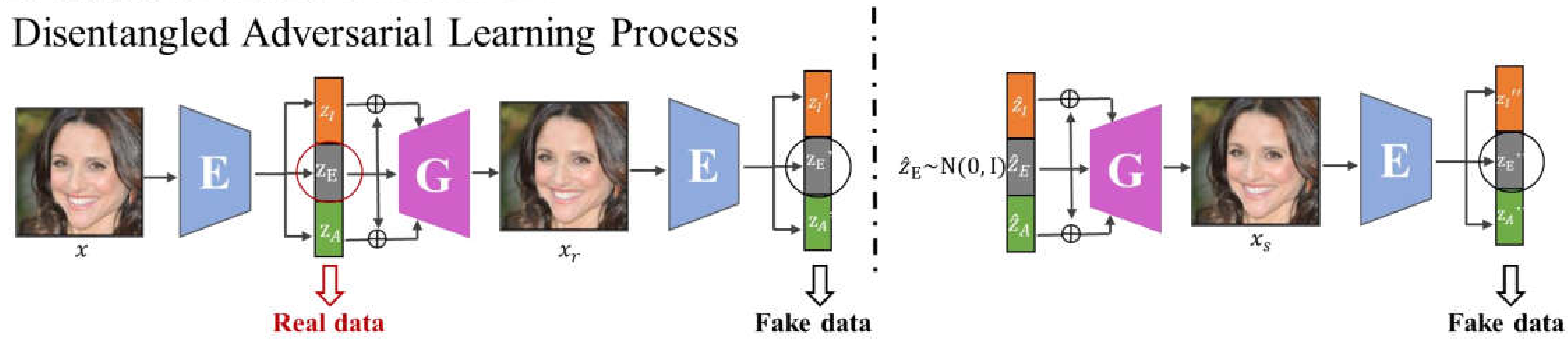
Peipei Li, Huaibo Huang, Yibo Hu, Xiang Wu, Ran He, Zhenan Sun. "Hierarchical Face Aging through Disentangled Latent Characteristics." **ECCV** 2020 (Oral).

Li et al, Hierarchical Face Aging through Disentangled Latent Characteristics, ECCV, 2020.

Aging

Hierarchical Face Aging through Disentangled Latent Characteristics



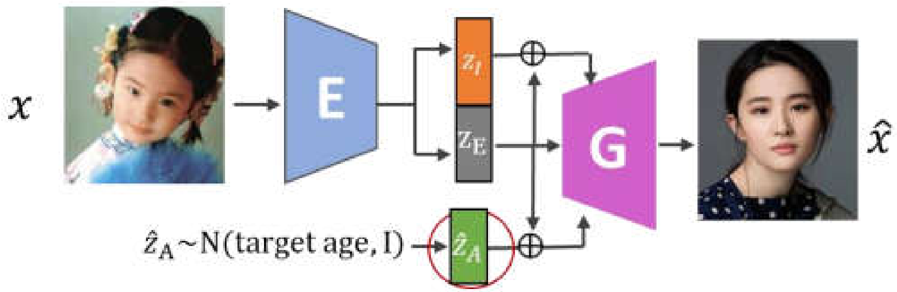


Aging

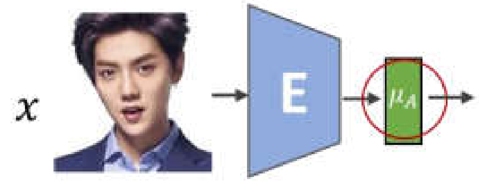
Hierarchical Face Aging through Disentangled Latent Characteristics

Inference and Sampling

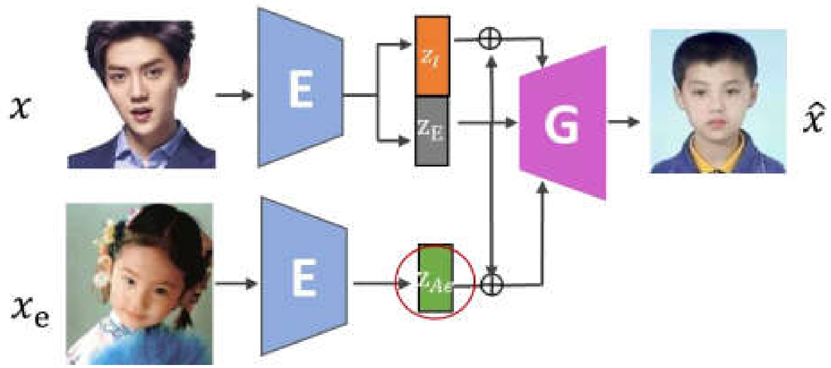
1) Face aging $\hat{x} = G(\hat{z}_A, z_I, z_E)$



3) Age estimation $\hat{y} = \frac{1}{C}\sum_{i=1}^{C} \mu_A^i$



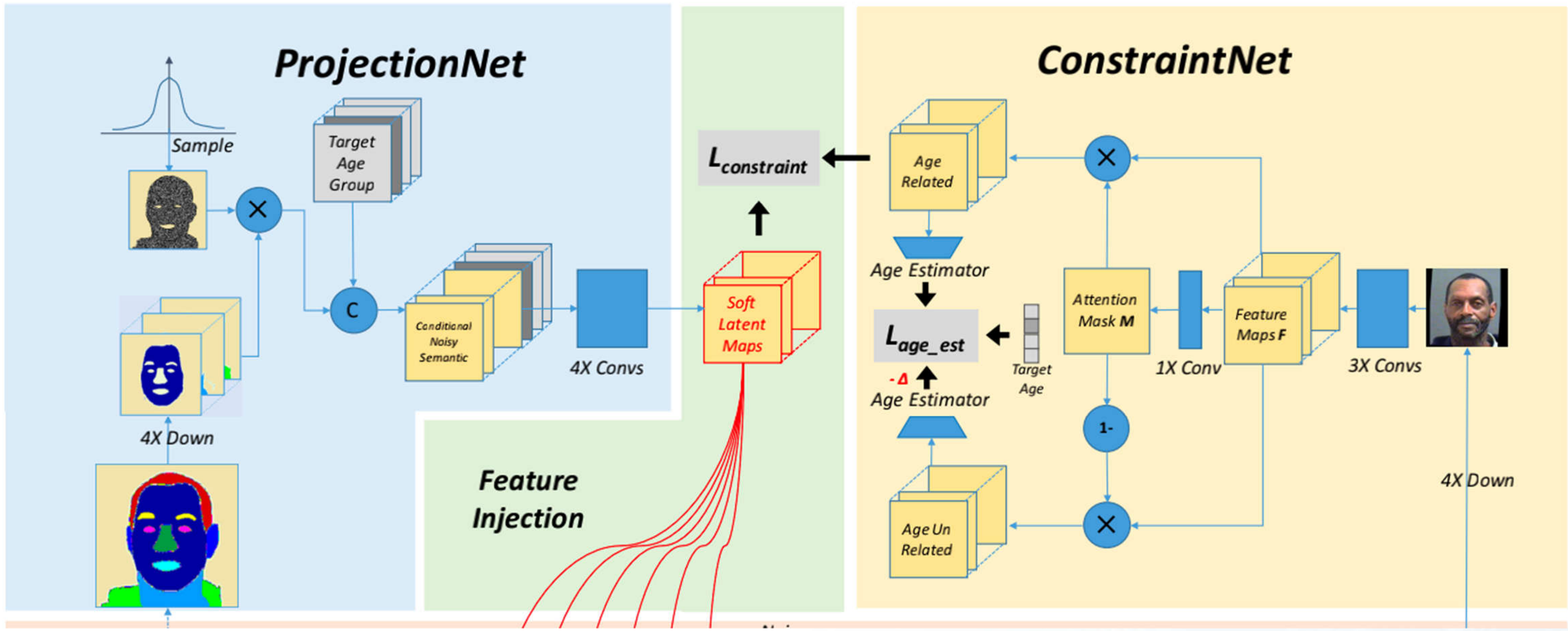
2) Exemplar-based face aging $\hat{x} = G(z_{A\_e}, z_I, z_E)$



Aging

A Unified Framework for Biphasic Facial Age Translation with Noisy-semantic Guided GANs

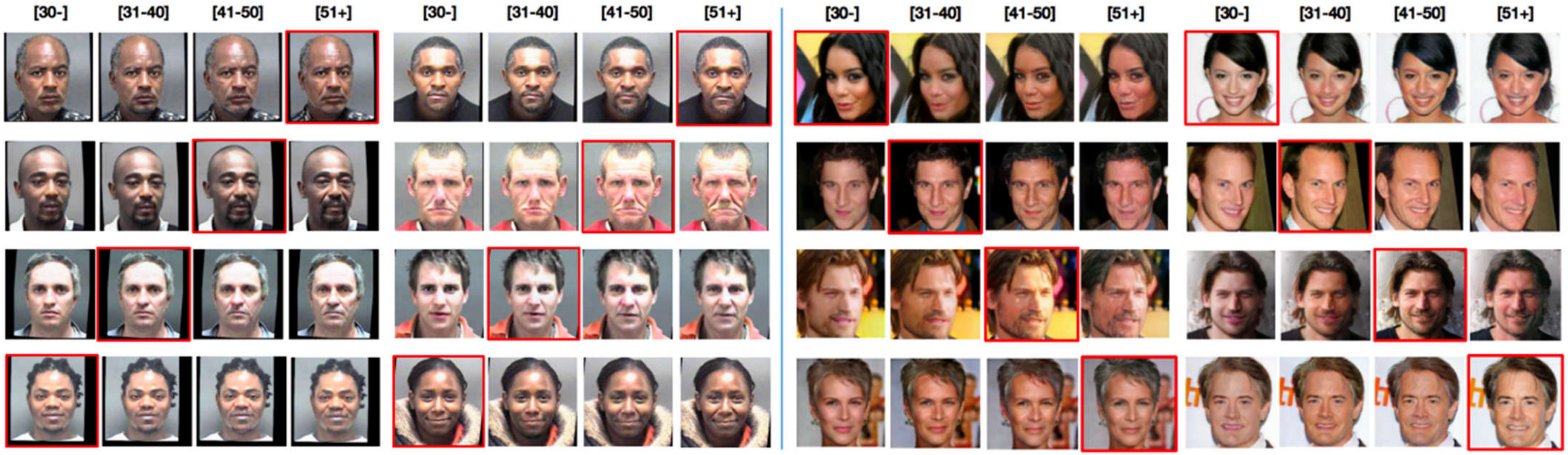
- Fine-grained Face Age Translation
- A Unified Framework for Data-Efficiency



Sun et al, A Unified Framework for Biphasic Facial Age Translation with Noisy-semantic Guided GANs, TIFS, 2022.

- **Comparison with Prior Works**

Train on MORPH and CACD Dataset 256*256 resolution



Train on FGNet dataset

Method	Year	MORPH Age Estimation Error	MORPH Identity Verification Rate	MORPH Image Quality	CACD Age Estimation Error	CACD Identity Verification Rate	CACD Image Quality
CAAE	2017	10.34 ± 5.63	34.83 (71.75)	-	5.16 ± 7.08	3.59 (59.90)	-
IPC-GAN	2017	1.74 ± 7.44	99.86 (94.04)	-	8.11 ± 9.69	99.19 (91.60)	-
Dual cGAN	2018	2.44 ± 6.03	99.99 (93.15)	-	3.28 ± 8.01	99.88 (93.85)	-
SPT-GAN	2020	1.53 ± 6.50	100.00 (95.67)	40.12	1.78 ± 7.53	99.92 (96.13)	46.73
NSG-GAN	2021	**1.20 ± 6.81**	99.99 (95.27)	**35.58**	**1.45 ± 8.02**	**99.93 (94.20)**	**40.24**

Face Parsing

Large-scale Image Database Generation for Face Parsing

Motivation

- It is expensive and time-consuming to construct a large-scale pixel-level manually annotated dataset for face parsing.
- We propose a D2VG, which can synthesize large-scale paired face images and parsing maps from a stand Gaussian distribution.

Peipei Li, Yinglu Liu, Hailin Shi, Xiang Wu, Yibo Hu, Ran He, Zhenan Sun. "Dual-structure Disentangling Variational Generation for Data-limited Face Parsing." **ACM MM**(Oral), 2020.

Li et al, Dual-Structure Disentangling Variational Generation for Data-Limited Face Parsing, ACM MM, 2020.

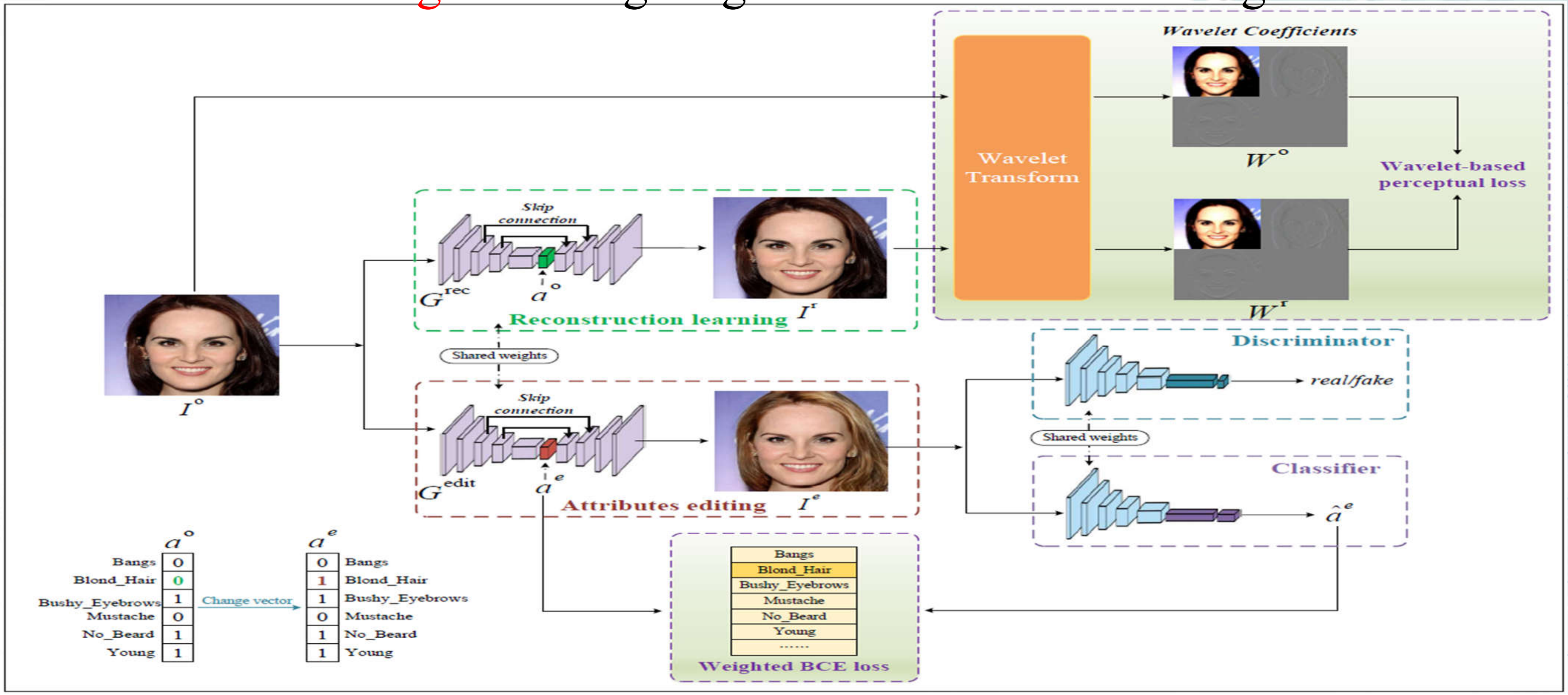
Facial Attribute Editing

www.ia.ac.cn

Controllable Multi-Attribute Editing of High-Resolution Face Images

Motivation:

- Most of existing methods have two main limitations:
  - Only applicable to face images with relative low resolutions
  - Undesired changes of image regions not related to the target attribute



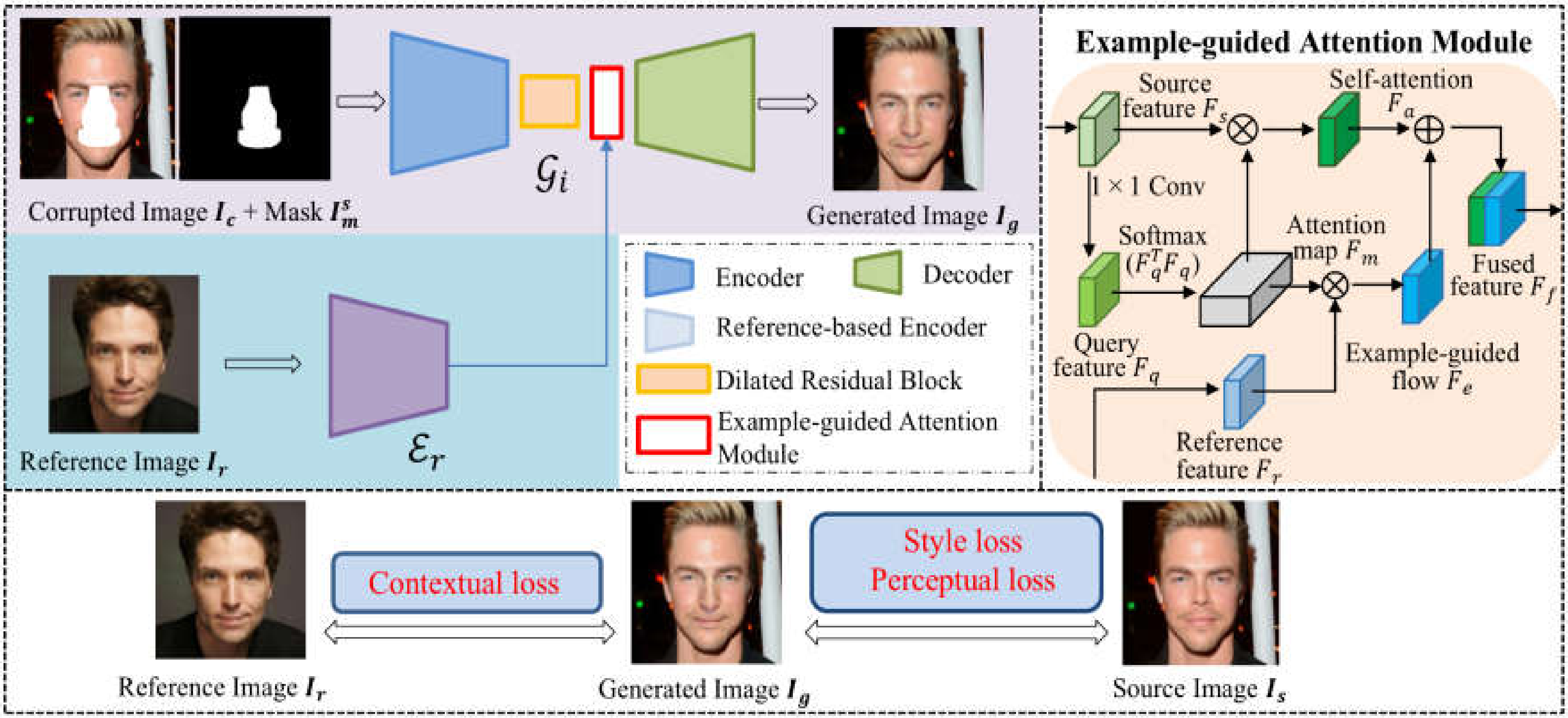
Deng et al, Controllable Multi-Attribute Editing of High-Resolution Face Images, TIFS, 2019.

Facial Attribute Editing

Controllable Multi-Attribute Editing of High-Resolution Face Images

Facial Attribute Editing

Reference-guided Face Component Editing

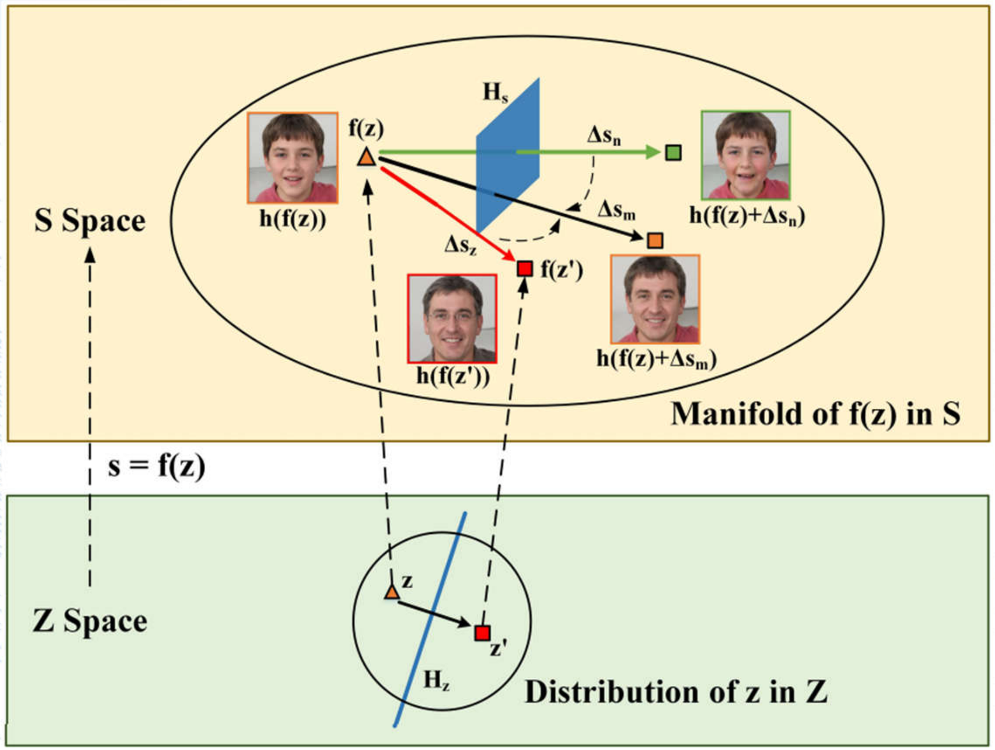


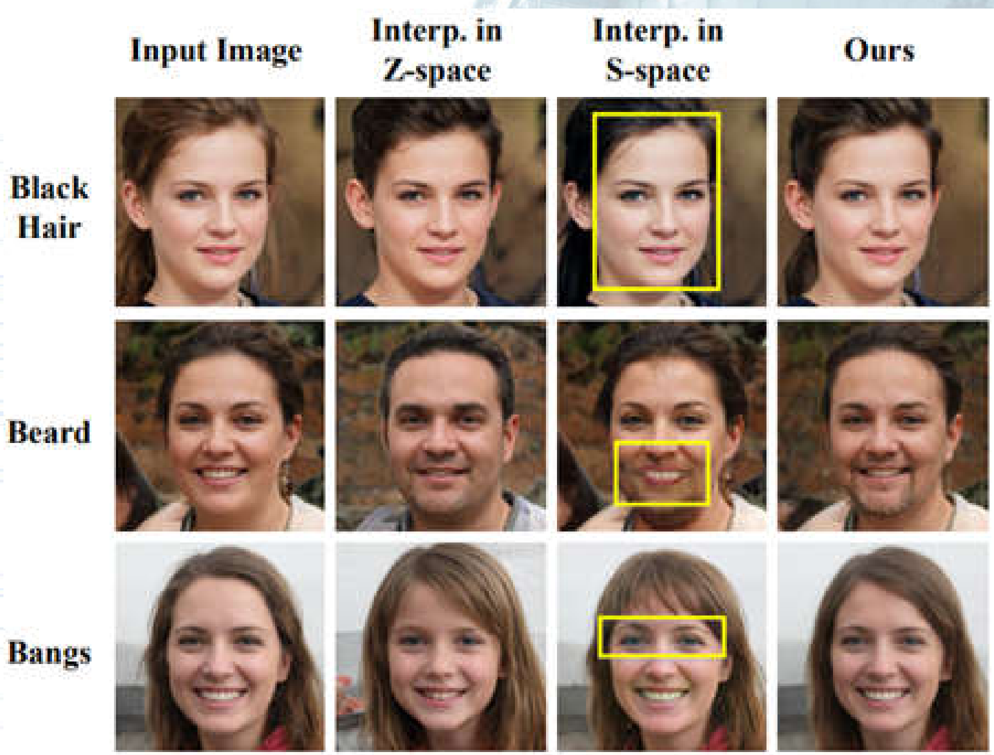
Deng et al, Reference-guided Face Component Editing, IJCAI, 2020.

Face Manipulation

Problem: Interpolating only in the Style Space would lead to disentangled but unnatural translation results.

Solution: Combining the translation effect in both Z and S space to make the best of both worlds.

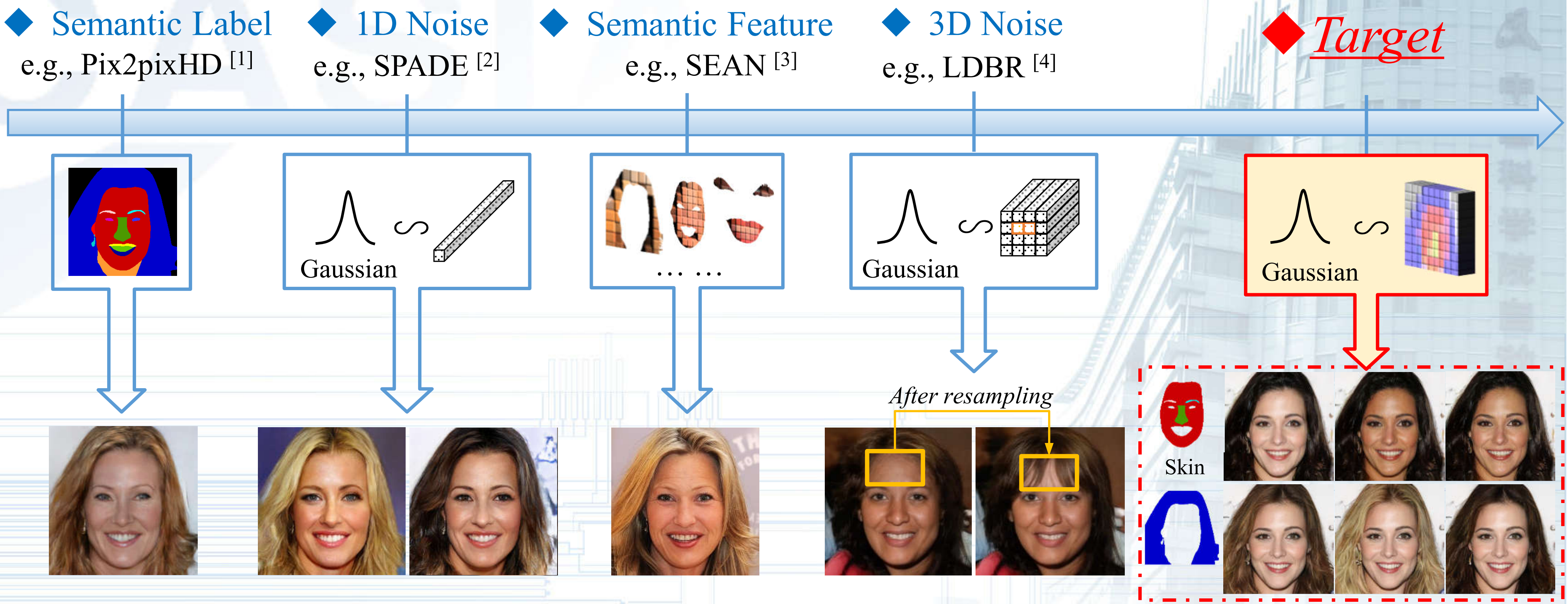




Liu et al, Towards Spatially Disentangled Manipulation of Face Images with Pre-trained StyleGANs, IEEE-TCSVT, 2022.

Face Synthesis

Semantic-aware Noise Driven Portrait Synthesis and Manipulation



How to achieve *3D Semantic Noise*

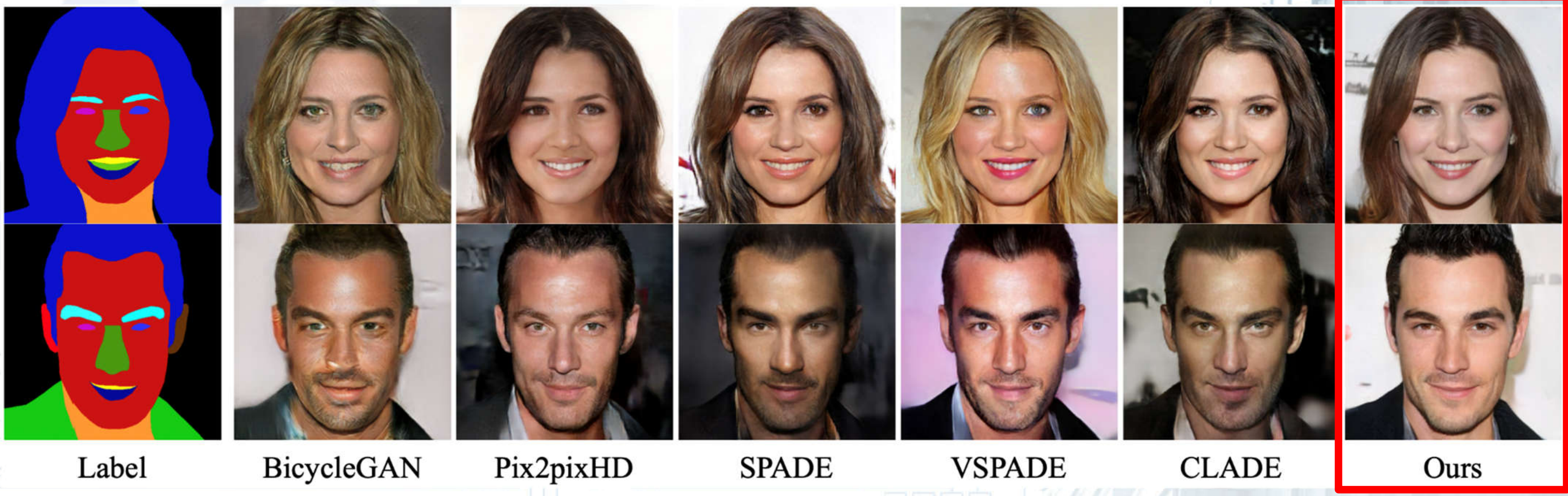
semantic controllability and style diversity

Deng et al, Semantic-aware Noise Driven Portrait Synthesis and Manipulation, TMM, 2022.

Face Synthesis

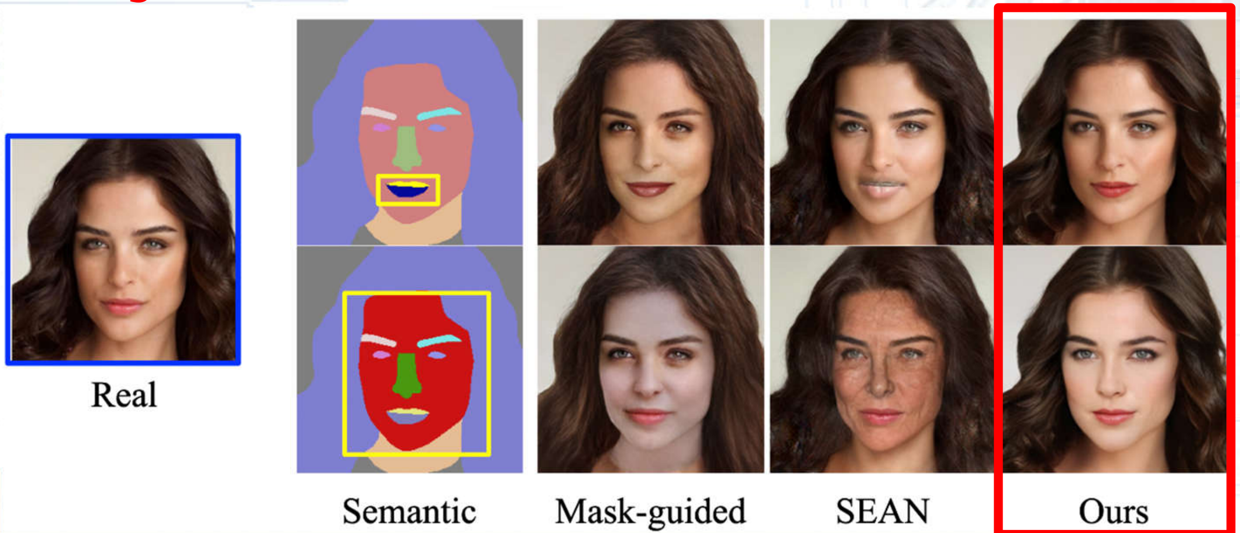
High-fidelity

- Semantic image synthesis



Controllability

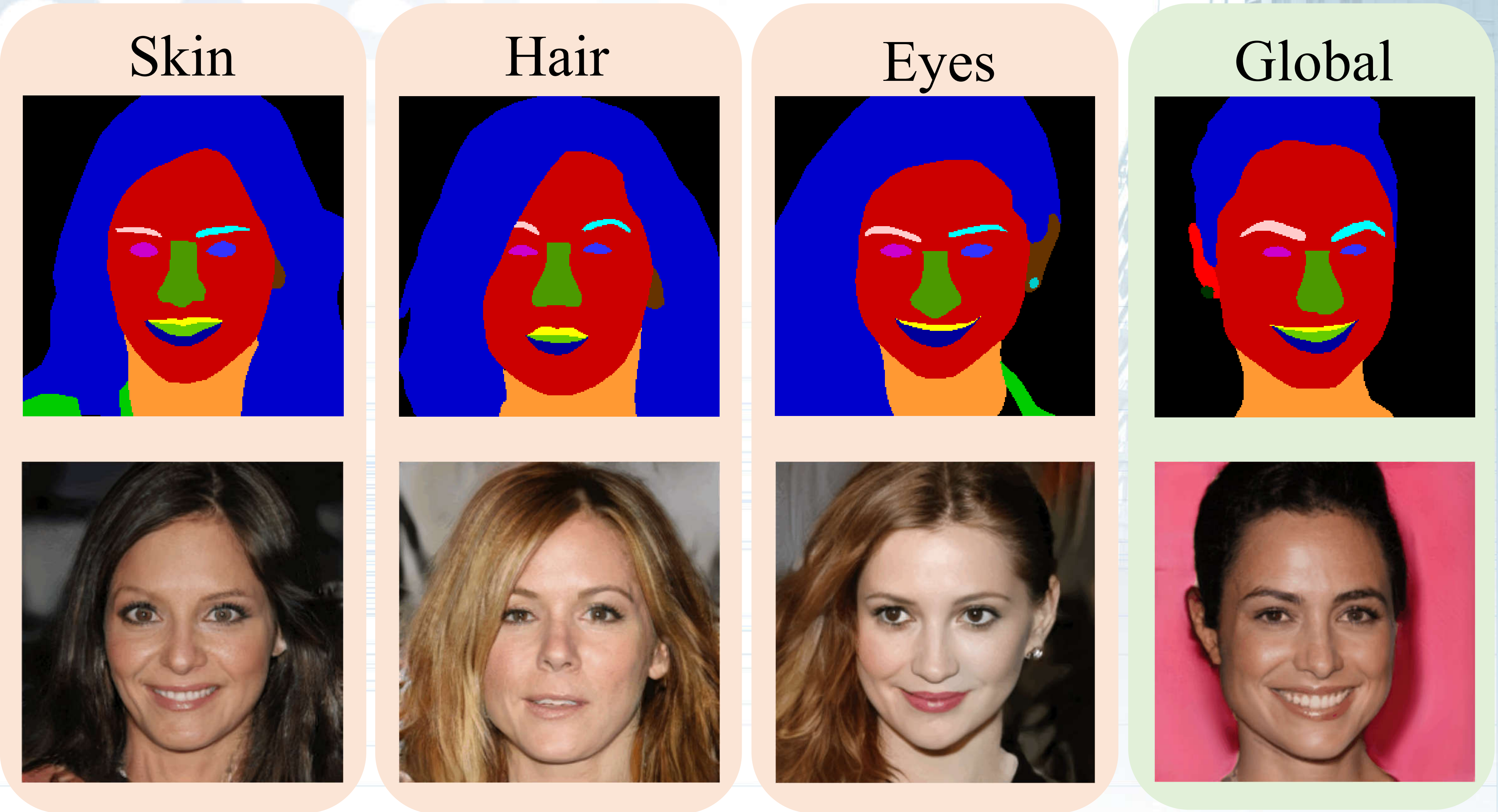
- Real portrait manipulation



Face Synthesis

Diversity

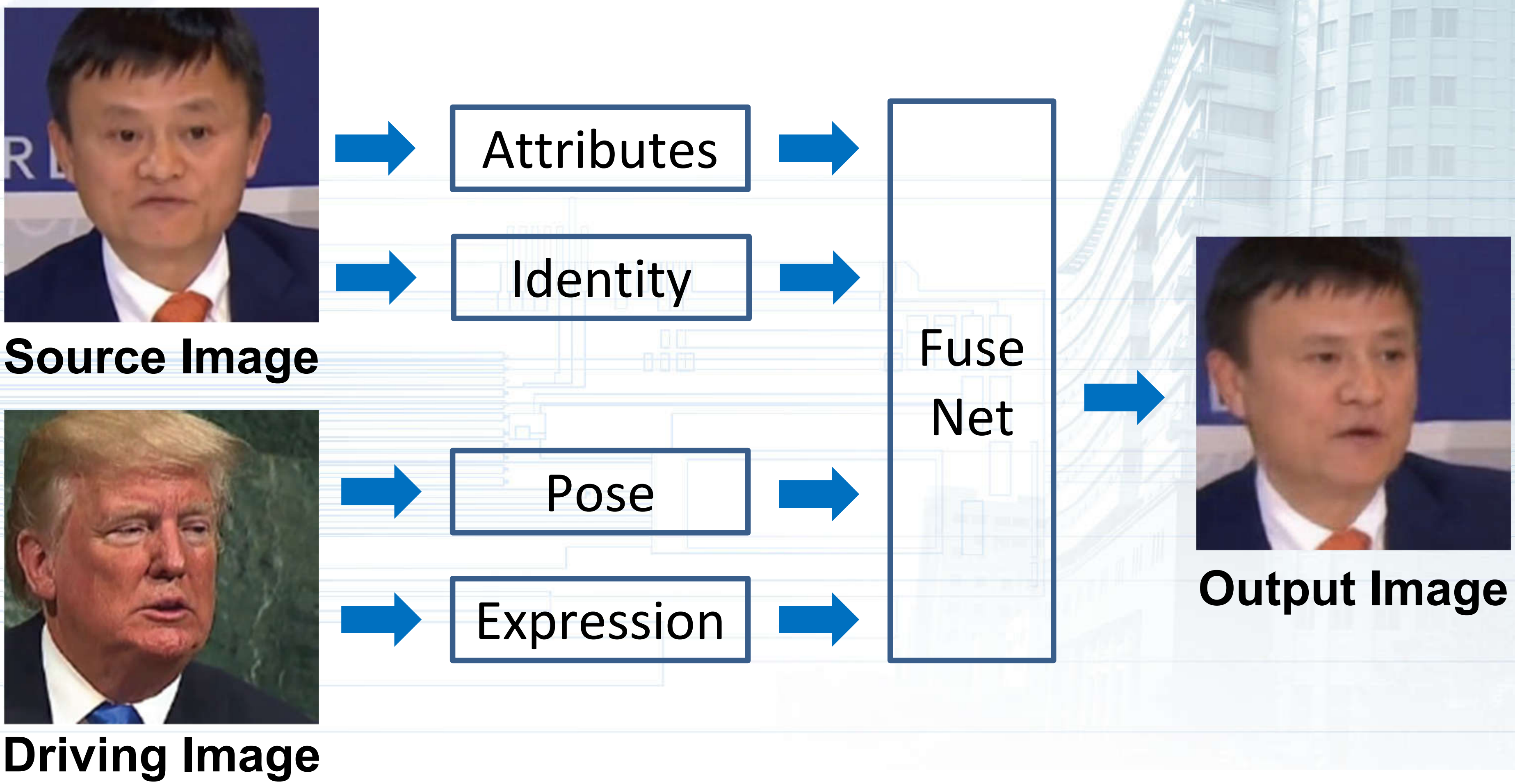
- Synthesis with diverse style



Face Reenactment

Semantic-aware One-shot Face Re-enactment with Dense Correspondence Estimation

- **Face re-enactment:** Control **3D-interpretable semantics** of an input face based on the reference image

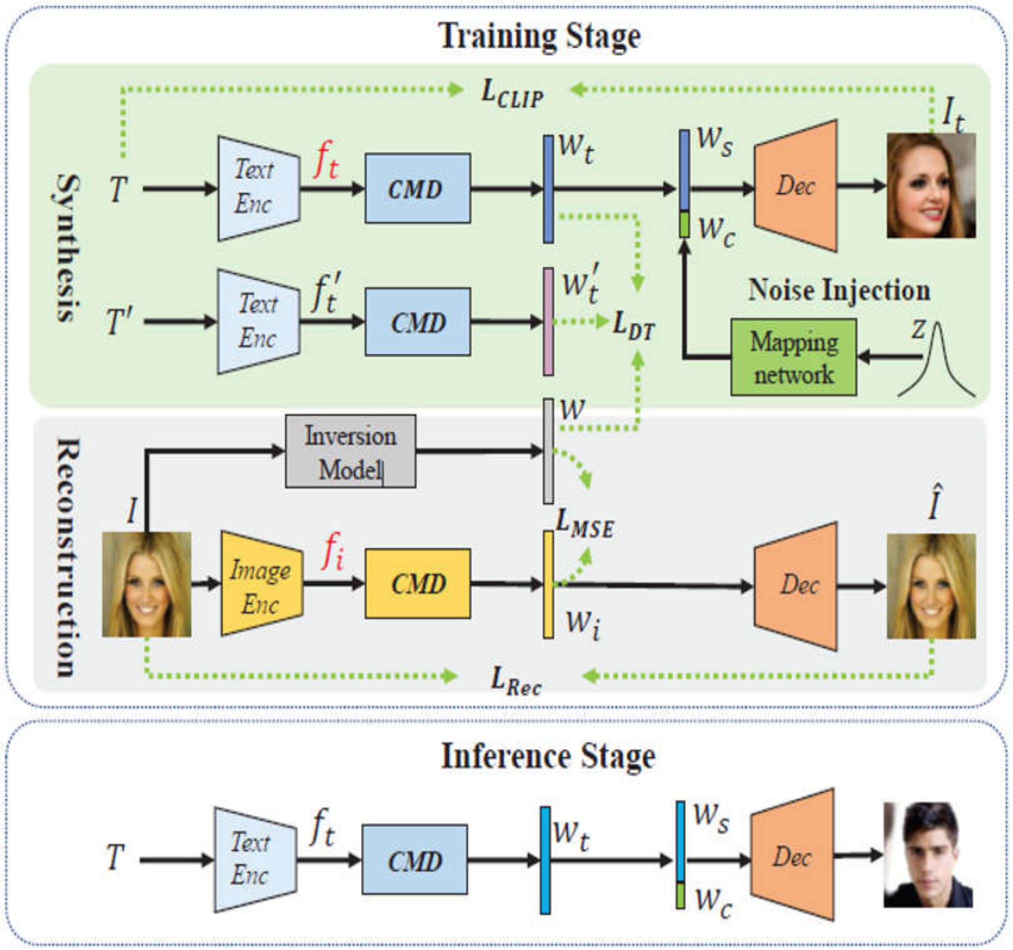


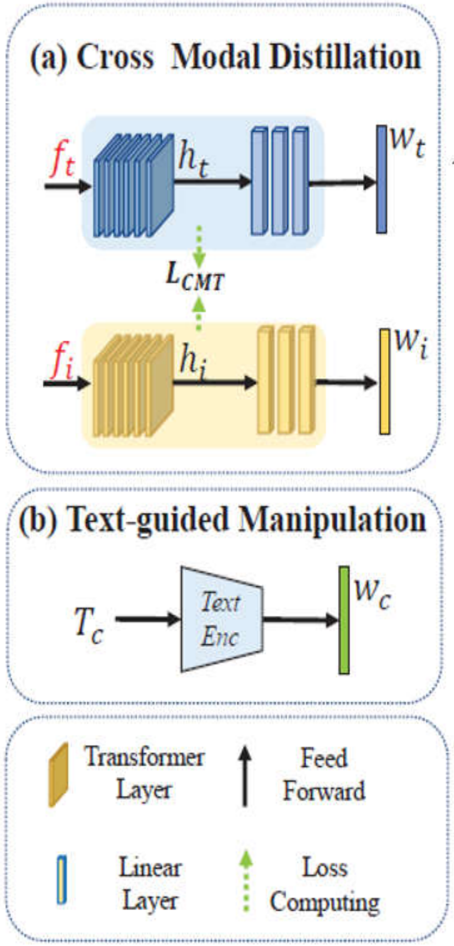
Liu et al, Semantic-aware One-shot Face Re-enactment with Dense Correspondence Estimation, MIR, 2022.

Text-to-Face

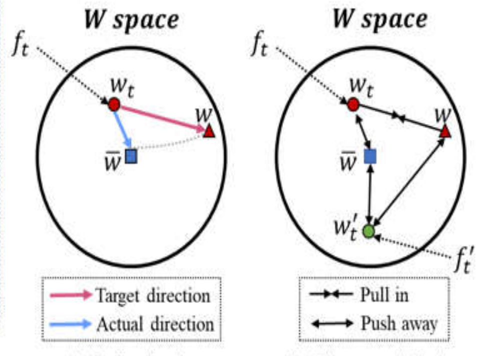
AnyFace: Free-style Text-to-Face Synthesis and Manipulation

■ Two Stream Framework





■ Diverse Triplet loss



(a) Pair-wise loss (b) Diverse triplet loss

$$\mathcal{L}_S = \mathcal{L}_{DT} + \lambda_{CMT}\mathcal{L}_{CMT}^{T} + \lambda_{CLIP}\mathcal{L}_{CLIP}$$

$$\mathcal{L}_T = \mathcal{L}_{MSE} + \lambda_{CMT}\mathcal{L}_{CMT}^{I} + \lambda_{Rec}\mathcal{L}_{Rec}$$

$$\mathcal{L}_{DT} = max\left\{\frac{\langle w_t, w\rangle}{\langle w_t, \overline{w}\rangle} - \frac{\langle w_t', w\rangle}{\langle w_t', \overline{w}\rangle} + m, 0\right\}$$

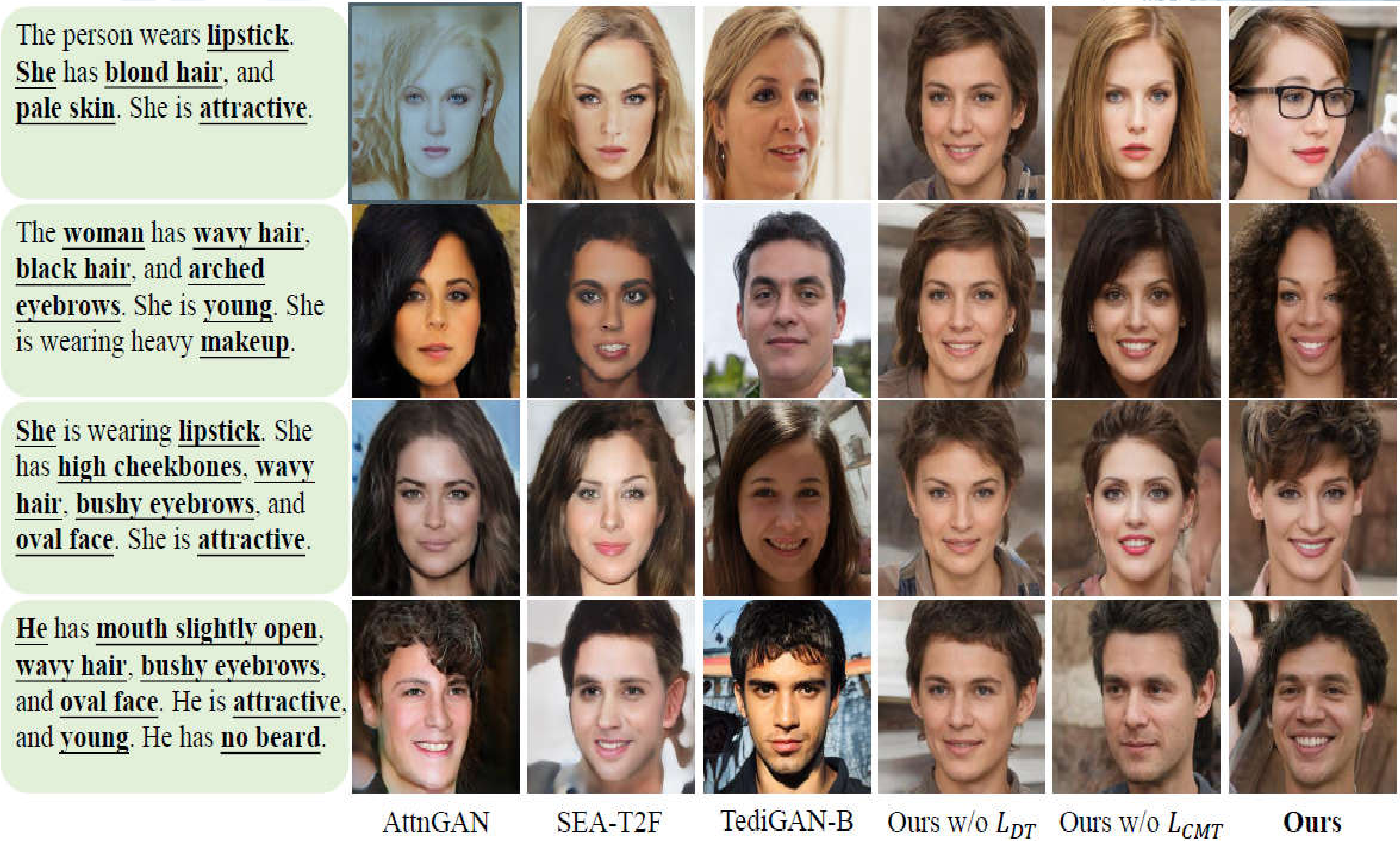
$$\mathcal{L}_{CLIP} = \frac{f_t \cdot f_i^t}{||f_t|| \times ||f_i^t||}$$

Sun et al, AnyFace: Free-style Text-to-Face Synthesis and Manipulation, CVPR, 2022.

Text-to-Face

AnyFace: Free-style Text-to-Face Synthesis and Manipulation

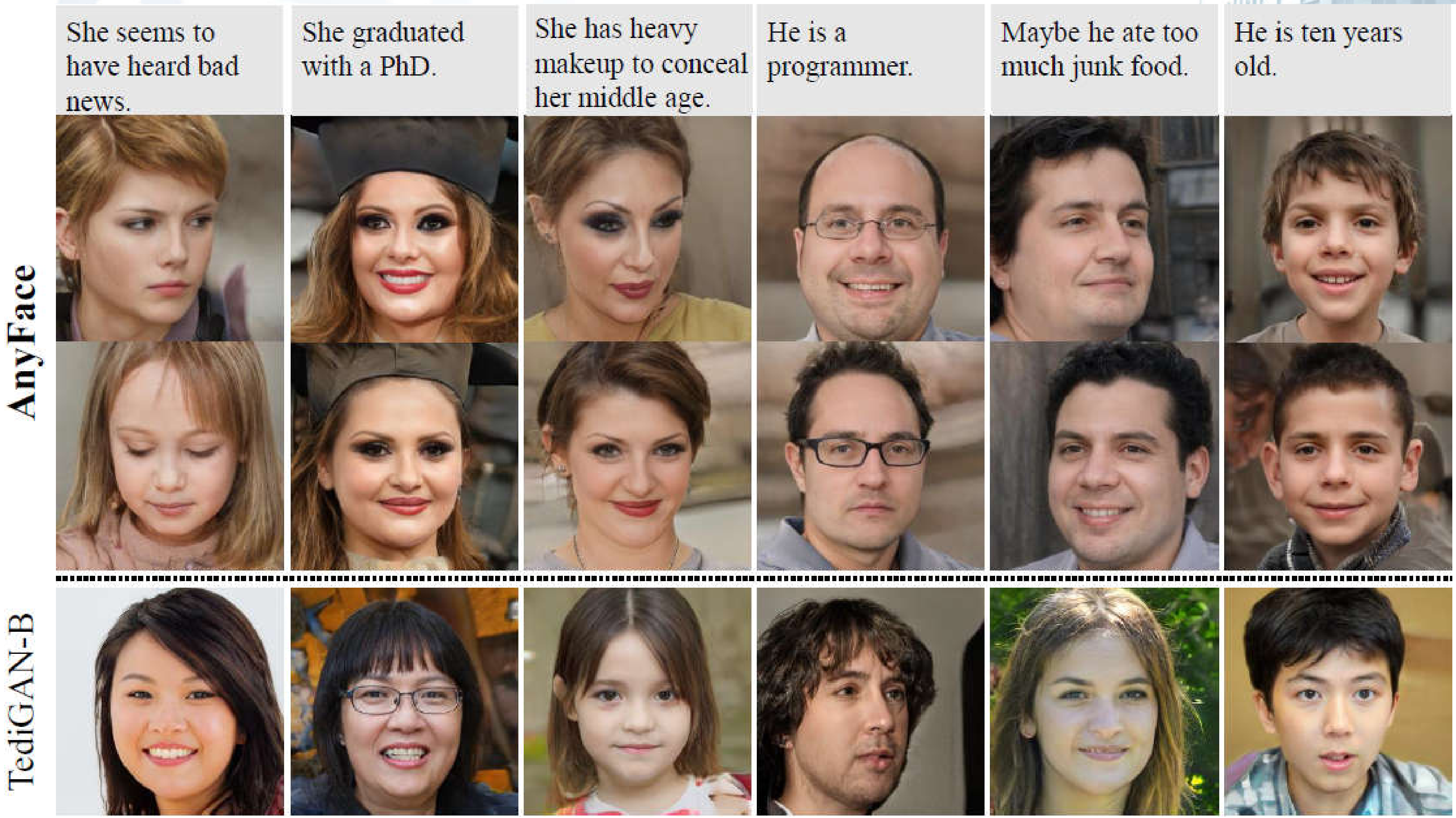
- **Comparison**



Text-to-Face

AnyFace: Free-style Text-to-Face Synthesis and Manipulation

- **Open-world Results**



AI enables face manipulation easier and has caused security risks

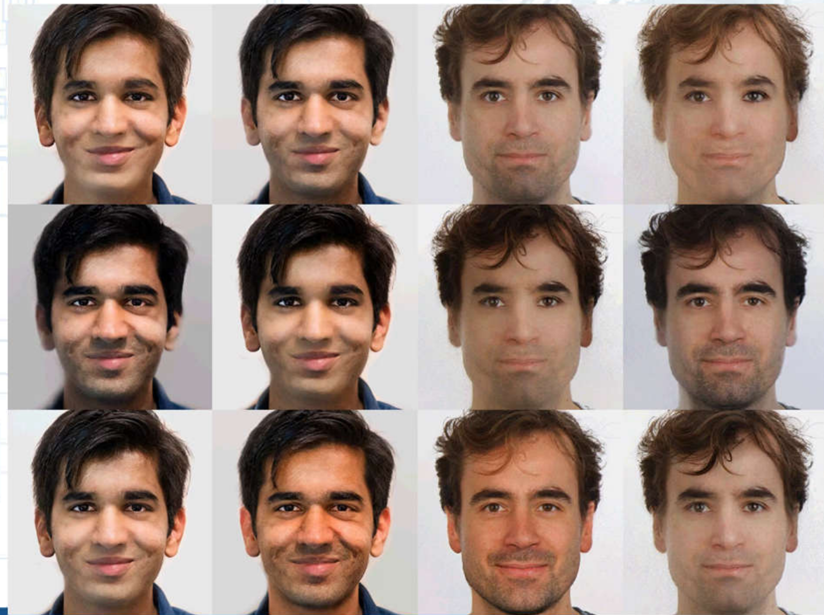
Fraudster Dimitri de Angelis Jailed for Fake Celebrity Friend Photoshop Scam

Conman scammed investors out of $8.5m by pretending to be friends with Queen, Pope, Bush and Clinton

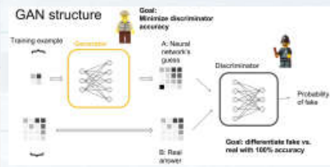
By Hannah Osborne
March 1, 2013 16:47 GMT

Dimitri de Angelis with Bill Clinton

Experts



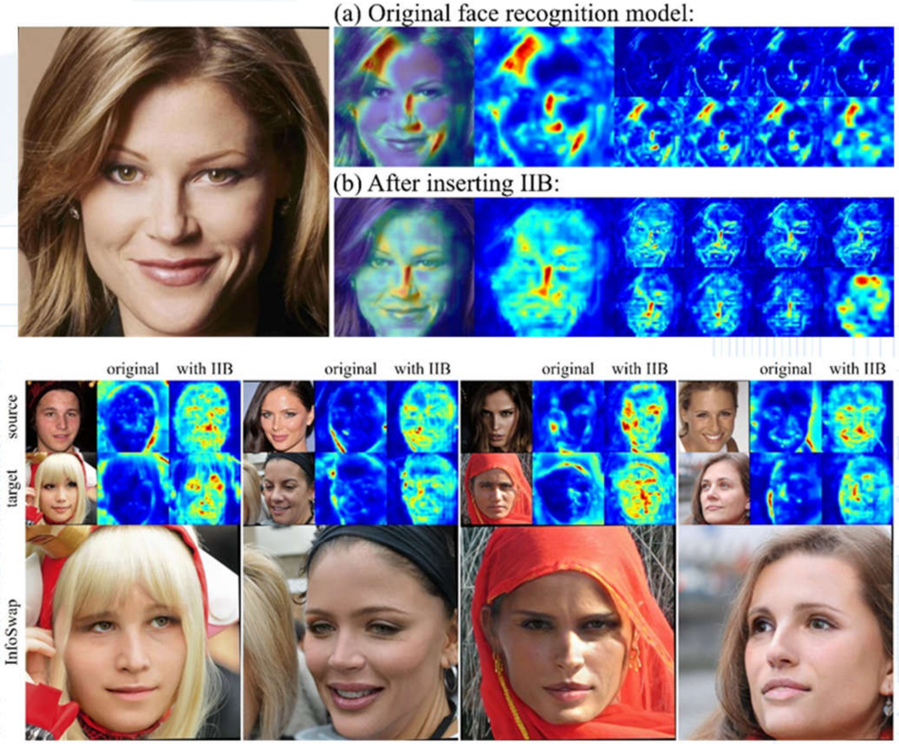
AI Enabled

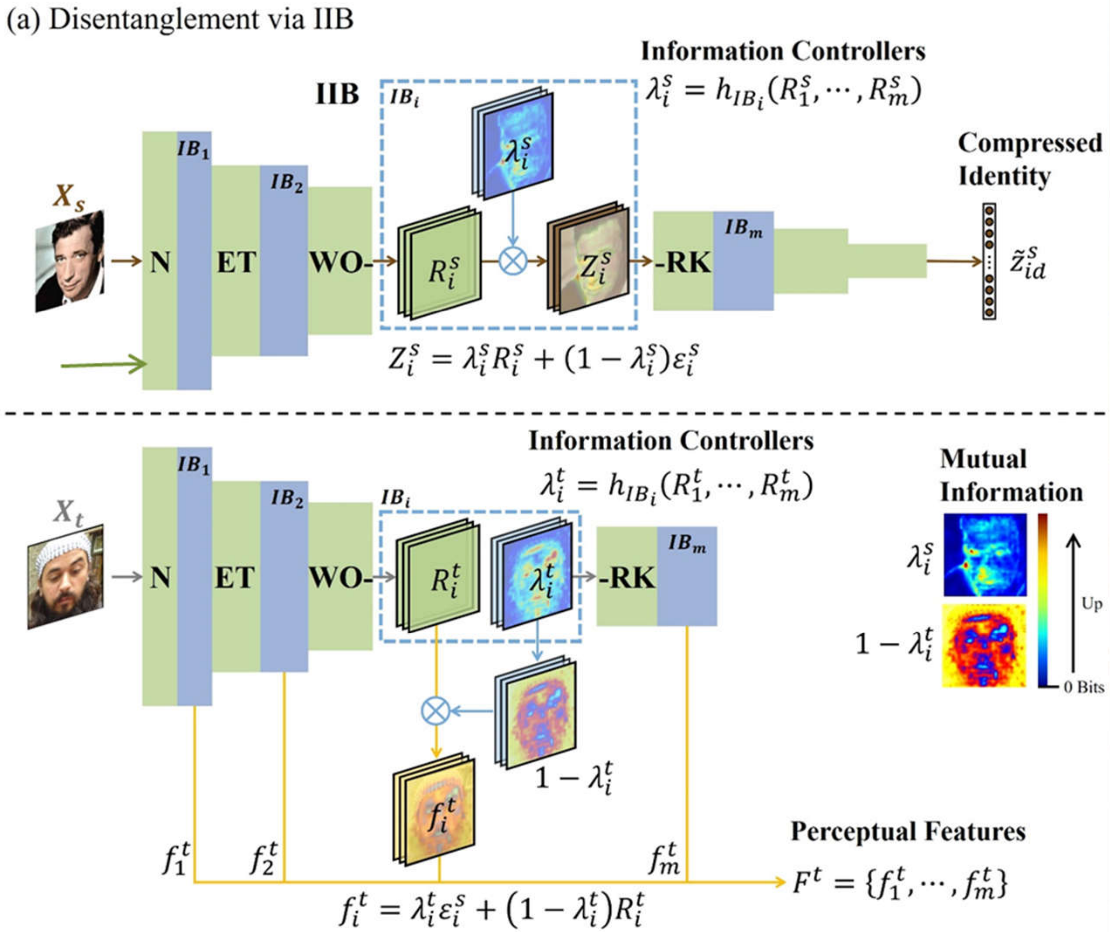


from Internet

Identity Swap

- **Identity leakage: Information Bottleneck network**

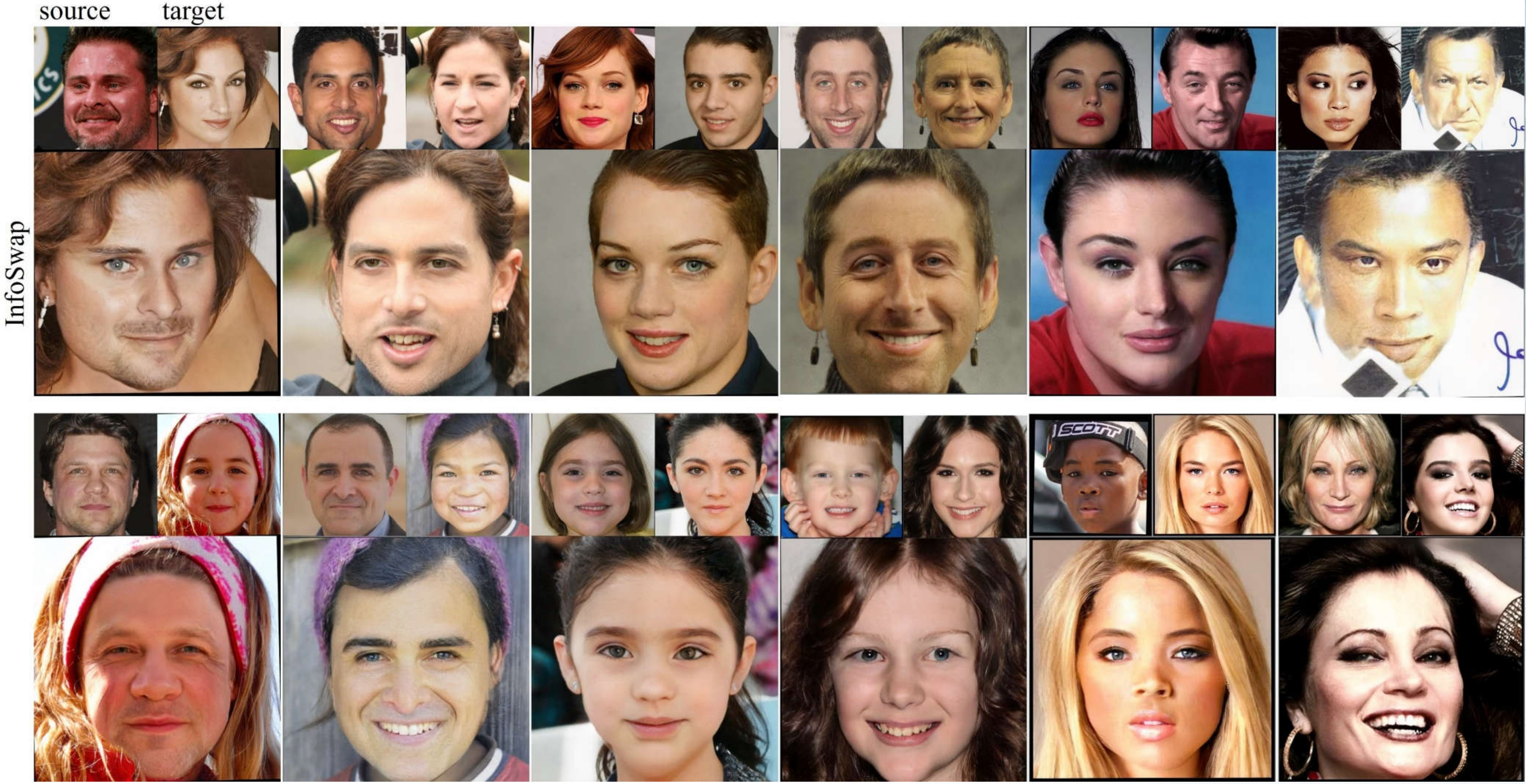




Gege Gao et al. Information Bottleneck Disentanglement for Identity Swapping. IEEE Computer Vision and Pattern Recognition, 2021.

Identity Swap

- **Identity leakage: Information Bottleneck network**



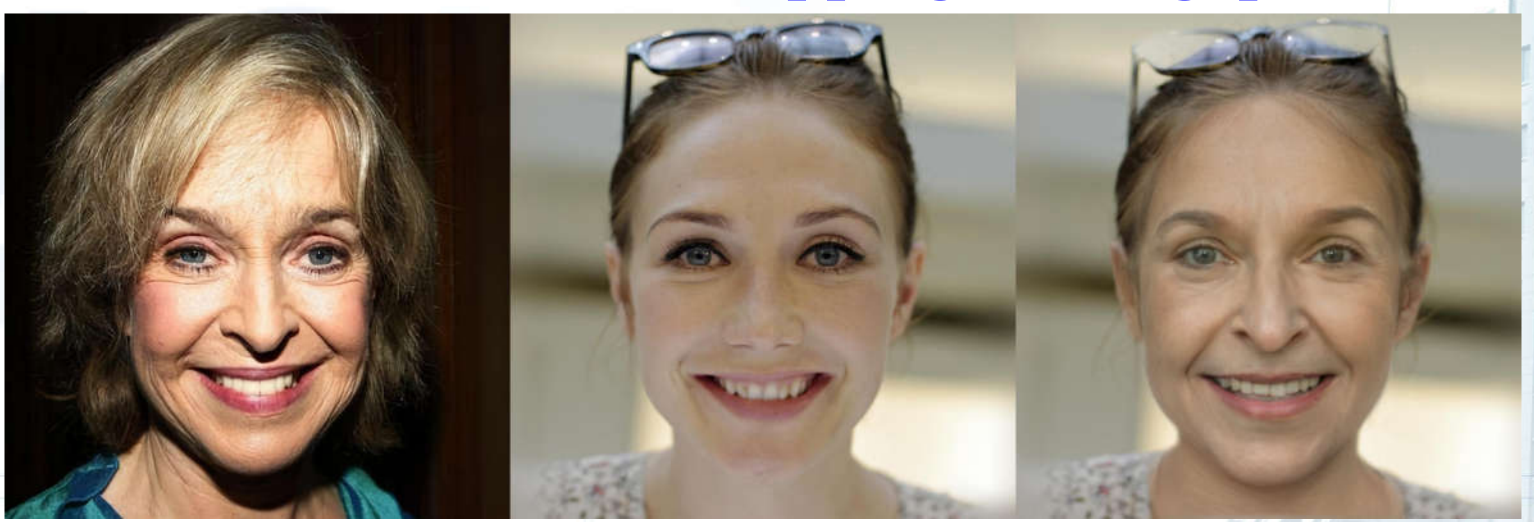
Gege Gao et al. Information Bottleneck Disentanglement for Identity Swapping. IEEE Computer Vision and Pattern Recognition, 2021.

中国科学院
自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

www.ia.ac.cn

Identity Swap

One Shot Face Swapping on Megapixels



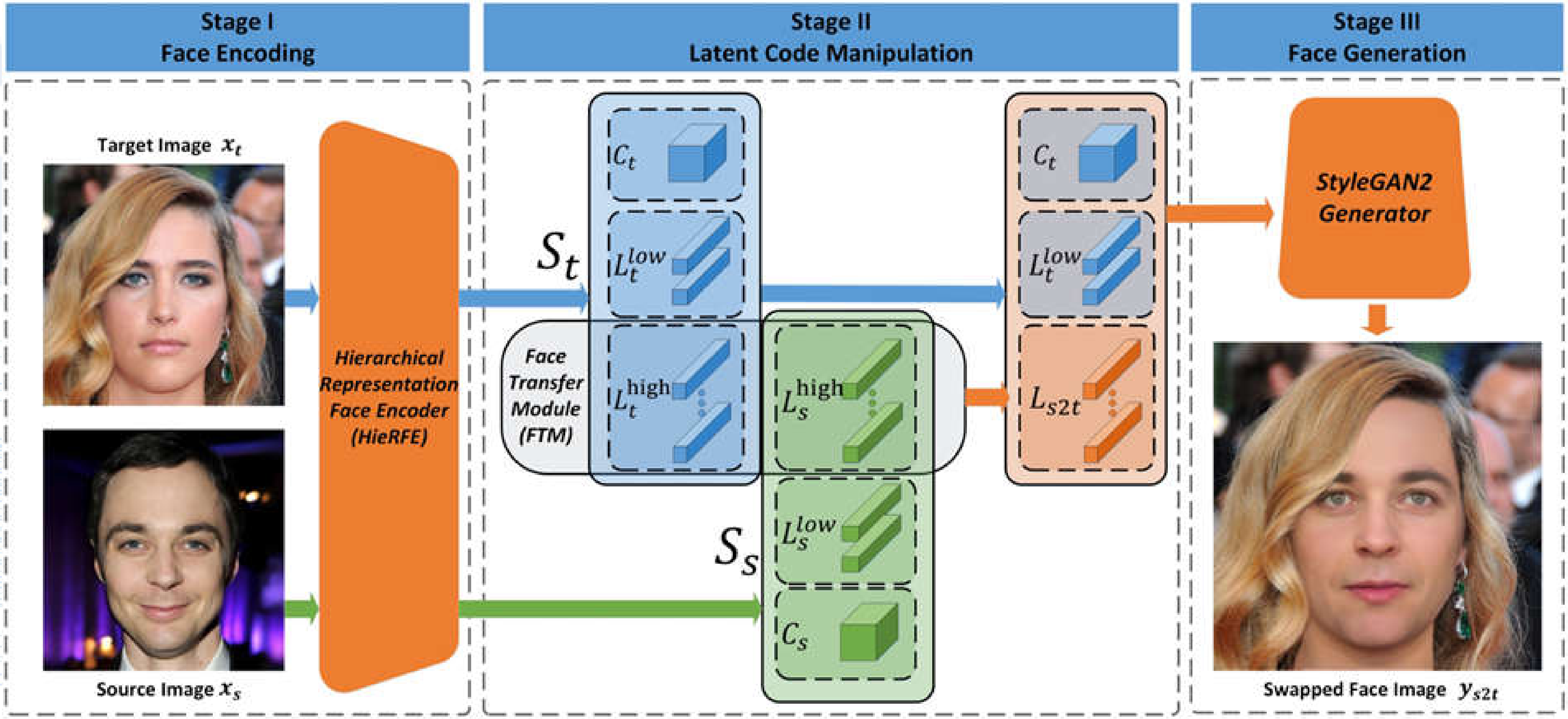
1. Problems
   - Only faces at 256 can be one shot swapped previously
   - How to swap faces using high resolution images?
2. Key Issues
   - Incapable of high-quality face generation based on compressed representations
   - Adversarial training is unstable
   - Hardware constraints(GPU memory)
3. Solution
   - StyleGAN2 + Its Appendages (Face Encoder & Face Transfer Module)

Yuhao Zhu, Qi Li, Jian Wang, Cheng-Zhong Xu and Zhenan Sun, One Shot Face Swapping on Megapixels, CVPR 2021.

Identity Swap

www.ia.ac.cn

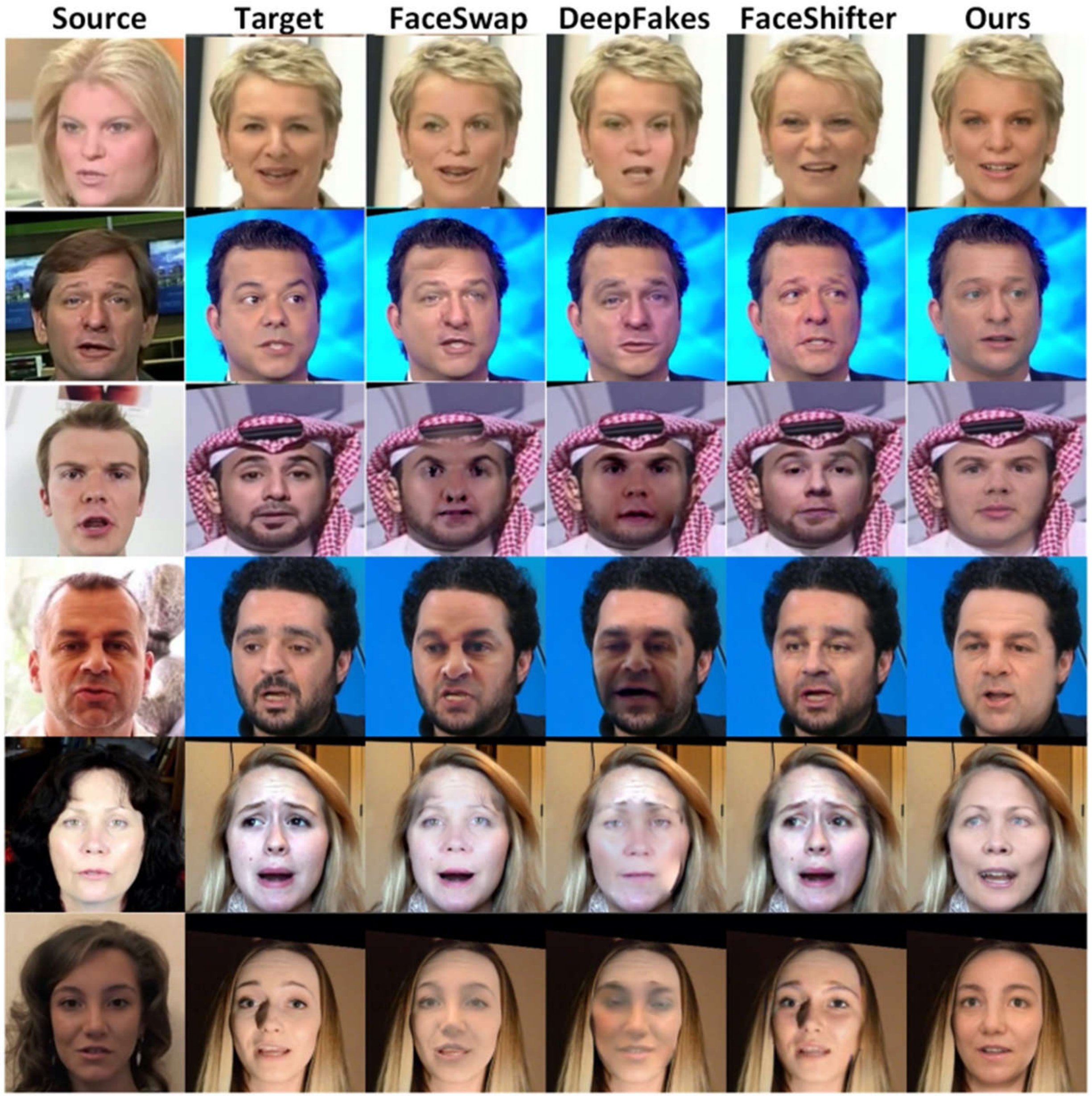
One Shot Face Swapping on Megapixels



Yuhao Zhu, Qi Li, Jian Wang, Cheng-Zhong Xu and Zhenan Sun, One Shot Face Swapping on Megapixels, CVPR 2021.

Identity Swap

One Shot Face Swapping on Megapixels



Identity Swap

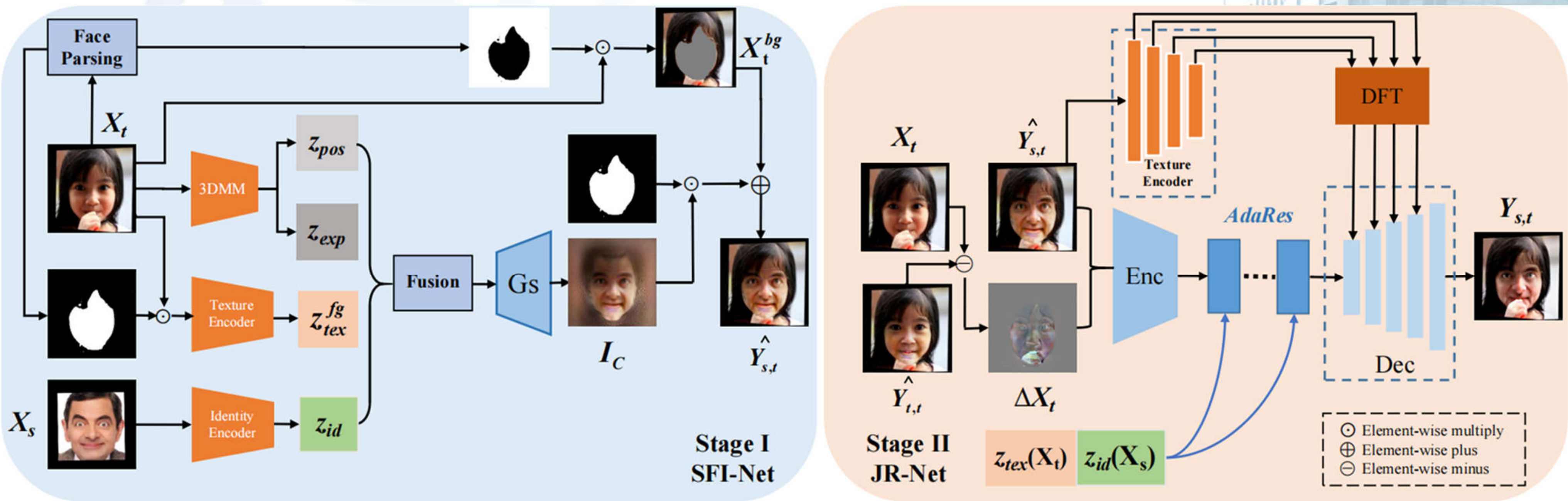
One Shot Face Swapping on Megapixels

Experiments on CelebA-HQ

Method	ID similarity ↑	pose ↓	expression ↓	FID ↓
Ours	0.5014	3.58	2.87	10.16

Identity Swap

- **Heterogenous domain**



$\mathcal{L}_{id}$ $\mathcal{L}_{exp}$ $\mathcal{L}_{pos}$ $\mathcal{L}_{GAN}$ $\mathcal{L}_{rec}$ $\mathcal{L}_{CX}$ $\mathcal{L}_{ppl}$ | $\mathcal{L}_{id}$ $\mathcal{L}_{GAN}$ $\mathcal{L}_{rec}$ $\mathcal{L}_{vgg}$ $\mathcal{L}_{CX}$

$\mathcal{L}_{id} = 1 - \langle z_{id}(\hat{Y}_{s,t}), z_{id}(X_s)\rangle$ $\mathcal{L}_{ppl} = \mathbb{E}_{w,y\sim N(0,I)}(\|J_w^T y\|_2 - a)^2$ | $\triangle X_t = X_t - \hat{Y}_{t,t}$

$\mathcal{L}_{CX} = -\log(CX(F^l_{vgg}(\hat{Y}_{s,t}), F^l_{vgg}(X_t)))$ $J_w = \frac{\partial \hat{Y}_{s,t}}{\partial w},\ w \sim f(C)$ | $AdaIN(h^i,\gamma^i,\beta^i) = \gamma^i_{\{id,att\}} \odot \frac{h^i-\mu^i}{\sigma^i} + \beta^i_{\{id,att\}}$

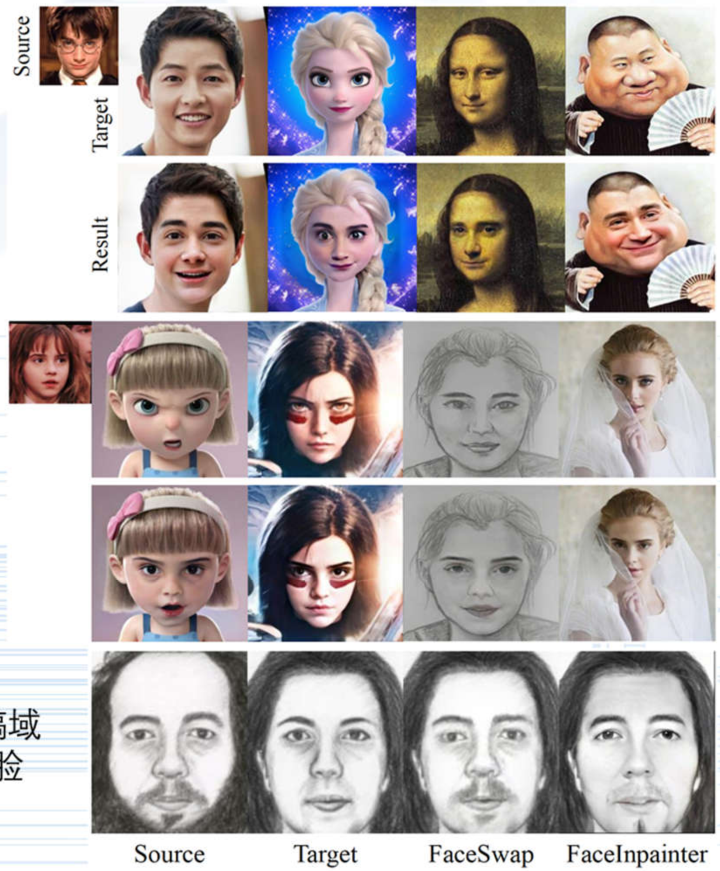
Jia Li, Zhaoyang Li, Jie Cao, Xingguang Song, Ran He. FaceInpainter: High Fidelity Face Adaptation to Heterogeneous Domains. CVPR 2021: 5089-5098

Identity Swap

- **Heterogenous domain**

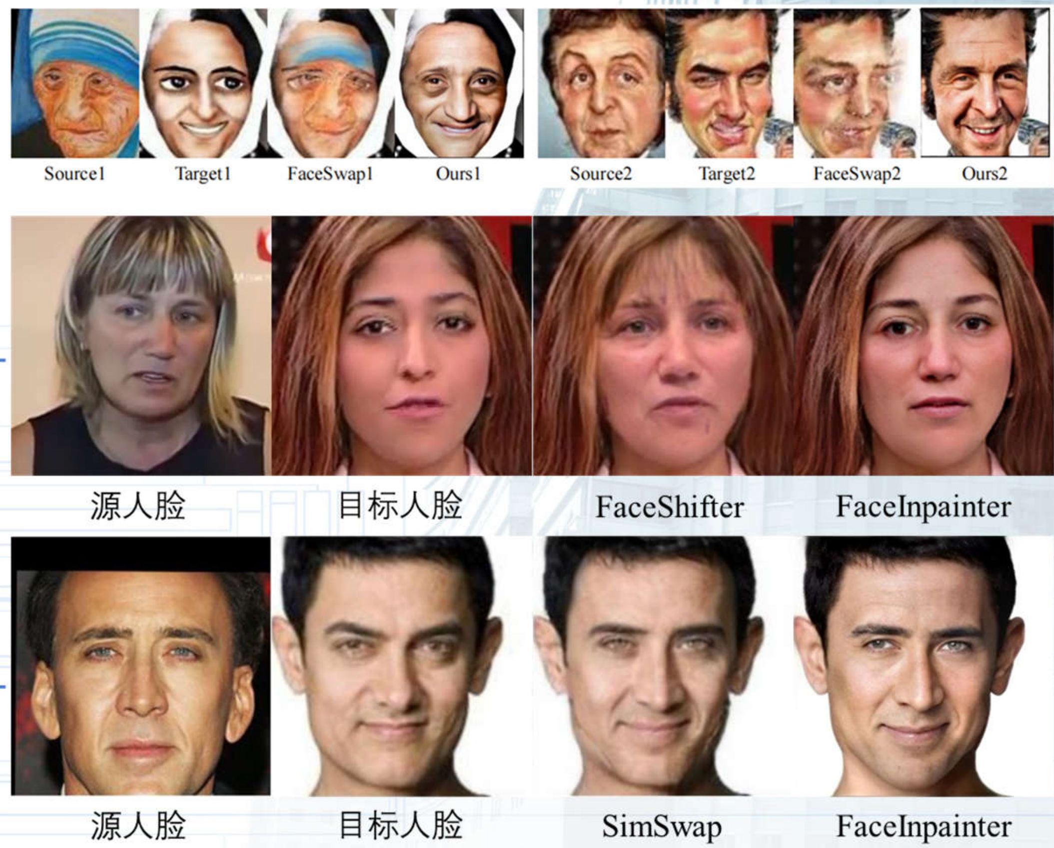
跨域换脸

线稿域换脸



卡通画换脸

真实人脸域换脸

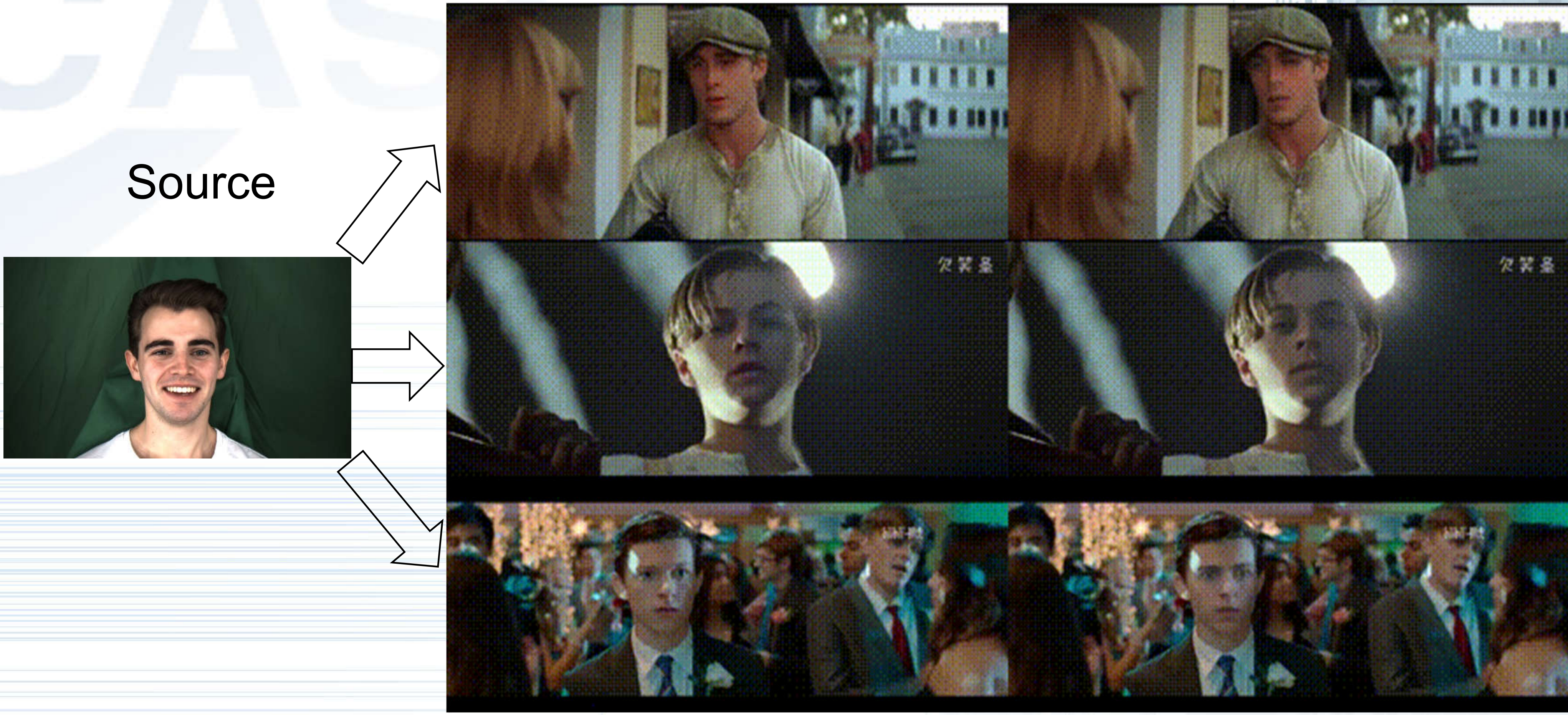


Jia Li, Zhaoyang Li, Jie Cao, Xingguang Song, Ran He. FaceInpainter: High Fidelity Face Adaptation to Heterogeneous Domains. CVPR 2021: 5089-5098

High-quality Identity Swap

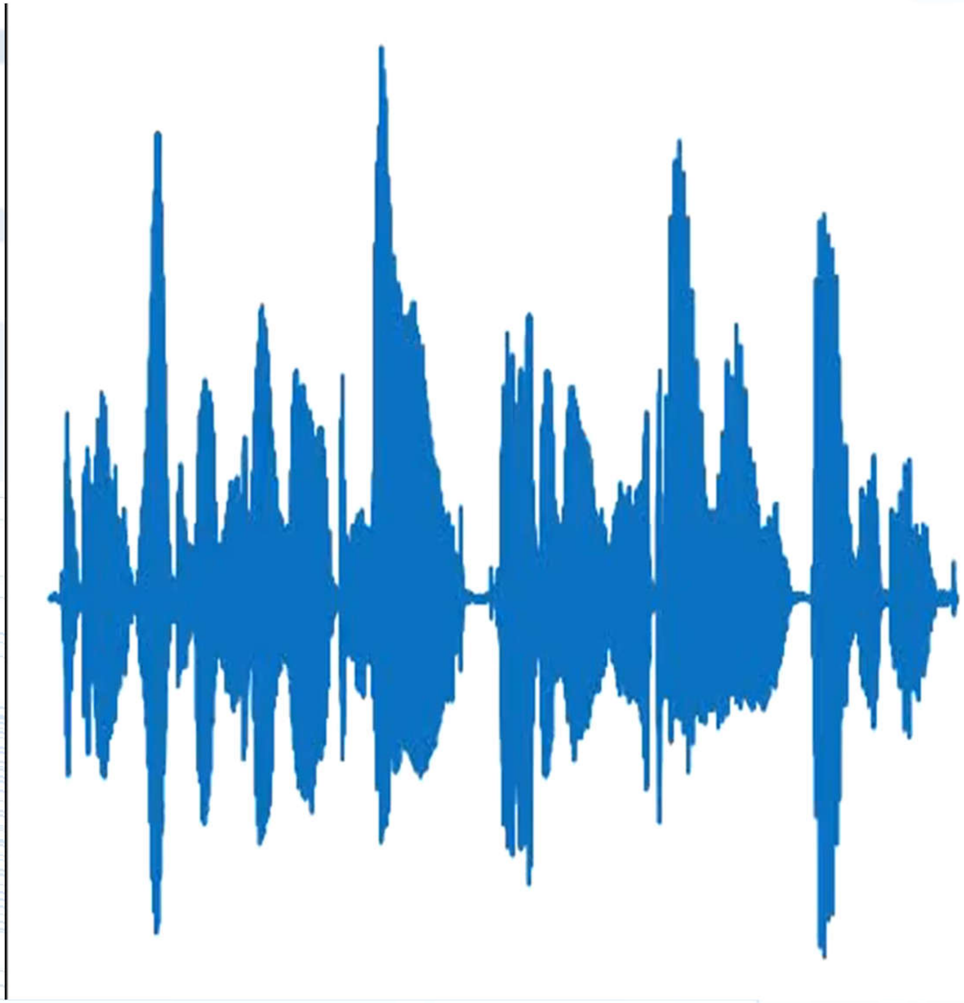
Target Swapped

Source



Hao Zhu, Ran He et al. AOT: Appearance Optimal Transport Model for Face Swapping. NeurIPS 2020.

Talking Face Video Generation



Kaisiyuan Wang, Ran He, et al. MEAD: A Large-scale Audio-visual Dataset for Emotional Talking Face Generation. ECCV, 2020.
Hao Zhu, Ran He, et al. Arbitrary Talking Face Generation via Attentional Audio-Visual Coherence Learning. IJCAI, 2020.

Possible features for fake detection

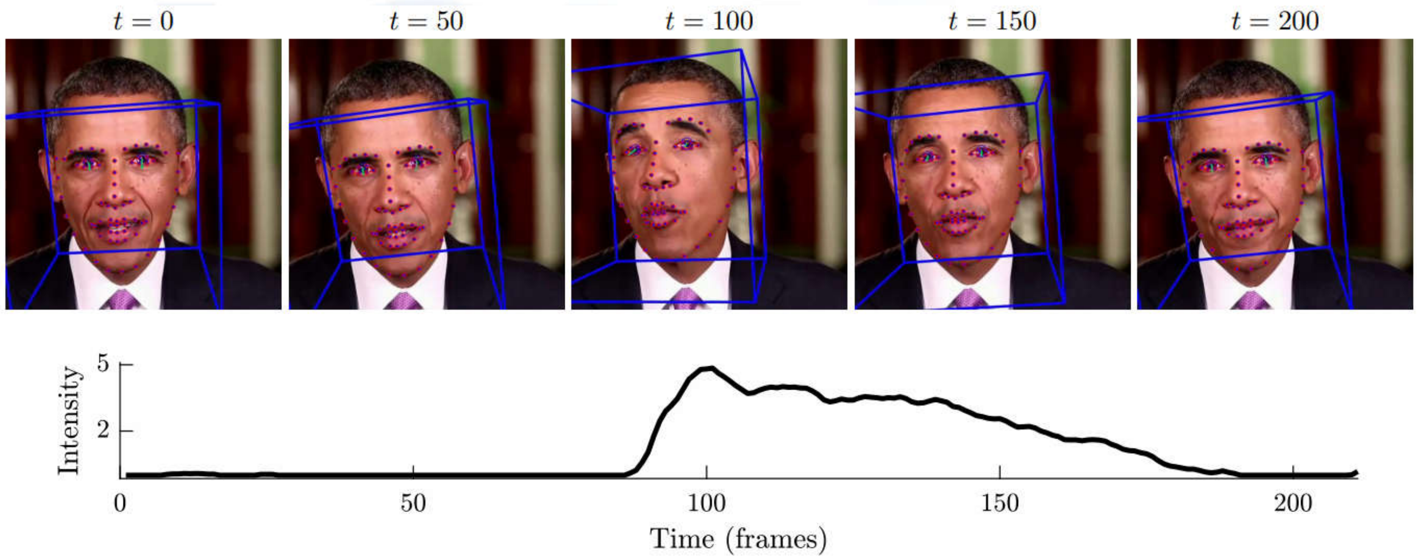
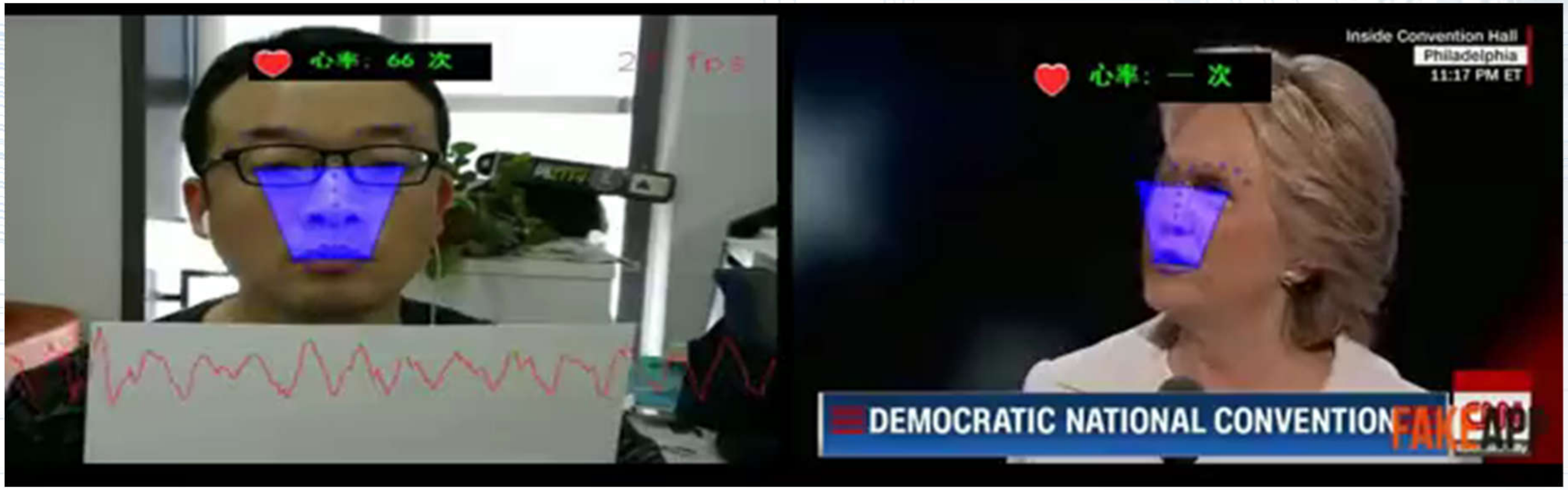


Figure 1. Shown above are five equally spaced frames from a 250-frame clip annotated with the results of OpenFace tracking. Shown below is the intensity of one action unit AU01 (eye brow lift) measured over this video clip.

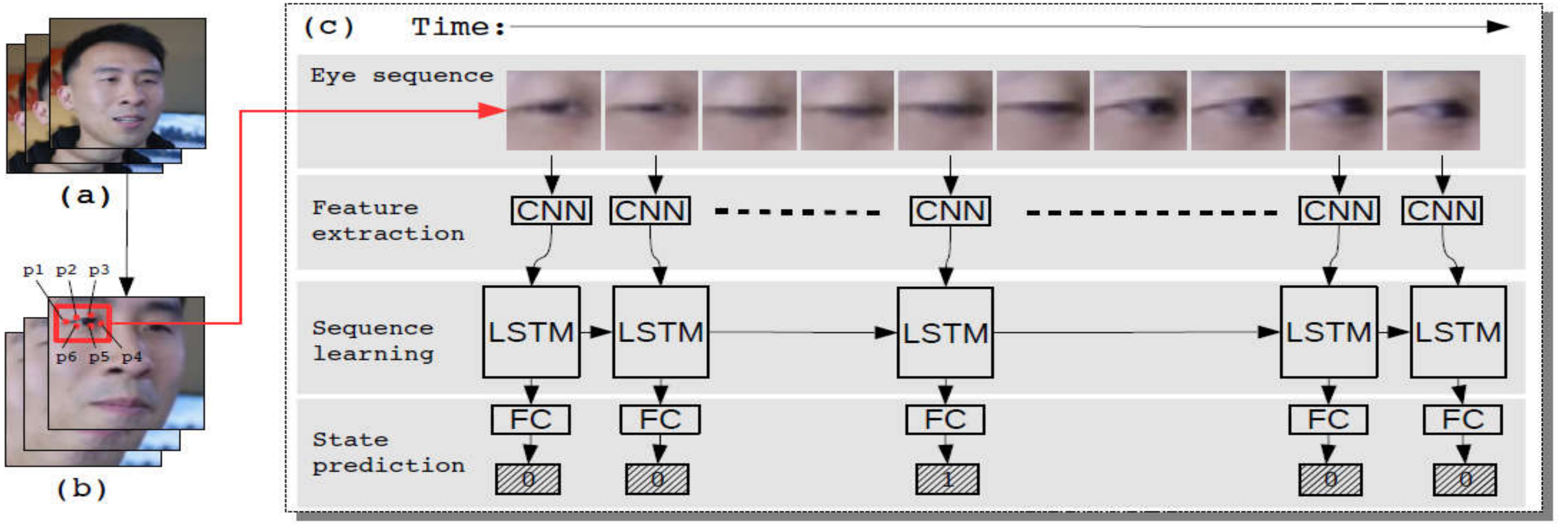
Facial Behavior Modeling



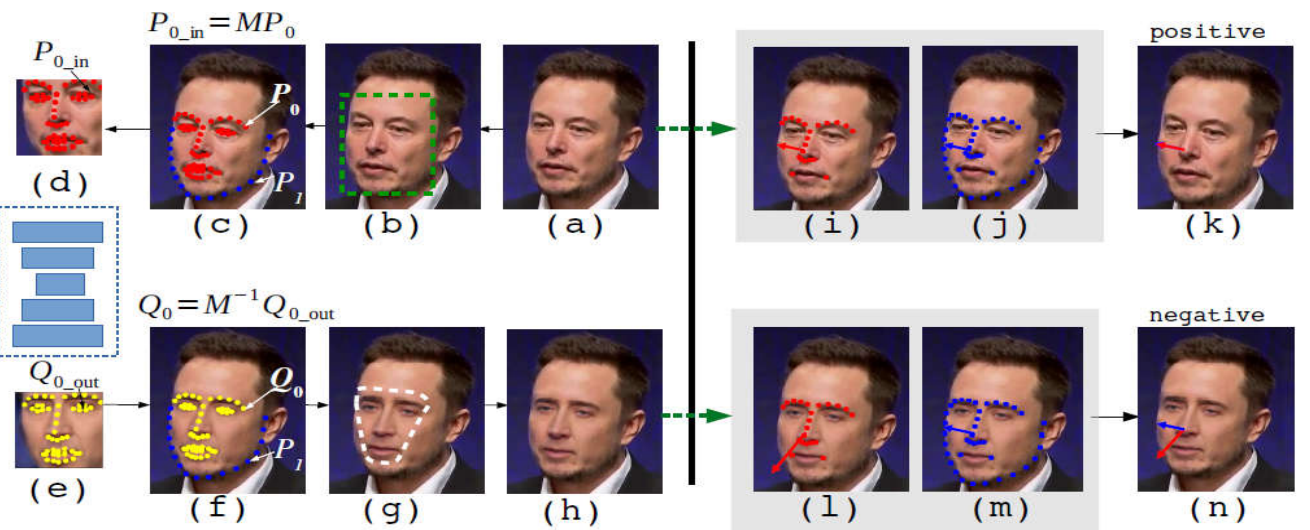
Physiological Indicator

Agarwal, Shruti, Hany Farid, et. al. "Protecting World Leaders Against Deep Fakes." CVPR 2019.

Possible features for fake detection



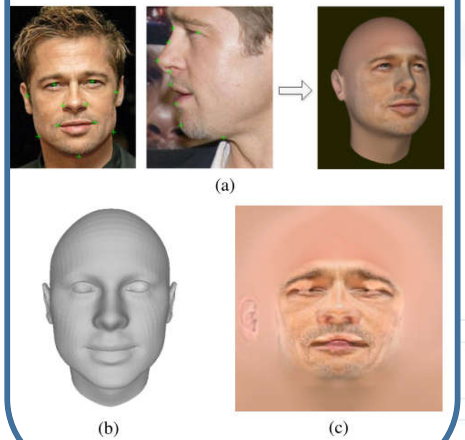
Eye-blinking clue

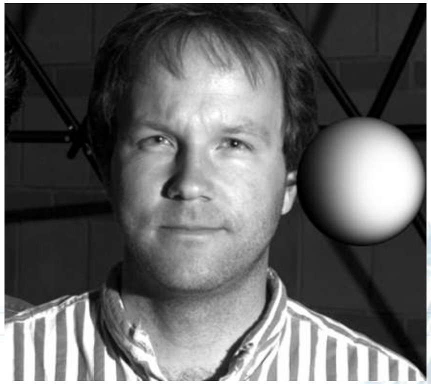


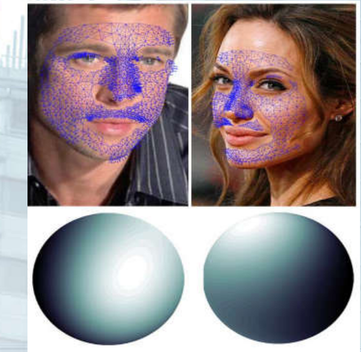
Head pose consistency

Inconsistency of lighting conditions

3D Face Fitting → Face Model Alignment → Optimized Lighting Estimation → Inconsistency Measure



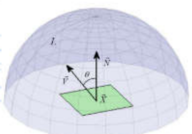


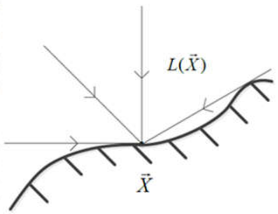


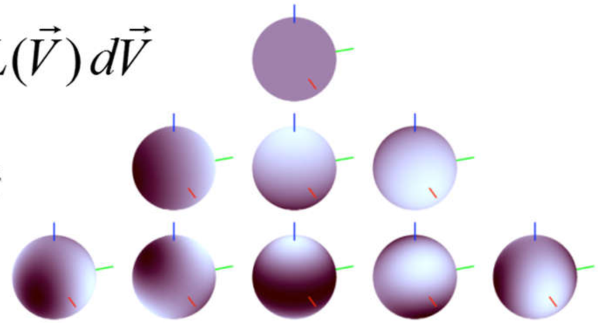
$$Dist(L_1, L_2) > T\ ?$$

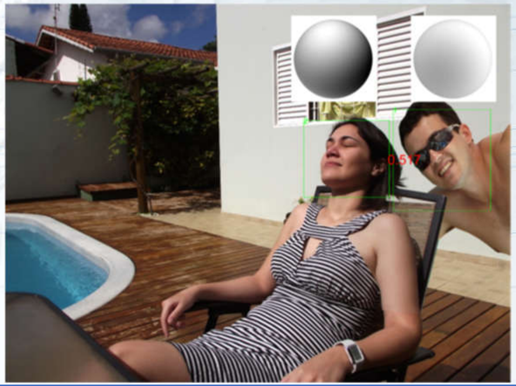
RELAXATION !

$$I(\vec{x}) = \int_{\Omega} \underbrace{\rho(\vec{X})}_{\text{Texture}} \underbrace{G(\vec{X},\vec{V})}_{\text{Occlusion}} R(\vec{V},\vec{N}(\vec{X})) L(\vec{V})\, d\vec{V}$$





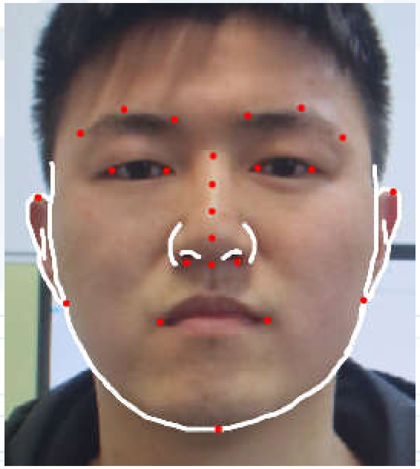




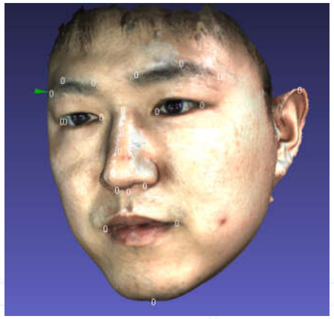
- Bo Peng, **Wei Wang**, Jing Dong, and Tieniu Tan, “Optimized 3D Lighting Environment Estimation for Image Forgery Detection,” IEEE Transactions on Information Forensics and Security, 2016.
- Bo Peng, **Wei Wang**, Jing Dong, and Tieniu Tan, “Automatic detection of 3D lighting inconsistencies via a facial landmark based morphable model,” IEEE International Conference on Image Processing (ICIP), 2016, pp. 3932-3936.
- Bo Peng, **Wei Wang**, Jing Dong, and Tieniu Tan, “Improved 3D lighting environment estimation for image forgery detection,” IEEE International Workshop on Information Forensics and Security (WIFS), 2015, pp. 1-6.

Invalidation of projective geometry laws

S1: Landmark & Contour Obervation



S2: Estimation by Landmark Correspondence



- Minimize projection error of landmarks
- Gold Standard Method

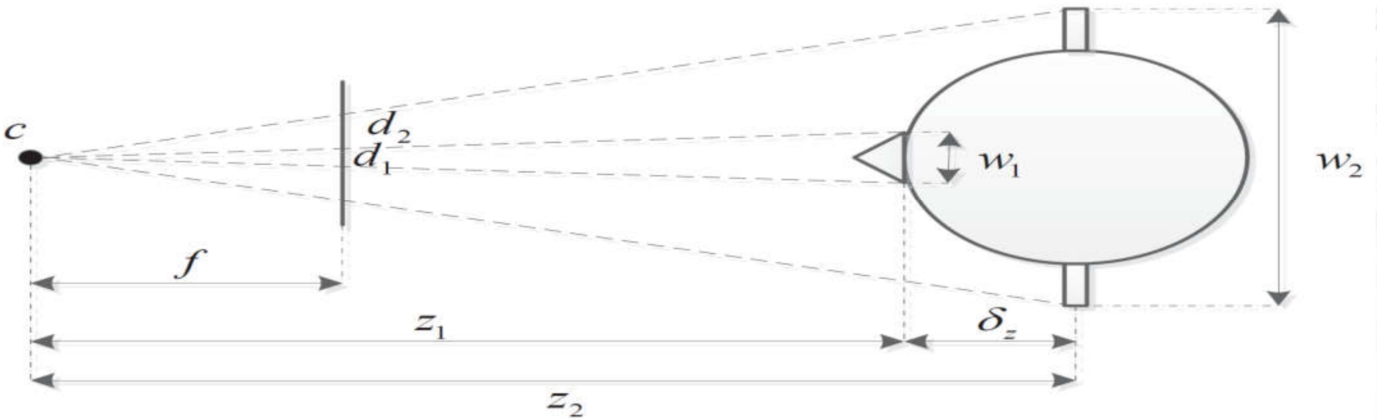
S3: Fine Tune using Contour

- Minimize projection error of contours
- ICP algorithm

S4: Consistency between Intrinsic paras

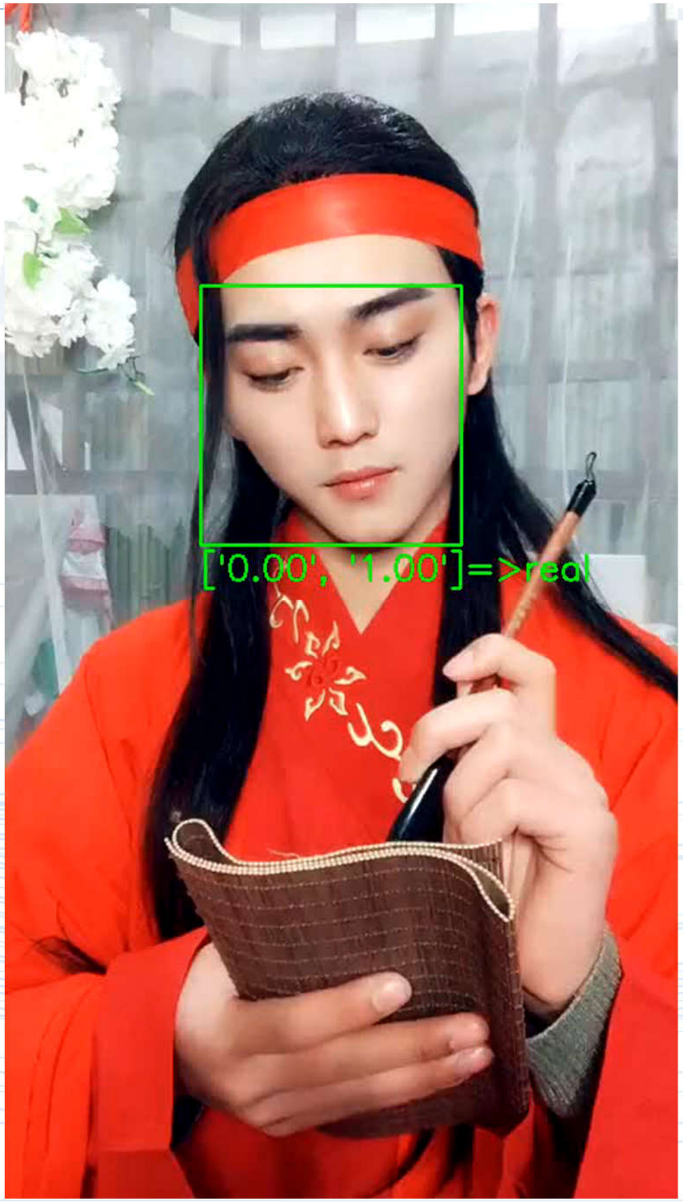
$$Dist(\theta, \{\hat{\theta}\}) > T\,?$$

- Random perturbation
- Mahalanobis distance



Bo Peng, Wei Wang, Jing Dong, and Tieniu Tan, "Position Determines Perspective: Investigating Perspective Distortion for Image Forensics of Faces," CVPR Workshop on Media Forensics 2017.

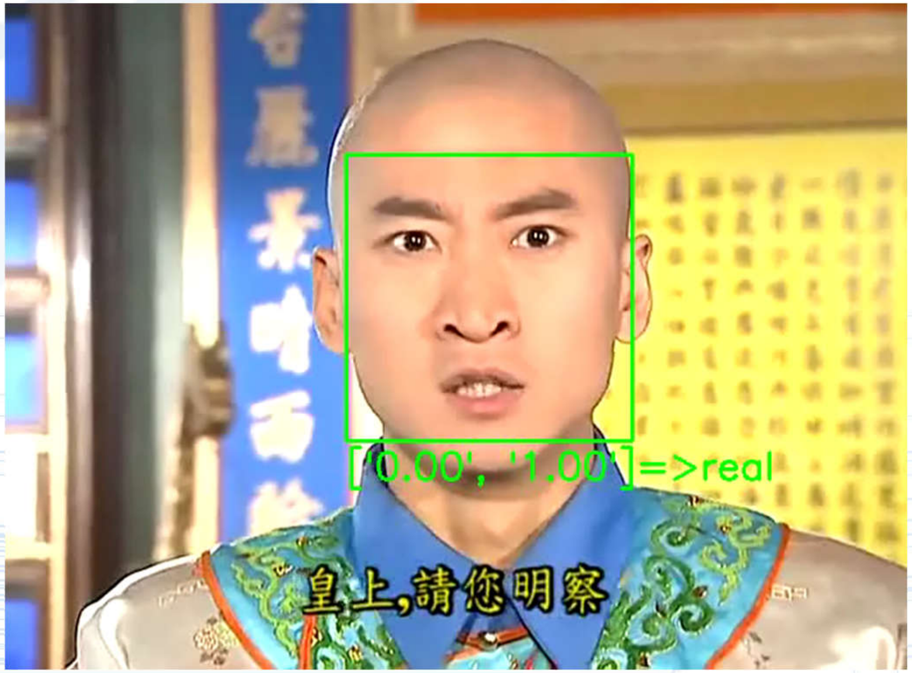
Fake Detection of Face Videos Generated by ZAO



Genuine video

ZAO video

Fake Detection of Face Videos Generated by ZAO



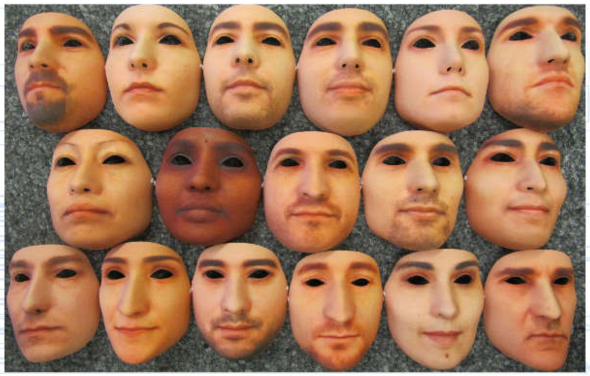
Genuine video

ZAO video

Open Problems of Face Recognition

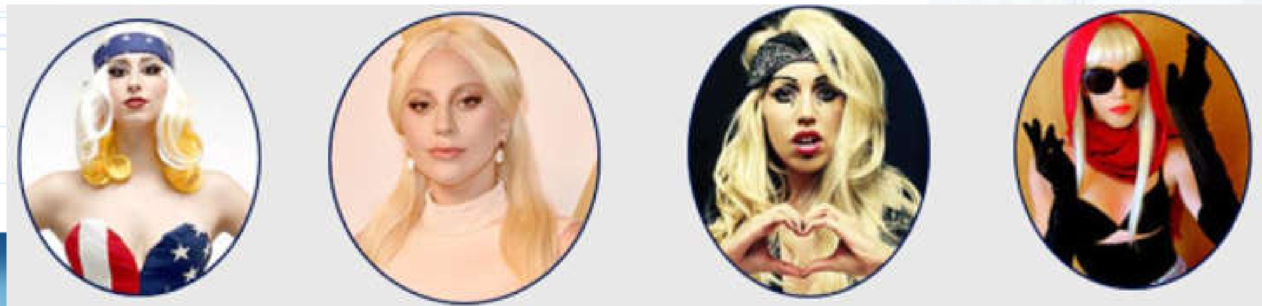
PIE (Pose, Illumination, Expression)

Face recognition in surveillance



Spoof-attack

Face recognition of twins



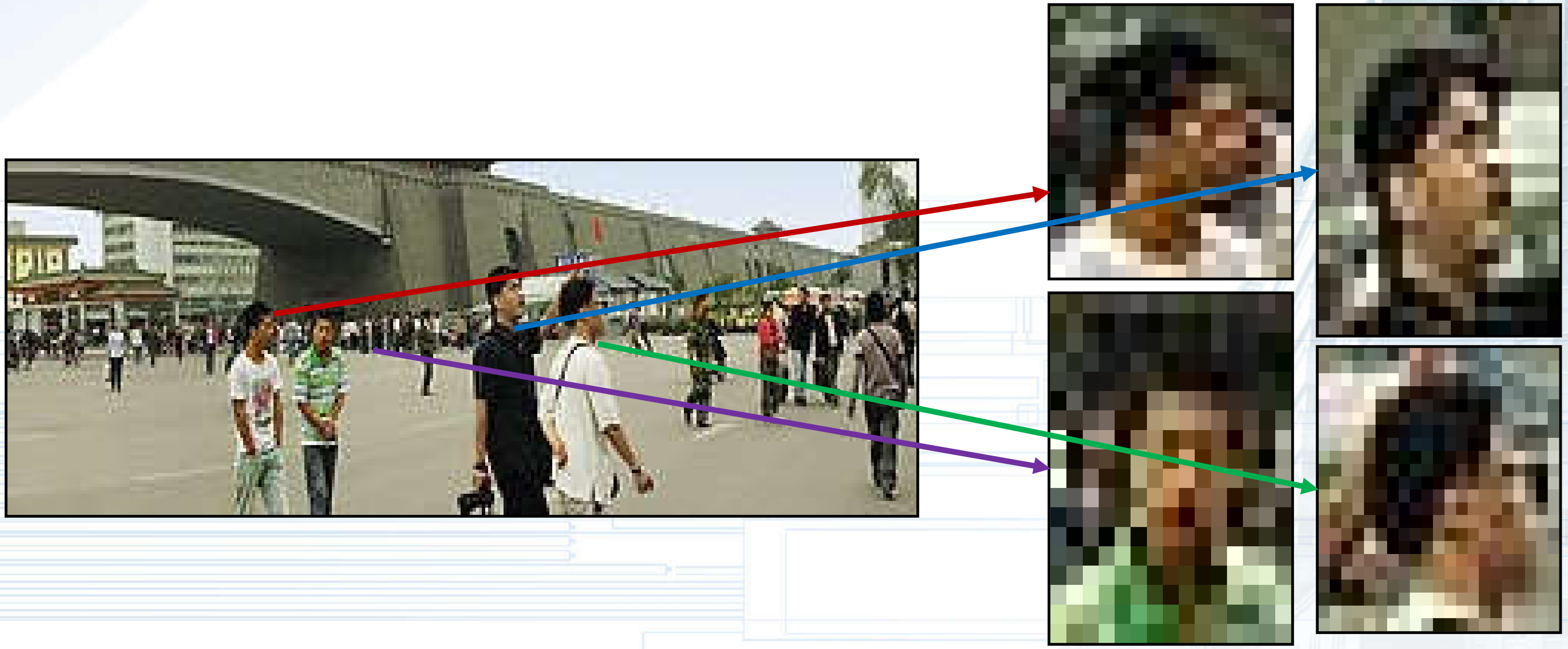
Facial disguise

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

Advantages of gait recognition

As a biometric, gait is still available at a distance when other biometrics are obscured or at too low resolution.



Advantages: robust against imaging distance, resolution, view, illumination

History of gait recognition

First gait biometrics paper - Cunado, Nixon and Carter (AVBPA 1997) - 90% CCR

DARPA Program: Human ID at a distance

Learning Representative Deep Features for Image Set Analysis， TMM

Cross-view gait based human identification with deep CNNs， TPAMI

GEINET: view-invariant gait recognition, ICB

1997

2000

2015

2016

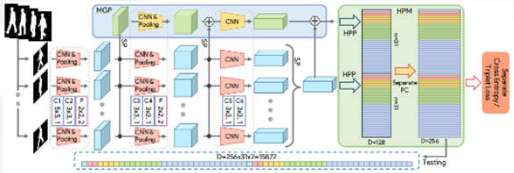
Design hand-crafted features for gait recognition

Deep learning for gait recognition

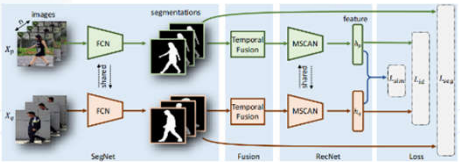
Recent Progress of Gait Recognition

Algorithms

Appearance-based



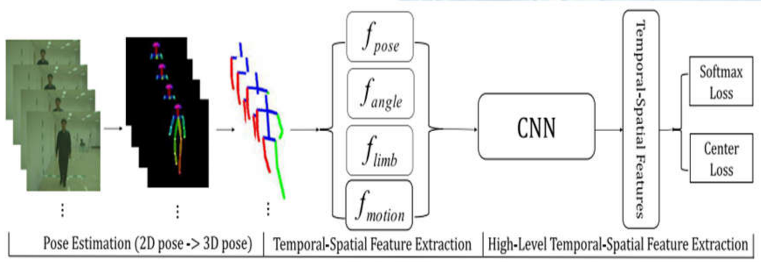
GaitSet



GaitNet

Model-based

PoseGait



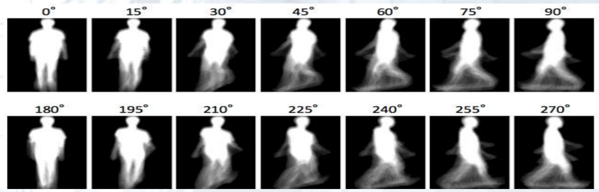
Database

CASIA-B (cross-view)

The first cross-view and cross-dressing database in the world: 124 people, 11 views per person, covering backpack and clothing changes

CASIA-E (the biggest ever)

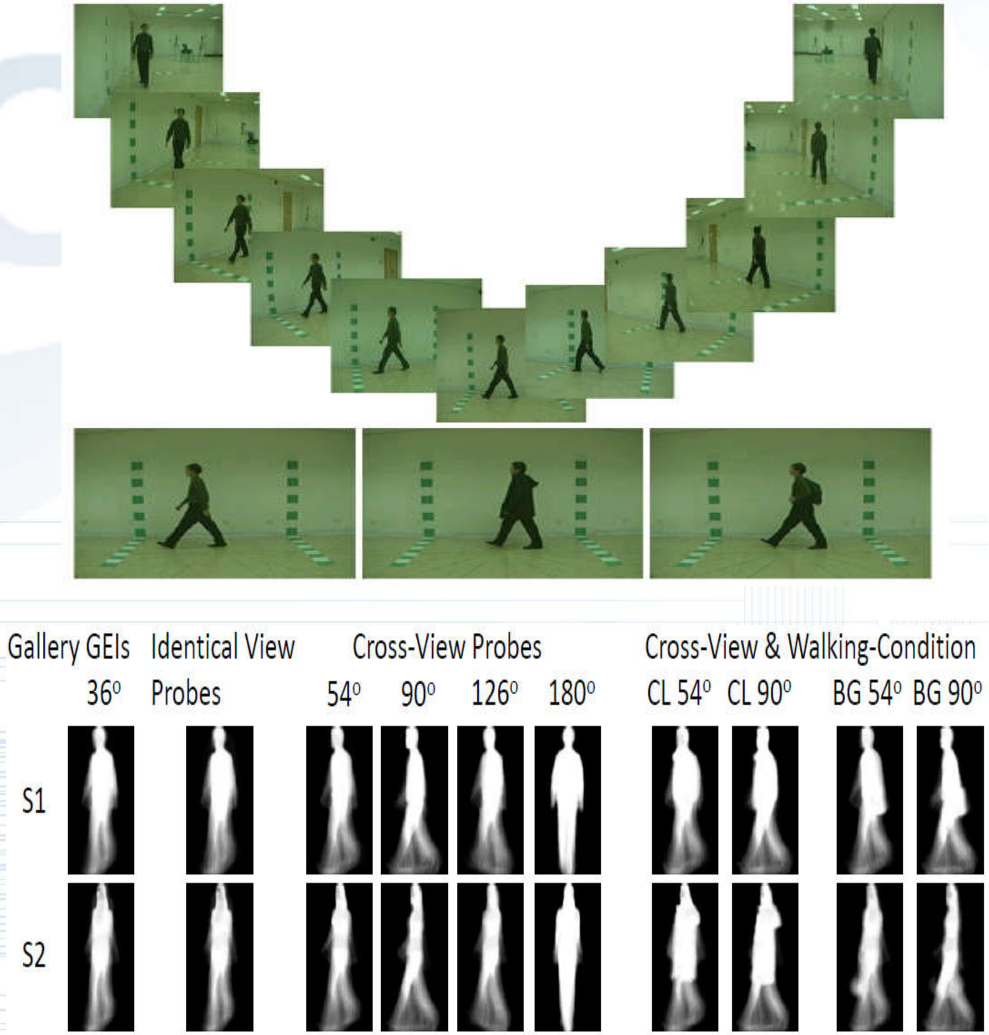
■ more than 1000 persons, nearly 1 million video clips
■3 kinds of clothing、3 kinds of scenes、2 kinds of walking patterns
■13 horizontal views, 2 vertical views

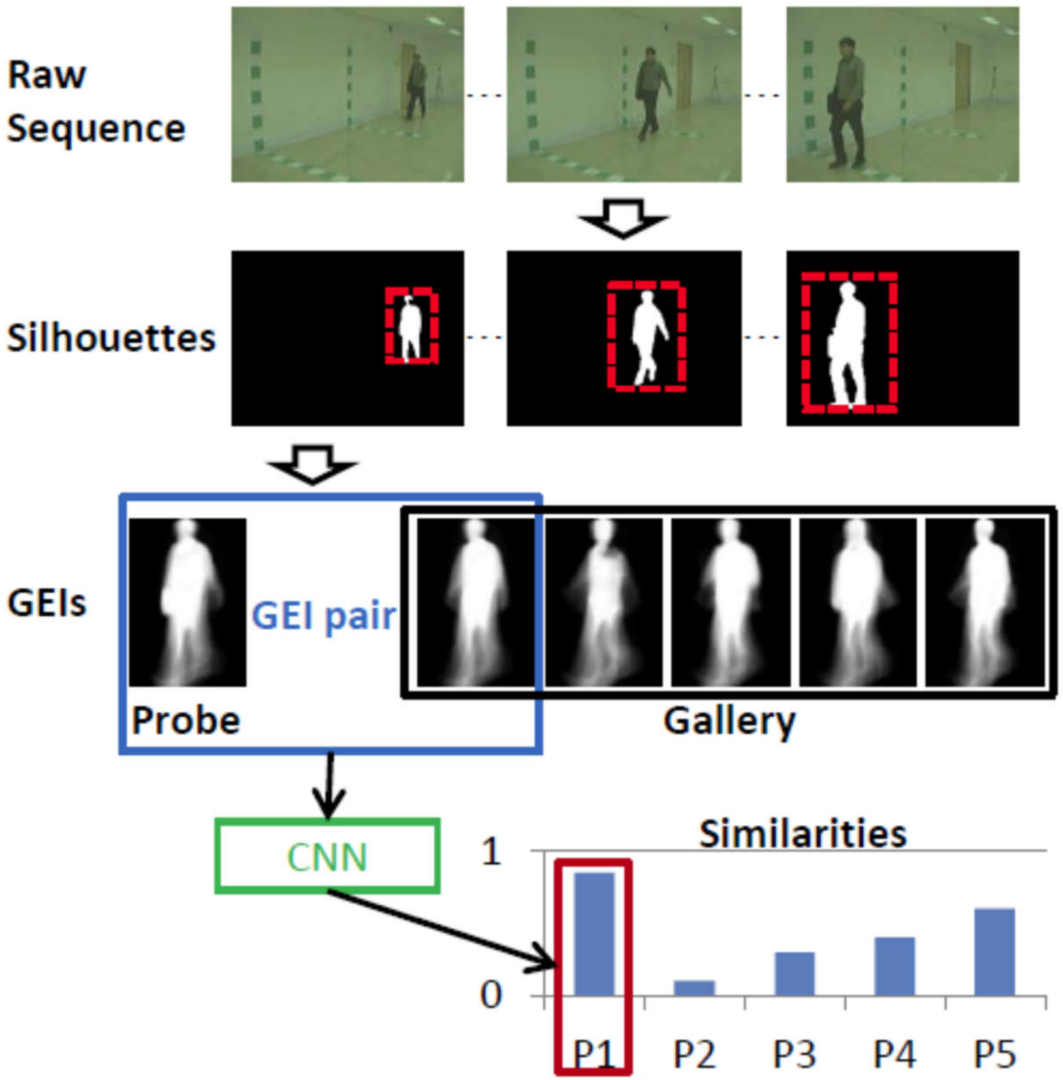


OU-MVLP

Multi-View Large Population Dataset

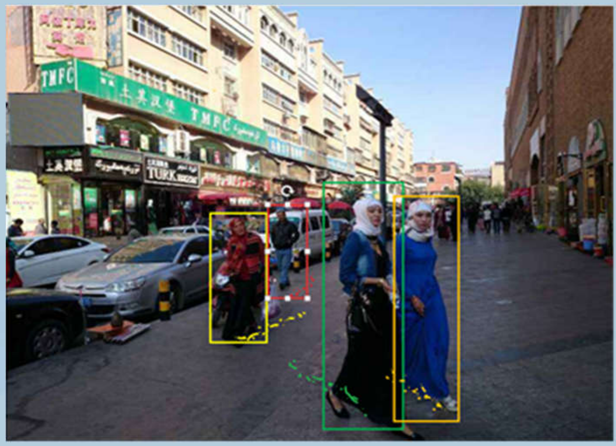
Multi-view Gait Recognition



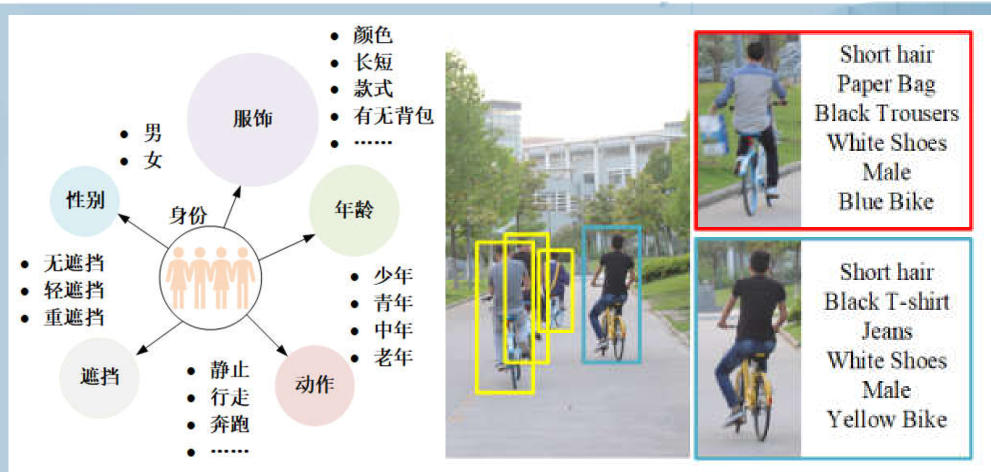


Zifeng Wu, Yongzhen Huang, Liang Wang, Xiaogang Wang, and Tieniu Tan, A comprehensive study on cross-view gait based human identification with deep CNNs, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2017.

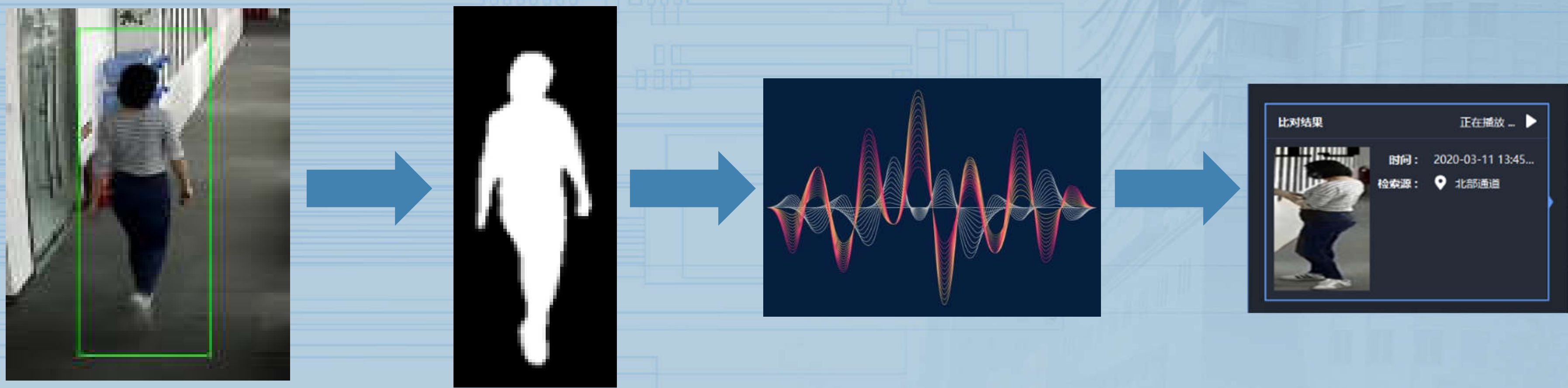
Core Techniques of Gait Recognition



Multi-object cross-view gait recognition



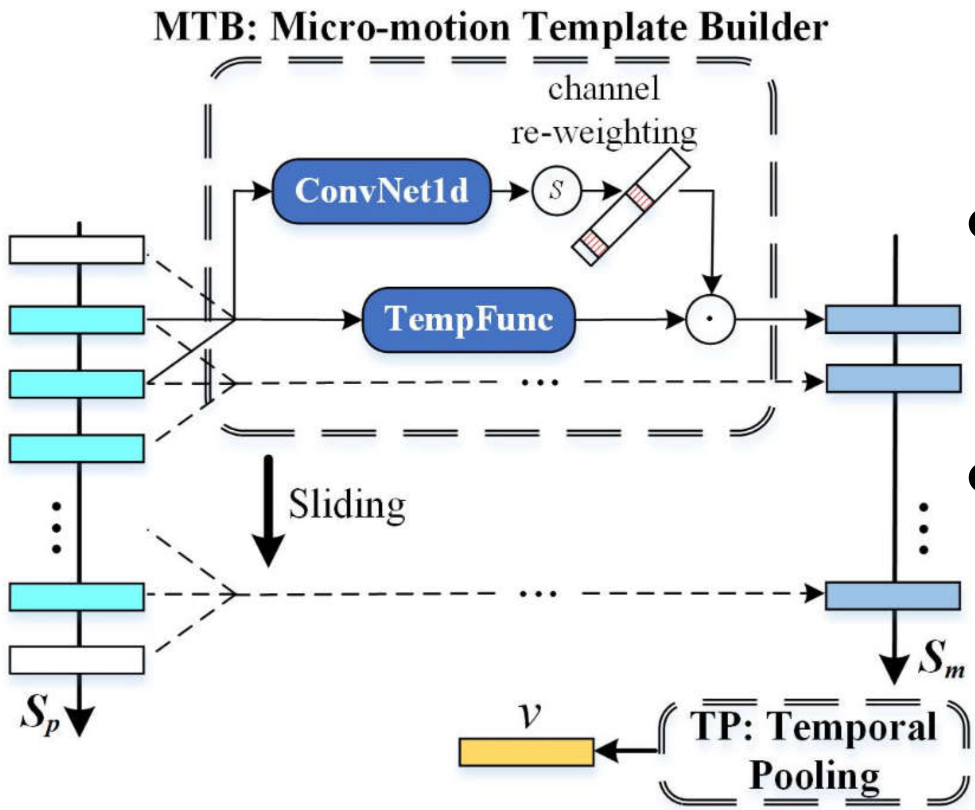
Gait attribute recognition, classification and tracking



Gait Detection, Gait Segmentation and Gait Recognition Integrated Technology (end-to-end)

MTB: Micro-motion Template Builder

Spatio-temporal attention mechanism design



- Short-range spatio-temporal representations (***micro-motion features***) are the most discriminative features of periodic gaits
- A **micro-motion capture module** maps part of the feature vectors of each frame to the micro motion feature vectors, and successfully improves the recognition performance.

Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He, GaitPart: Temporal Part-Based Model for Gait Recognition, in CVPR 2020

MTB: Micro-motion Template Builder

Experimental Results of CASIA-B

Table 3. Averaged rank-1 accuracies on **CASIA-B**, excluding identical-view cases. CNN-LB:[26], GaitSet[5], GaitNet[30].

Gallery NM#1-4		0° – 180°											mean
Probe		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM #5-6	CNN-LB[26]	82.6	90.3	96.1	94.3	90.1	87.4	89.9	94.0	94.7	91.3	78.5	89.9
	GaitSet[5]	90.8	97.9	**99.4**	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
	GaitNet[30]	91.2	92.0	90.5	95.6	86.9	**92.6**	93.5	96.0	90.9	88.8	89.0	91.6
	GaitPart(ours)	**94.1**	**98.6**	99.3	**98.5**	**94.0**	92.3	**95.9**	**98.4**	**99.2**	**97.8**	**90.4**	**96.2**
BG #1-2	CNN-LB[26]	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
	GaitSet[5]	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	**94.4**	79.0	87.2
	GaitNet[30]	83.0	87.8	88.3	93.3	82.6	74.8	**89.5**	91.0	86.1	81.2	85.6	85.7
	GaitPart(ours)	**89.1**	**94.8**	**96.7**	**95.1**	**88.3**	**94.9**	89.0	**93.5**	**96.1**	93.8	**85.8**	**91.5**
CL #1-2	CNN-LB[26]	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
	GaitSet[5]	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	GaitNet[30]	42.1	58.2	65.1	70.7	68.0	70.6	65.3	69.4	51.5	50.1	36.6	58.9
	GaitPart(ours)	**70.7**	**85.5**	**86.9**	**83.3**	**77.1**	**72.5**	**76.9**	**82.2**	**83.8**	**80.2**	**66.5**	**78.7**

Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He, GaitPart: Temporal Part-Based Model for Gait Recognition, in CVPR 2020

MTB: Micro-motion Template Builder

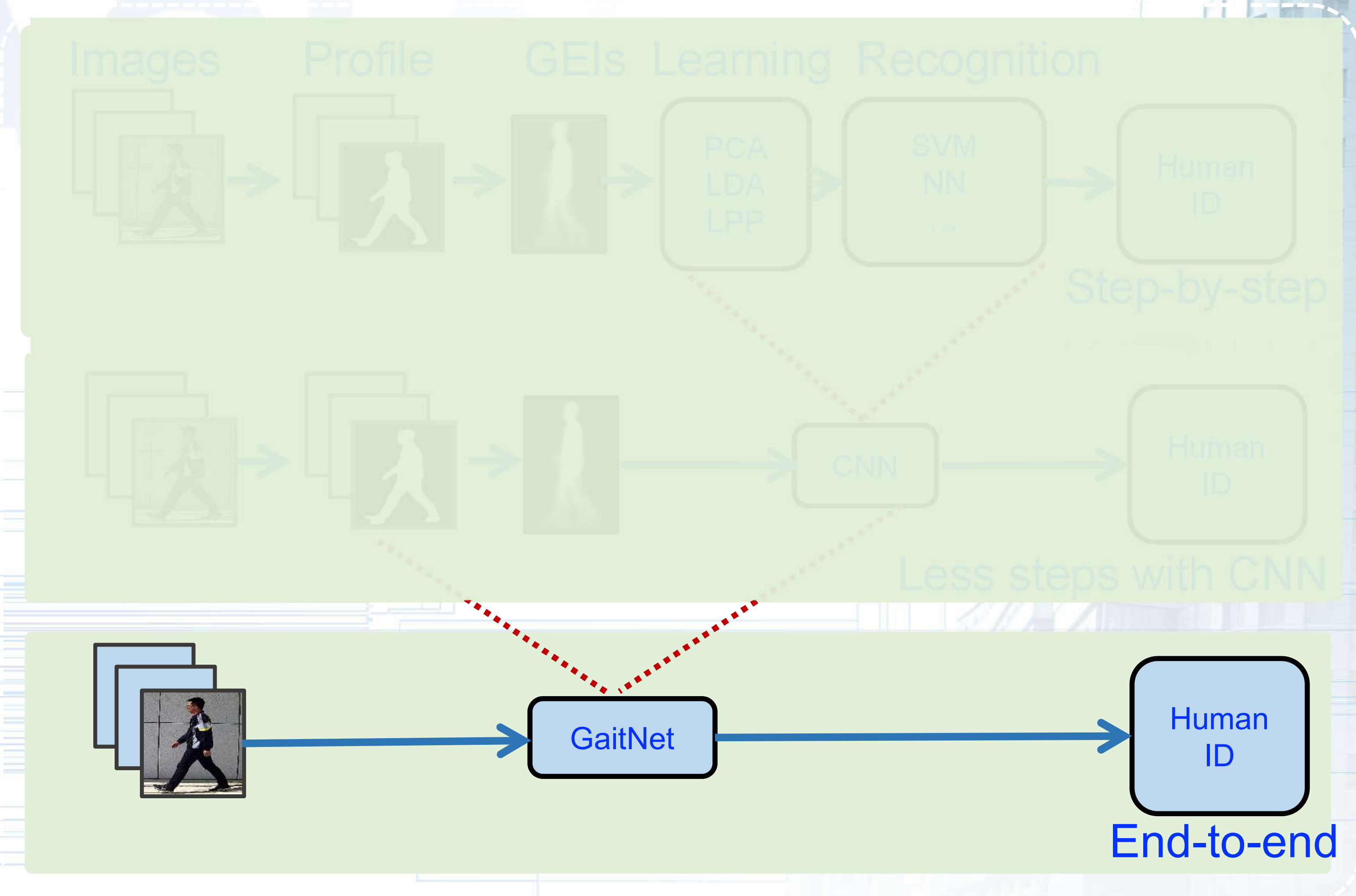
Experimental Results of OU-MVLP

Table 4. Averaged rank-1 accuracies on **OU-MVLP**, excluding identical-view cases. GEINet:[18], GaitSet:[5].

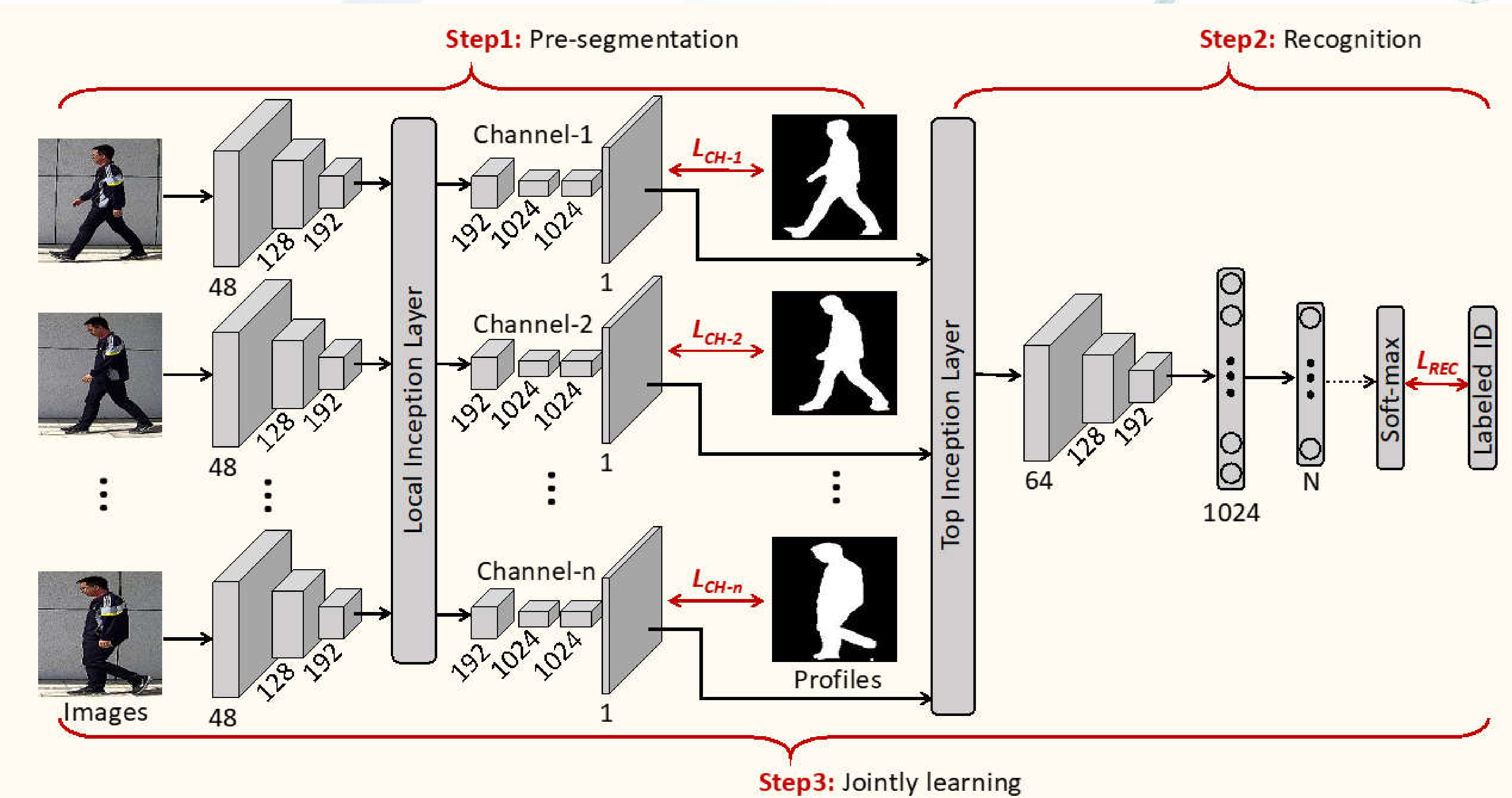
Probe	Gallery All 14 views		
	GEINet[18]	GaitSet[5]	GaitPart(ours)
0°	11.4	79.5	**82.6**
15°	29.1	87.9	**88.9**
30°	41.5	89.9	**90.8**
45°	45.5	90.2	**91.0**
60°	39.5	88.1	**89.7**
75°	41.8	88.7	**89.9**
90°	38.9	87.8	**89.5**
180°	14.9	81.7	**85.2**
195°	33.1	86.7	**88.1**
210°	43.2	89.0	**90.0**
225°	45.6	89.3	**90.1**
240°	39.4	87.2	**89.0**
255°	40.5	87.8	**89.1**
270°	36.3	86.2	**88.2**
mean	35.8	87.1	**88.7**

Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He, GaitPart: Temporal Part-Based Model for Gait Recognition, in CVPR 2020

An end to end gait recognition system



Flowchart of end-to-end gait recognition



C. Song, Y. Huang L. Wang, et al, GaitNet: An End-to-end Network for Video-based Human Identification, PR 2019.

Experiments-Results on Outdoor-Gait

Methods		SCENE-1			SCENE-2			SCENE-3			Mean
		NM	CL	BG	NM	CL	BG	NM	CL	BG	
GEI[9]	PCA	79.71	84.56	86.23	97.83	93.48	96.38	65.22	66.42	72.26	82.45
	LDA	88.41	87.50	86.23	97.10	94.93	97.10	60.87	61.94	71.53	82.85
	LPP	86.96	87.50	89.13	93.48	92.03	97.10	60.87	59.70	76.64	82.60
GEnI[3]	PCA	79.71	78.68	78.26	98.55	92.75	96.38	57.25	51.49	65.69	77.64
	LDA	82.61	86.03	84.78	97.10	92.75	95.65	58.70	57.46	69.34	80.49
	LPP	86.23	86.03	85.51	93.48	95.65	95.65	55.80	58.21	71.53	80.90
GFI[17]	PCA	81.16	83.82	87.68	95.65	91.30	94.93	66.67	58.96	72.26	81.38
	LDA	79.71	68.38	81.88	88.41	86.96	91.30	46.38	43.28	57.66	71.55
	LPP	66.67	69.85	78.26	81.88	86.23	86.96	44.93	50.75	53.29	68.76
CGI[28]	PCA	71.01	72.99	80.44	86.96	89.13	91.30	39.86	41.05	51.83	69.40
	LDA	71.01	68.61	78.99	84.78	88.41	90.58	31.88	39.55	50.37	67.13
	LPP	71.01	68.61	74.64	84.06	84.06	86.96	38.41	44.78	48.91	66.83
GEI-CNN[23]		86.23	90.55	93.48	96.01	95.65	96.74	70.65	70.55	76.81	86.30
GaitNet	Non-Joint	95.59	95.22	**99.26**	98.16	98.9	**100.0**	92.28	92.28	97.06	96.53
	Joint	**100.0**	**100.0**	98.9	**100.0**	**100.0**	99.63	**99.26**	**98.16**	**100.0**	**99.55**

中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

Applications of Gait Recognition

◆ **Public Security**

Gait Retrieval System

Shanghai/Beijing - Sample test

◆ **Commercial Security**

PetroChina - field drilling platform

Gait recognition for white list

◆ **Smart Home**

Midea(Fortune 500) air conditioner

Family member gait recognition

Gait Retrieval - Field Test

△ Crime scene (lateral side, shadow on face)

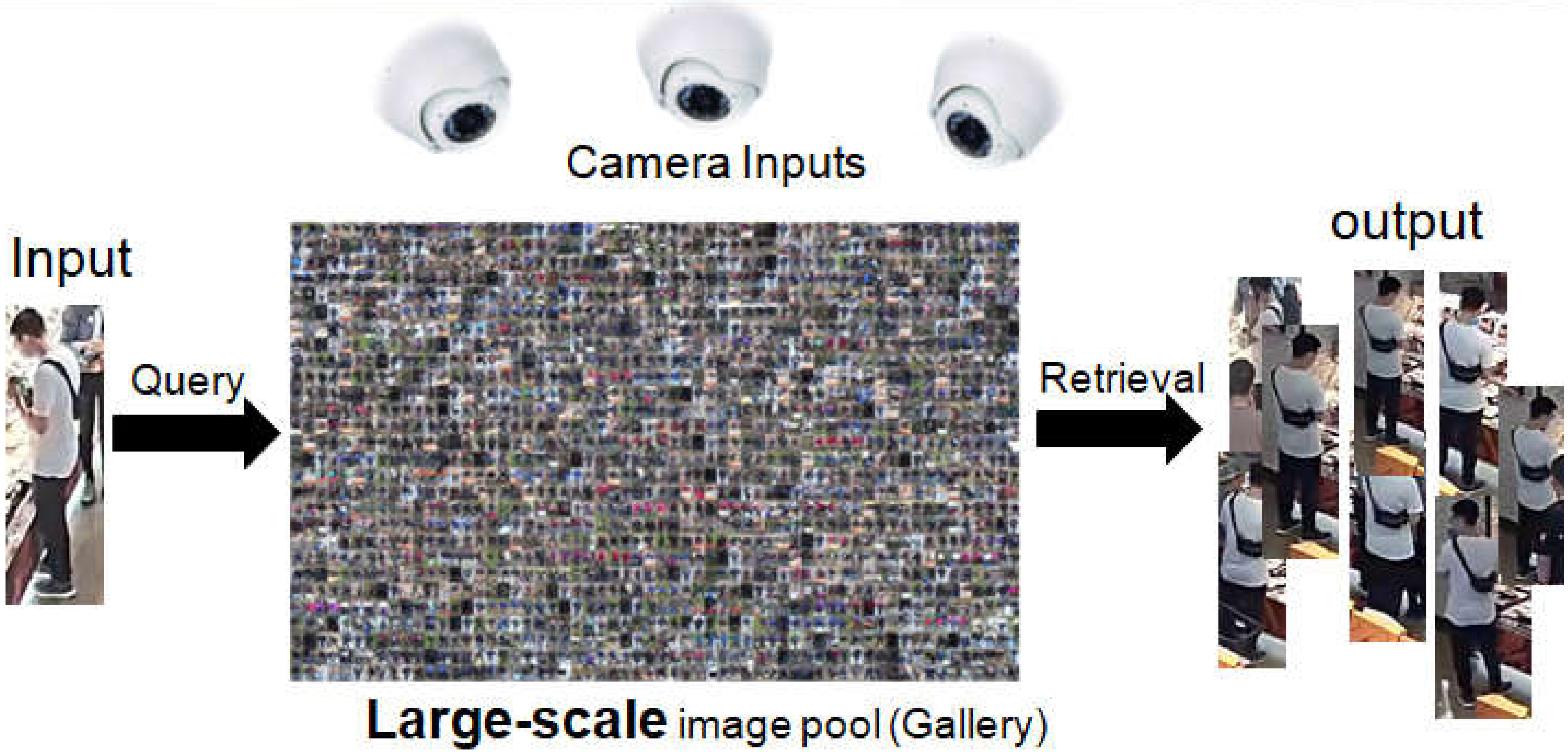
△ Retrieval result: similarity 0.97

Demo of Gait Recognition

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

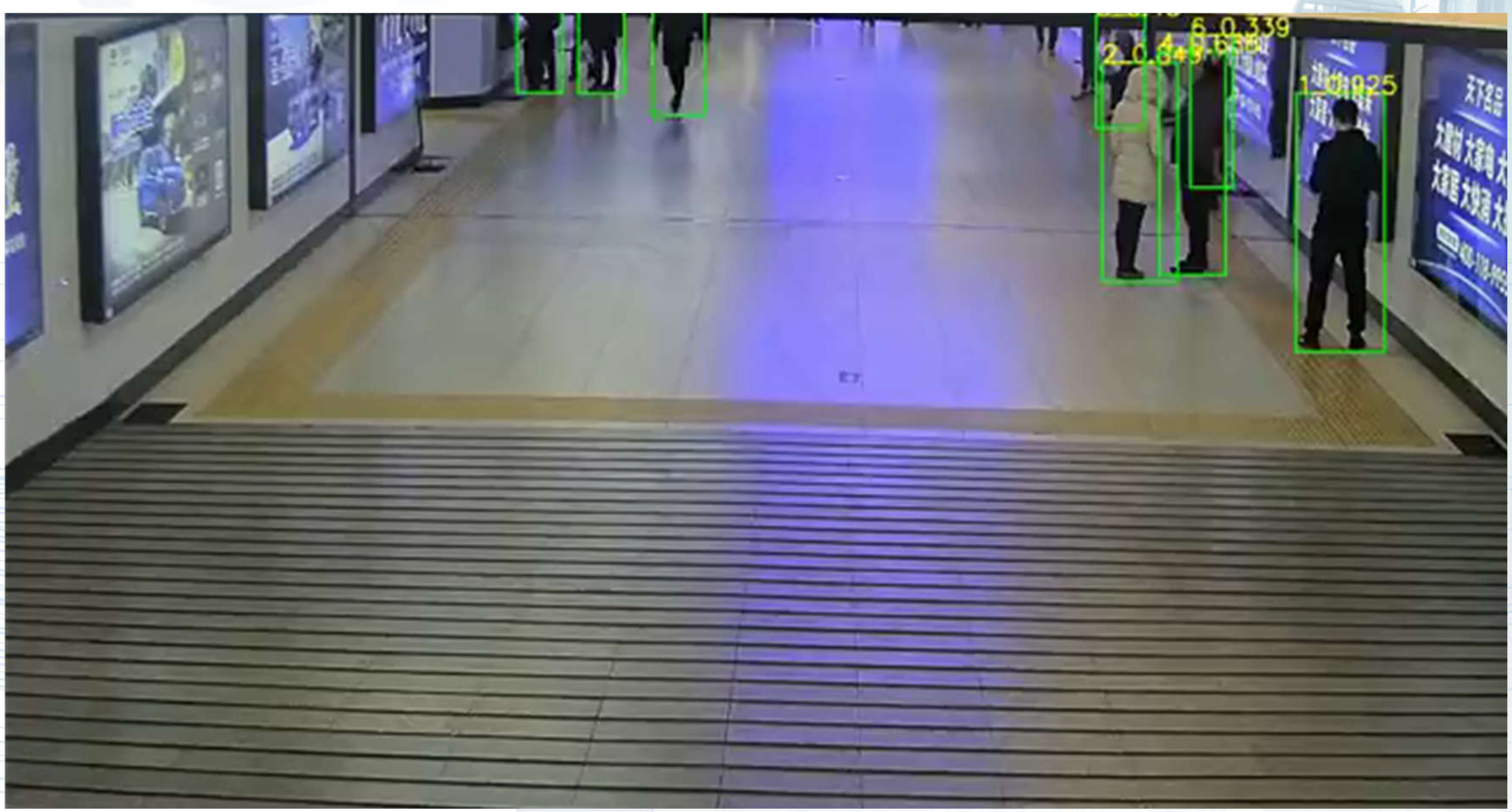
Person Re-identification



Person re-identification aims to match person images from non-overlapping cameras

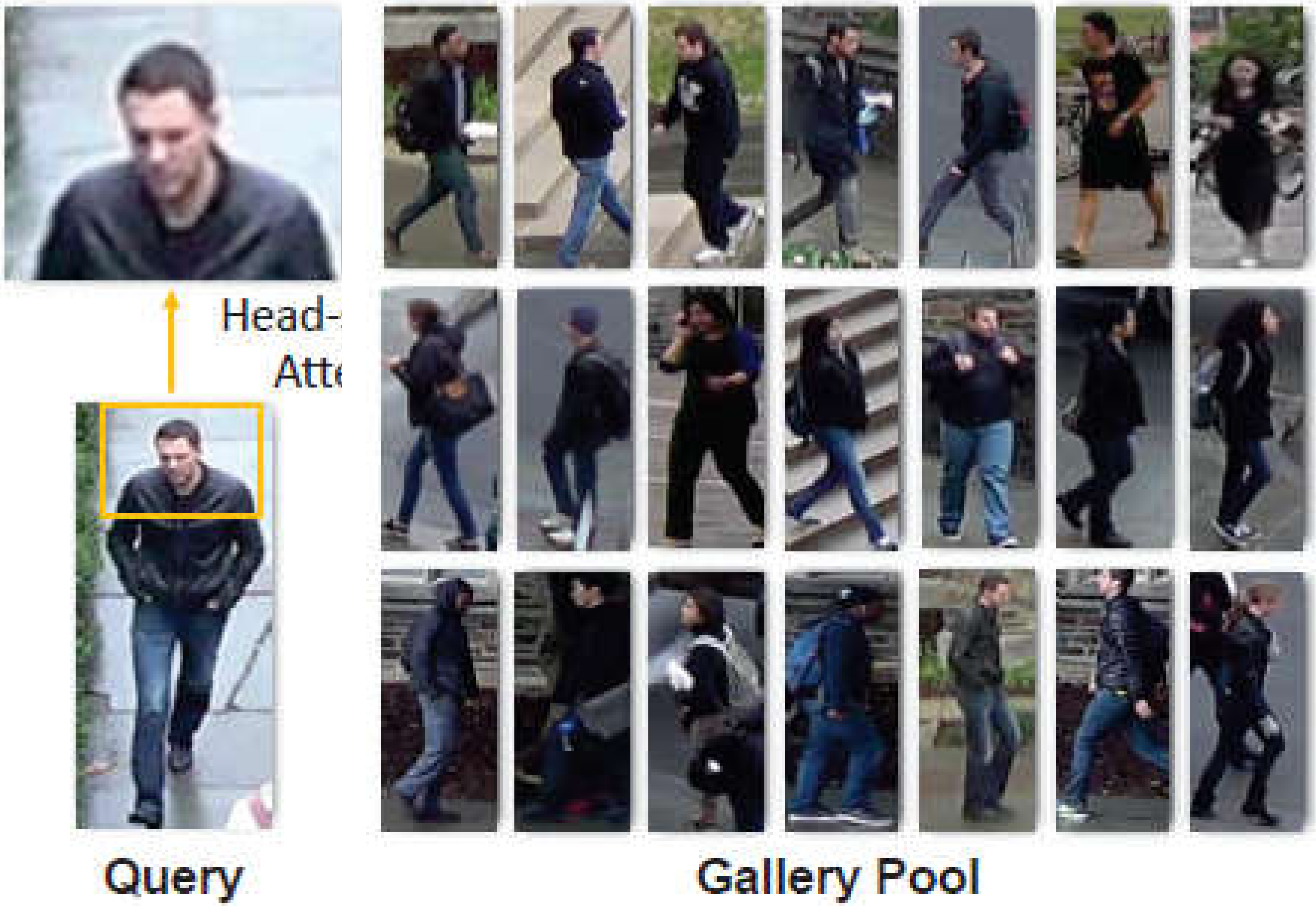
Black Re-ID

Black Re-ID problem: **When people wear black clothes or they are captured by surveillance systems in low light illumination, the attributes of the clothing are severely missing.**



Boqiang Xu, Lingxiao He, Xingyu Liao, Wu Liu, Zhenan Sun, Tao Mei. "Black Re-ID: A Head-shoulder Descriptor for the Challenging Problem of Person Re-Identification." ACM MM. 2020 (Oral).

Black Re-ID

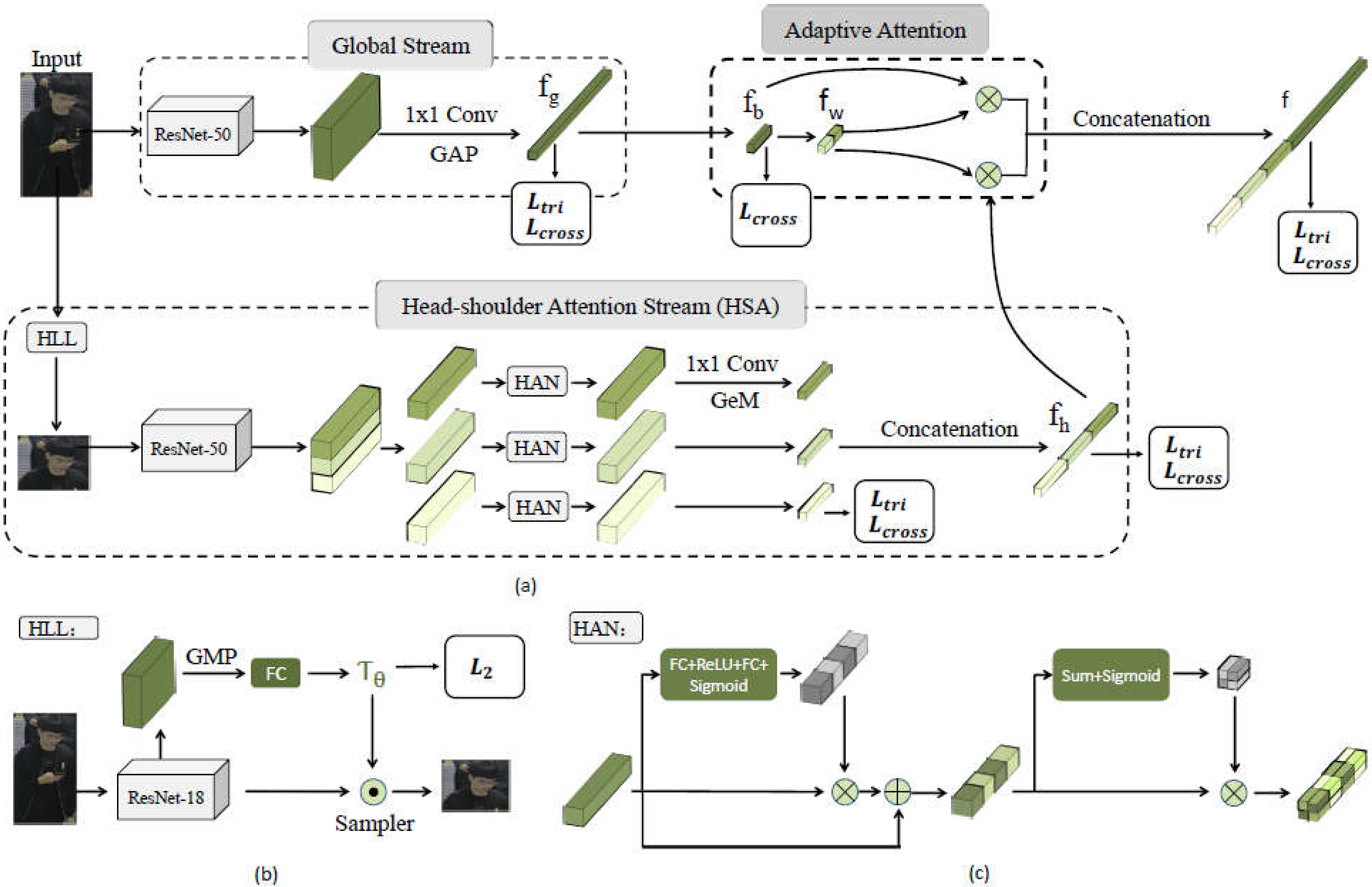


We exploit the head-shoulder feature to assist solving the Black Re-ID problem.

中国科学院
自动化研究所
CASIA
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

Black Re-ID

www.ia.ac.cn



Black Re-ID

Table 2: Quantitative comparison with the state-of-the-art methods in person re-id on Black-reID dataset. Bold number denote the best performance. We denote HAA (ResNet50) and HAA (MGN) by the method selecting ResNet50 and MGN as the backbone respectively.

Method	Black Group		White Group	
	mAP	Rank-1	mAP	Rank-1
ResNet50 [4]	70.8	80.9	75.8	89.5
PCB [29]	73.4	83.2	78.2	90.8
AlignedReID [34]	75.5	83.5	80.5	91.3
MGN [31]	79.1	86.7	85.8	94.3
HAA (ResNet50)	79.0	86.7	84.4	93.5
HAA (MGN)	**83.8**	**91.0**	**88.1**	**95.3**

Challenges of ReID

- **Cloth-Changing Re-ID.** In most Re-ID datasets each person is captured within a short period of time on the same day. As result, each wears the same outfit. However, in practical, we may need to match a person over a much longer period of time, e.g., days or even months. As a result, clothing changes are commonplace.
- **Efficient Model Deployment.** It is important to design efficient and adaptive models to address scalability issue for practical model deployment. How to retrieve fast and how to design a lightweight Re-ID model still need further study.
- **Dynamic Model Updating.** Fixed model is inappropriate for practical dynamically updated surveillance system. To alleviate this issue, dynamic model updating is imperative, either to a new domain/camera or adaptation with newly collected data.

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

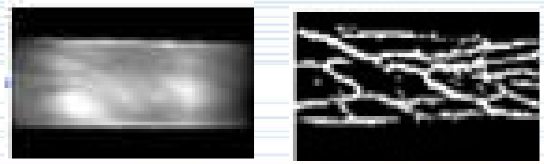
Hand Vein Patterns for Biometric Recognition

Unique, stable and secure biometric patterns underneath the skin surface

Finger vein

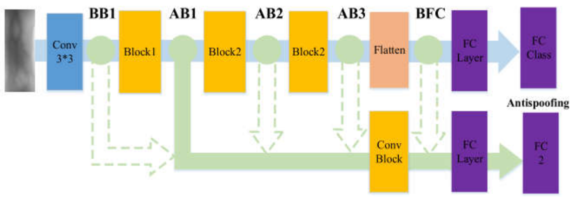
Palm vein

Hand vascular pattern

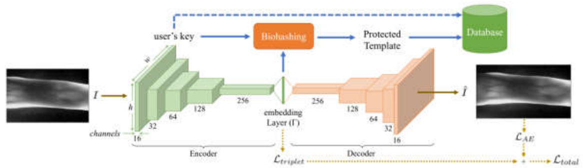


Recent Work on Vein Recognition

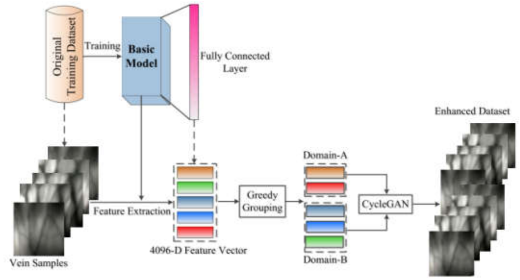
Anti-Spoof Vein Recognition



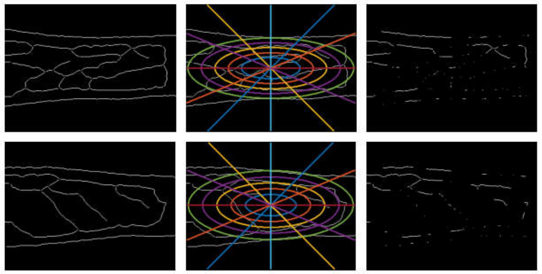
Template Protection Vein Recognition



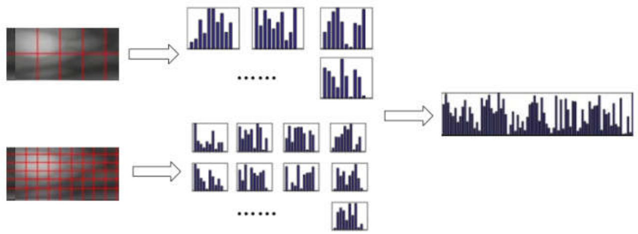
GAN based Vein Image Augmentation



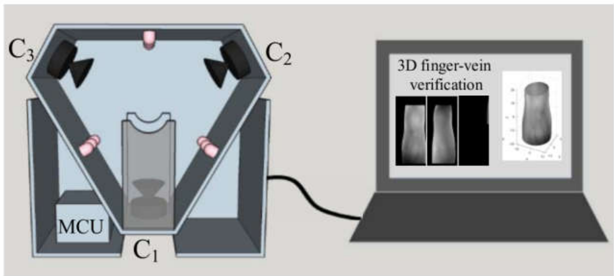
Efficient Vein Recognition



Soft Biometrics for Vein Recognition



3D finger vein Recognition

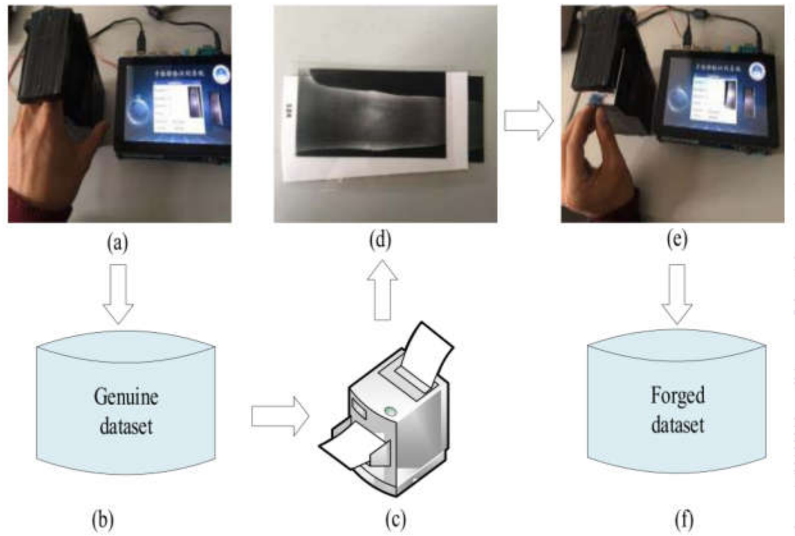


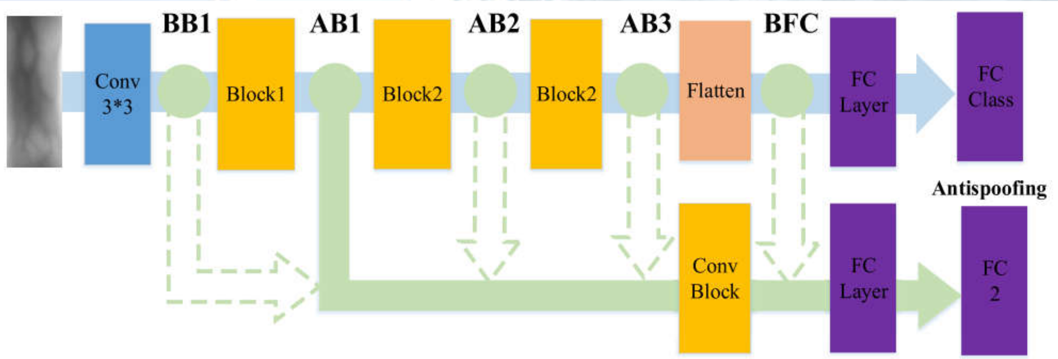
Anti-Spoof Vein Recognition

Anti-spoof vein recognition aims to integrates the recognition task and the anti-spoof task into a unified system.

Two problems：

- Design a Multi-task leaning strategy.
- Balance the performance of both recognition and anti-spoof tasks.





Procedures for forging vein image

Structure of FVRAS-Net

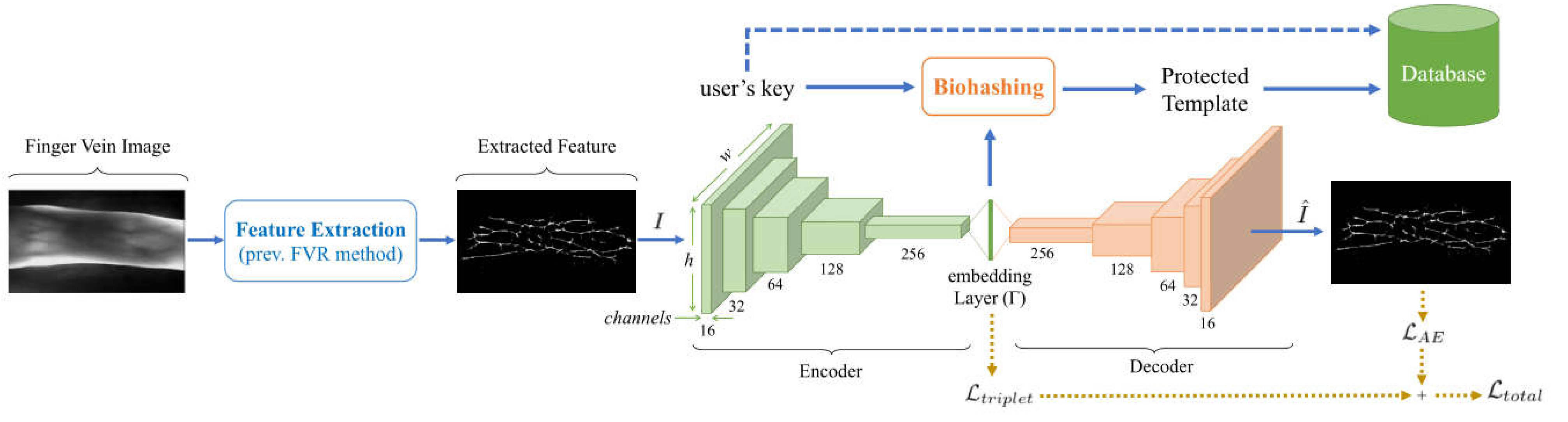
Yang, W., Luo, W., Kang, W., Huang, Z., & Wu, Q.. FVRAS-Net: An Embedded Finger-Vein Recognition and AntiSpoofing System Using a Unified CNN. ***IEEE TIM 2020****.*

Template Protection Vein Recognition

Template protection has been a crucial concern in biometric recognition systems, because biometric trait usually are irreplaceable.

Two problems：

- Consider both raw and pre-processed vein image.
- Consider both the normal and the stolen scenario.



The architecture of the template protection vein recognition system

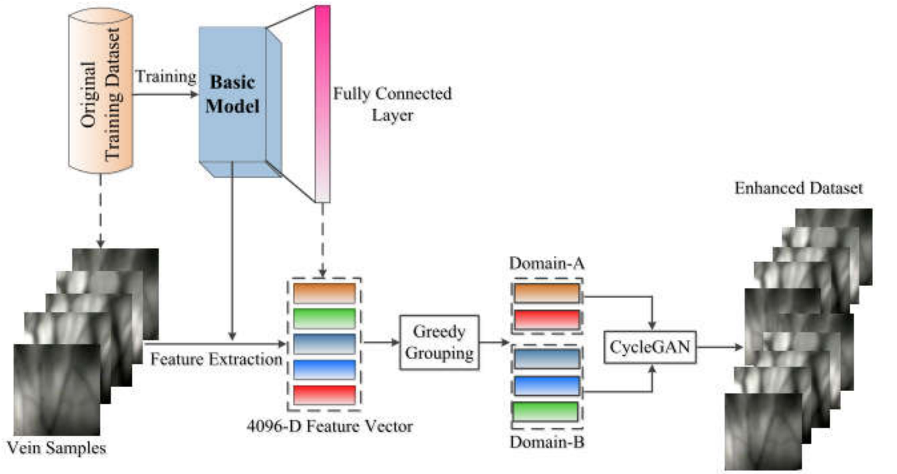
Shahreza, H.O., & Marcel, S. Towards Protecting and Enhancing Vascular Biometric Recognition Methods via Biohashing and Deep Neural Networks. ***IEEE TBIOM 2021****.*

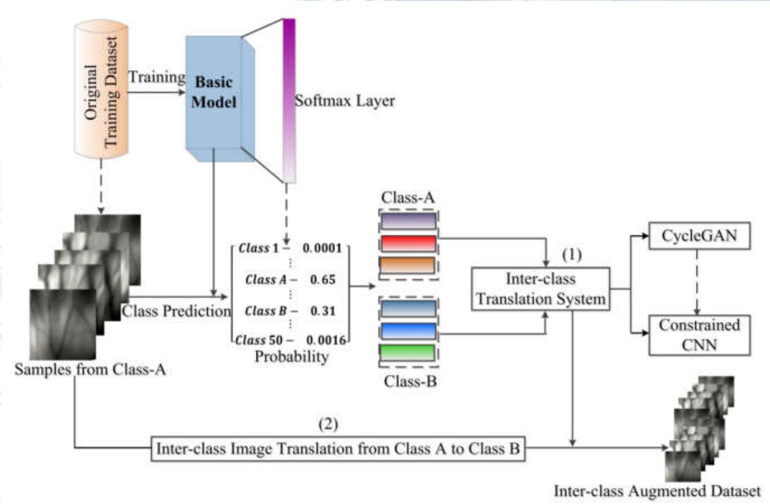
GAN based Vein Image Augmentation

GAN based vein image augmentation aim to alleviate the problem of insufficient training vein data for the application of CNN model.

Key problems：

- Consider both the intra-class augmentation and the inter-class augmentation for vein images.





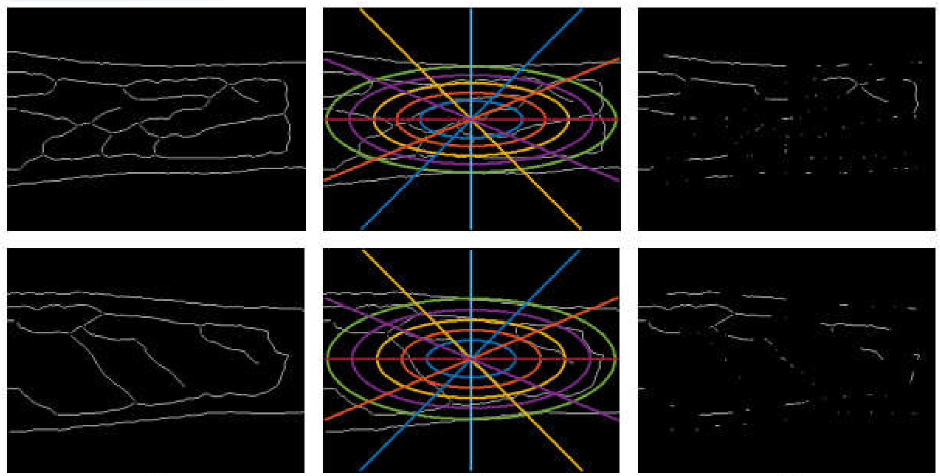
Intra-class vein data augmentation

Inter-class vein data augmentation

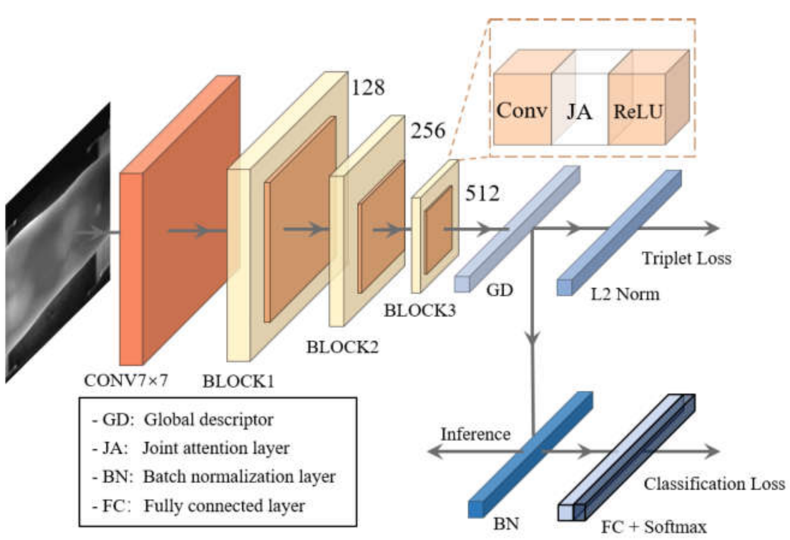
Wang, G., Sun, C., & Sowmya, A. Learning a Compact Vein Discrimination Model With GANerated Samples. ***IEEE TIFS 2020****.*

Efficient Vein Recognition

Efficient vein recognition aim to balance the recognition accuracy and the time cost of the vein recognition system.



Vein code

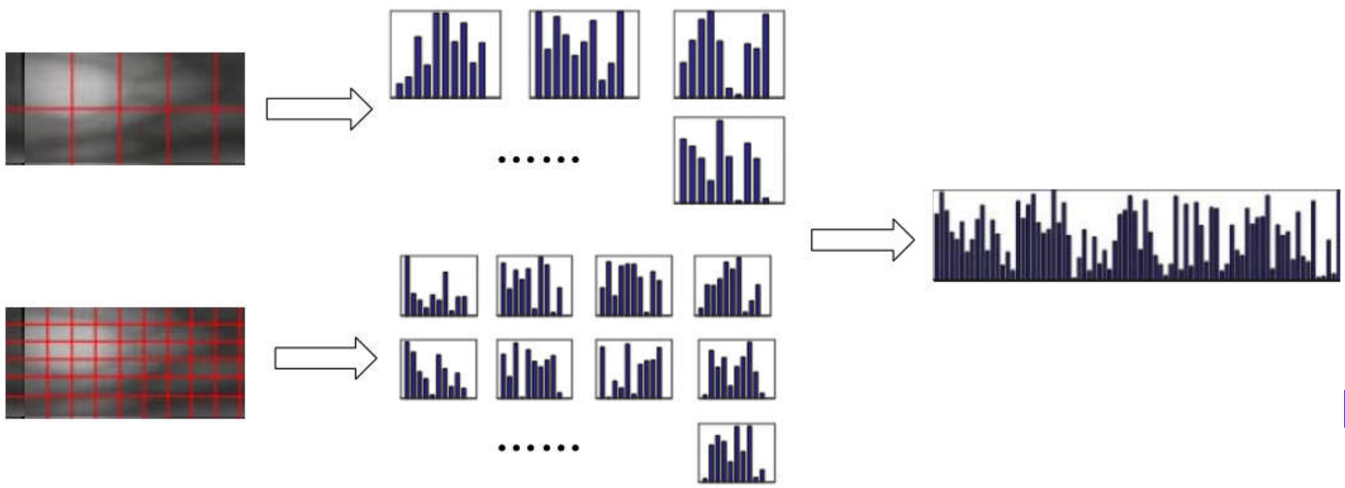


Lightweight model

Yang, L., Yang, G., Xi, X., Su, K., Chen, Q., & Yin, Y. Finger Vein Code: From Indexing to Matching. ***IEEE TIFS 2019.***
Huang, J., Tu, M., Yang, W., & Kang, W. Joint Attention Network for Finger Vein Authentication. ***IEEE TIM 2021***

Soft Biometrics for Vein Recognition

Soft biometrics for vein recognition aim to exploit the kind of “noise”-intensity distribution except the vein pattern



Mean and variance

Array of mean and variance

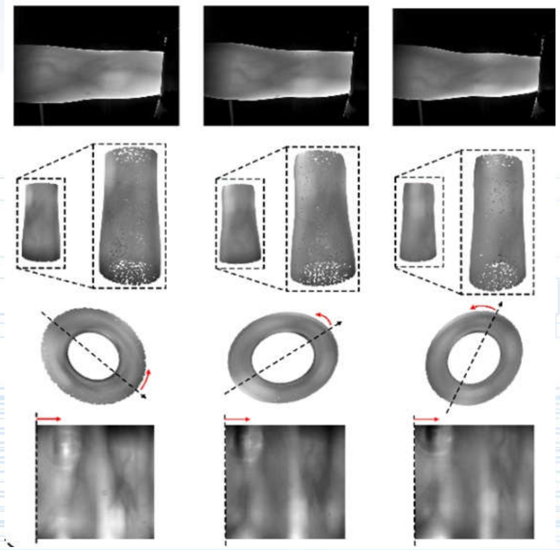
Histogram of spatial pyramid

Soft biometric trait extraction with the spatial pyramid.

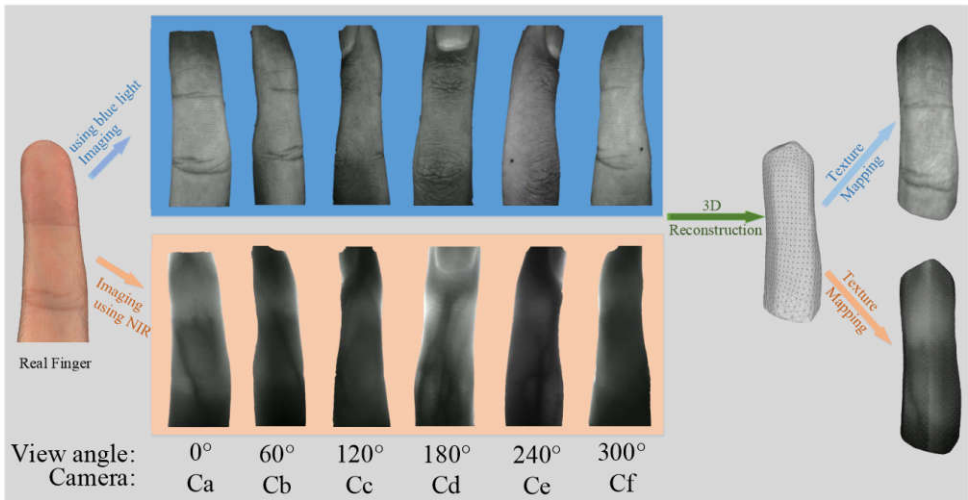
Kang, W., Lu, Y., Li, D., & Jia, W. From Noise to Feature: Exploiting Intensity Distribution as a Novel Soft Biometric Trait for Finger Vein Recognition. ***IEEE TIFS 2019.***

3D finger vein Recognition

3D vein recognition utilize full view cameras to capture vein images around the entire range of the finger and then reconstruct the 3D finger vein model for recognition.



Point cloud and unfolded image



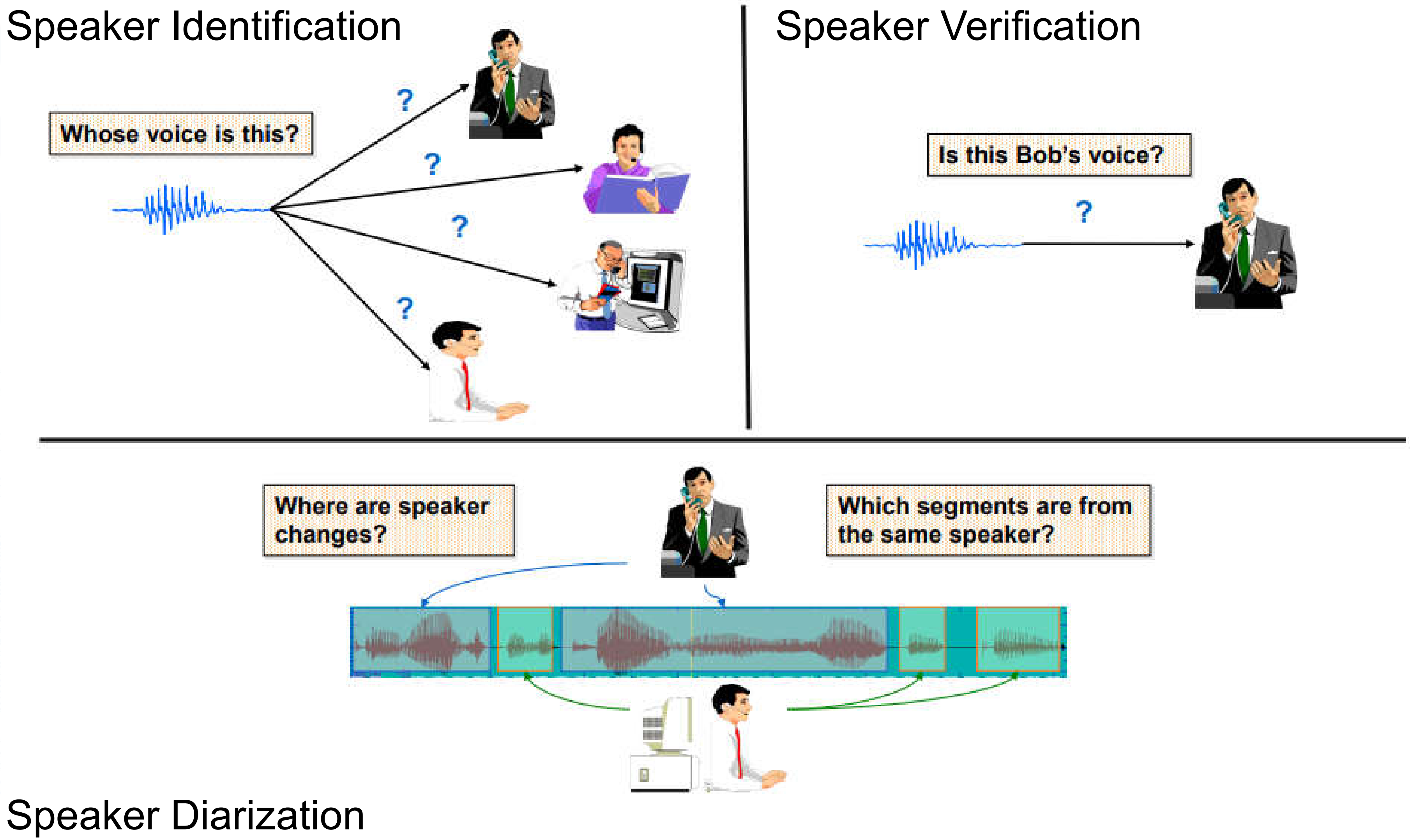
Full view image and 3D model

Kang, W., Liu, H., Luo, W., & Deng, F. Study of a Full-View 3D Finger Vein Verification Technique. ***IEEE TIFS 2020***
Yang, W., Chen, Z., Huang, J., Wang, L., & Kang, W. LFMB-3DFB: A Large-scale Finger Multi-Biometric Database and Benchmark for 3D Finger Biometrics. ***IEEE IJCB 2021***
H. Xu, W. Yang, Q. Wu, W. Kang, Endowing Rotation Invariance for 3D Finger Shape and Vein Verification. ***FCS 2021.***

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

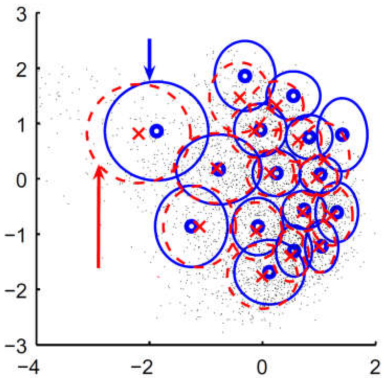
Voiceprint Biometrics



M. Redmond, "Speaker verification: From research to reality," Tutorial of Int.conf.acoustics Speech & Signal Processing May, 2001

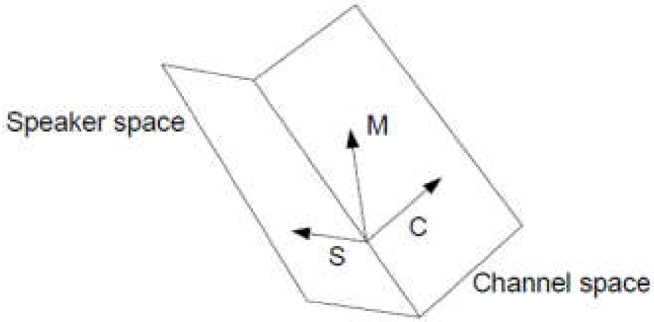
Methods for Voiceprint Biometrics

GMM-UBM

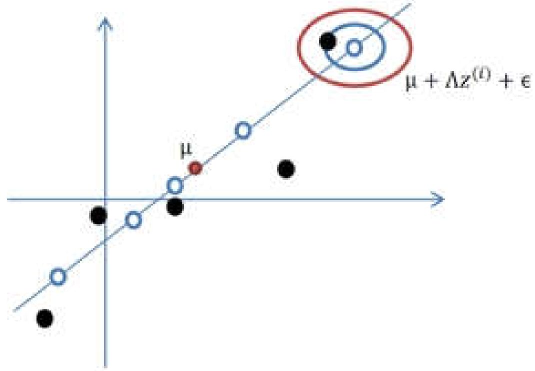


GMM-UBM + MAP

Factor Analysis

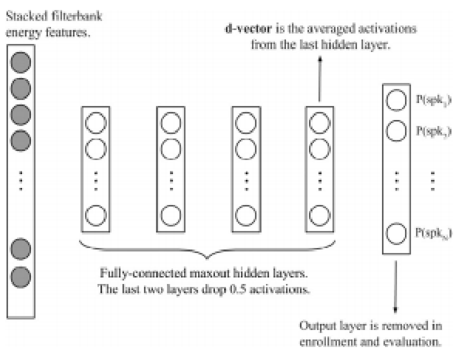


JFA

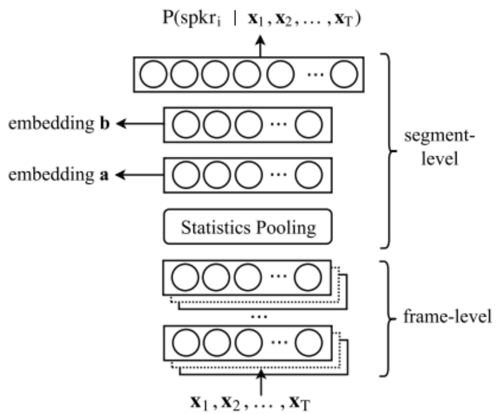


i-vector

Deep Feature Learning

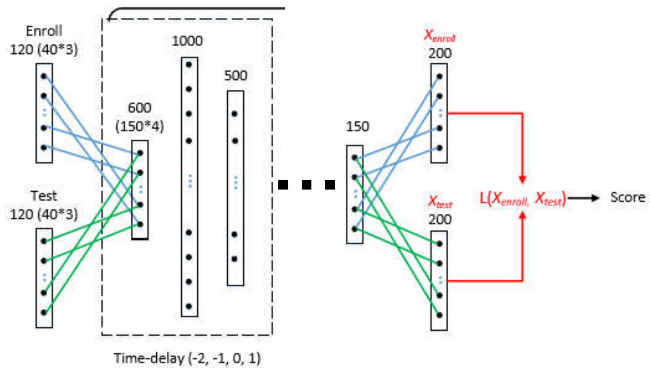


d-vector



x-vector

End-to-End Learning



t-vector

Spoofing ASV System

Impersonation

No obvious threat
(Can only cheat humans)

Speech synthesis

Voice conversion

Medium to high threat

Replay

High threat
Easy to implement

Adversarial Examples

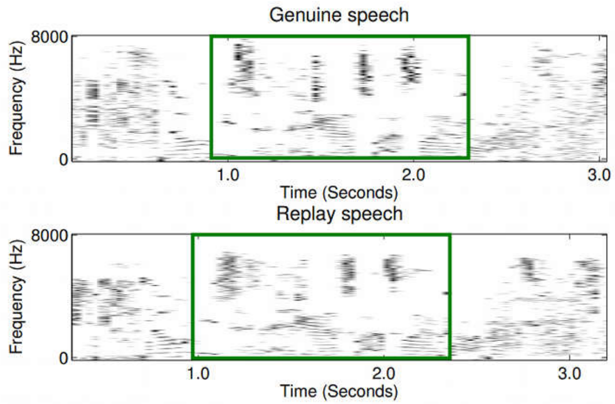
High threat
Difficult to implement in reality

Countermeasure of Replay Spoofing

Challenge and response

Random prompt string

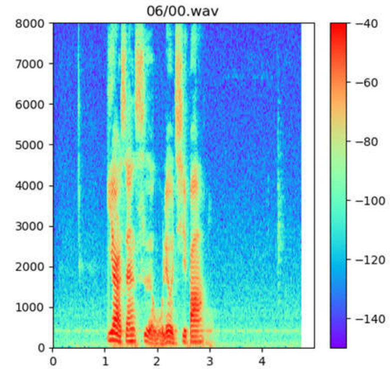
Template match



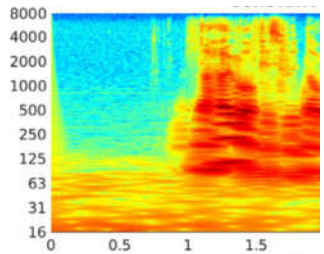
Compared with the voice in the database

Distortion detection

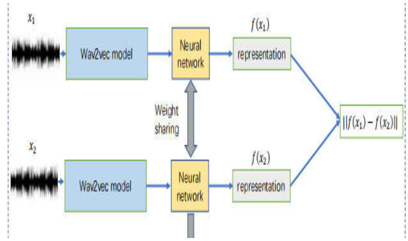
Try to **directly** detect the **distortion** introduced by the playback process (devices).



Feature

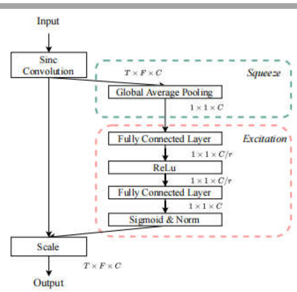


Hand-craft feature

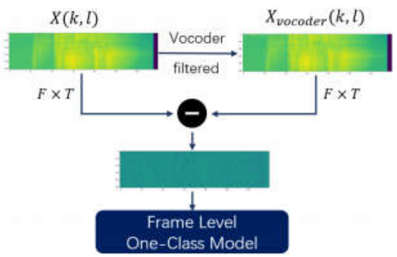


Data-driven learning

Model



Binary classification



One-class learning

Challenges of Voiceprint Biometrics

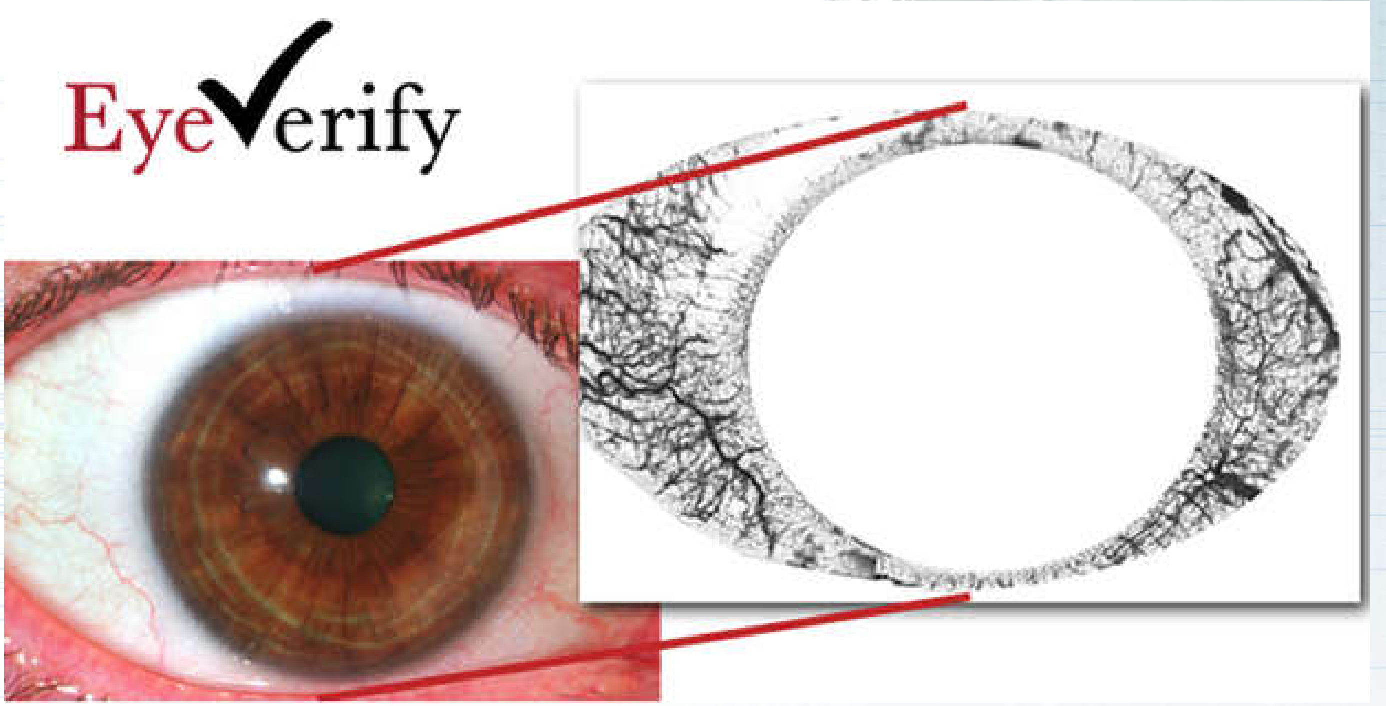
- **Short speech robustness.** How to build a robust speaker model based on a limited duration of enrollment speech, and how to achieve accurate confidence measure and judgment based on a ultra-short duration of test speech.
- **Anti-spoofing.** The performance of speech synthesis technology improved day by day, and the quality of the playback device can be very high. How to protect the system from being deceived under the latest deception technology.
- **Integration with other modalities.** Single-modal automatic speaker recognition technology is limited in accuracy and security in certain scenarios, such as cocktail party scenes, spoofing attacks, etc. Therefore, how to effectively integrate it with other modalities, such as video, is also important.

Outline

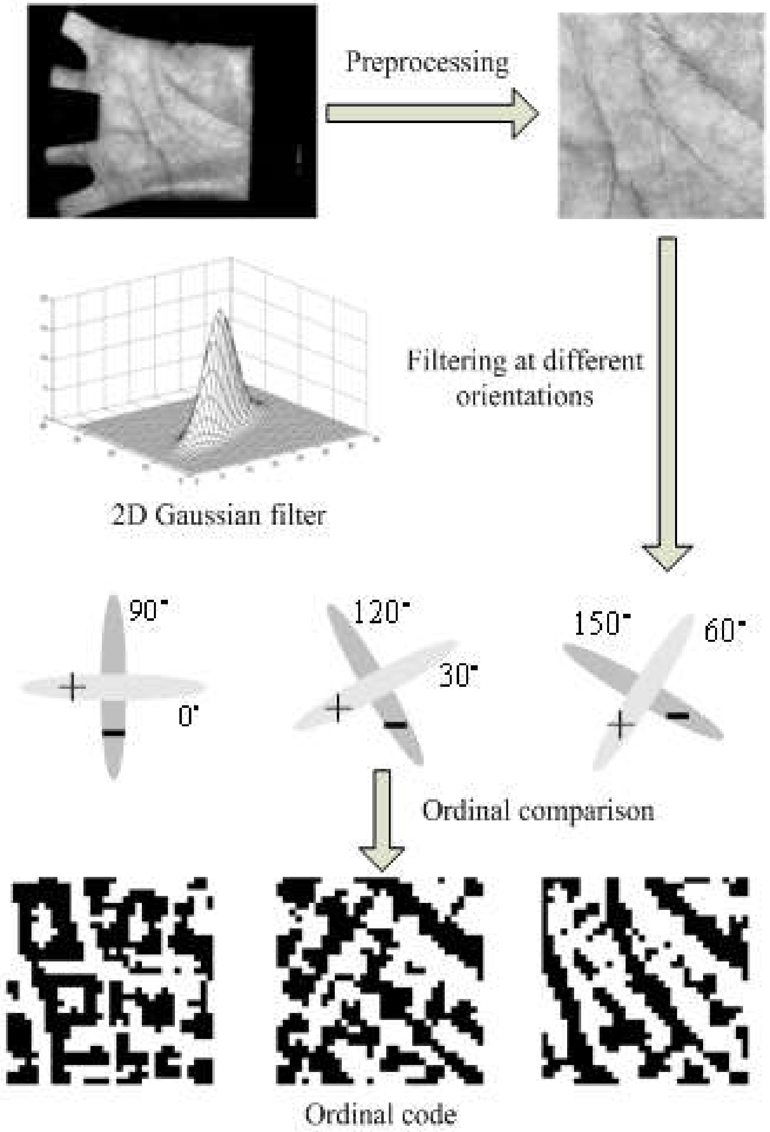
- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

Eyeprint Recognition

Regular front-facing smartphone cameras can create an cryptographic key used to authenticate users based on the micro features in and around their eyes, the most important of which are the blood vessels visible in the whites of the eyes.



Ordinal Measure-based Palmprint Recognition



Ear Biometrics

BBC | Menu | Search

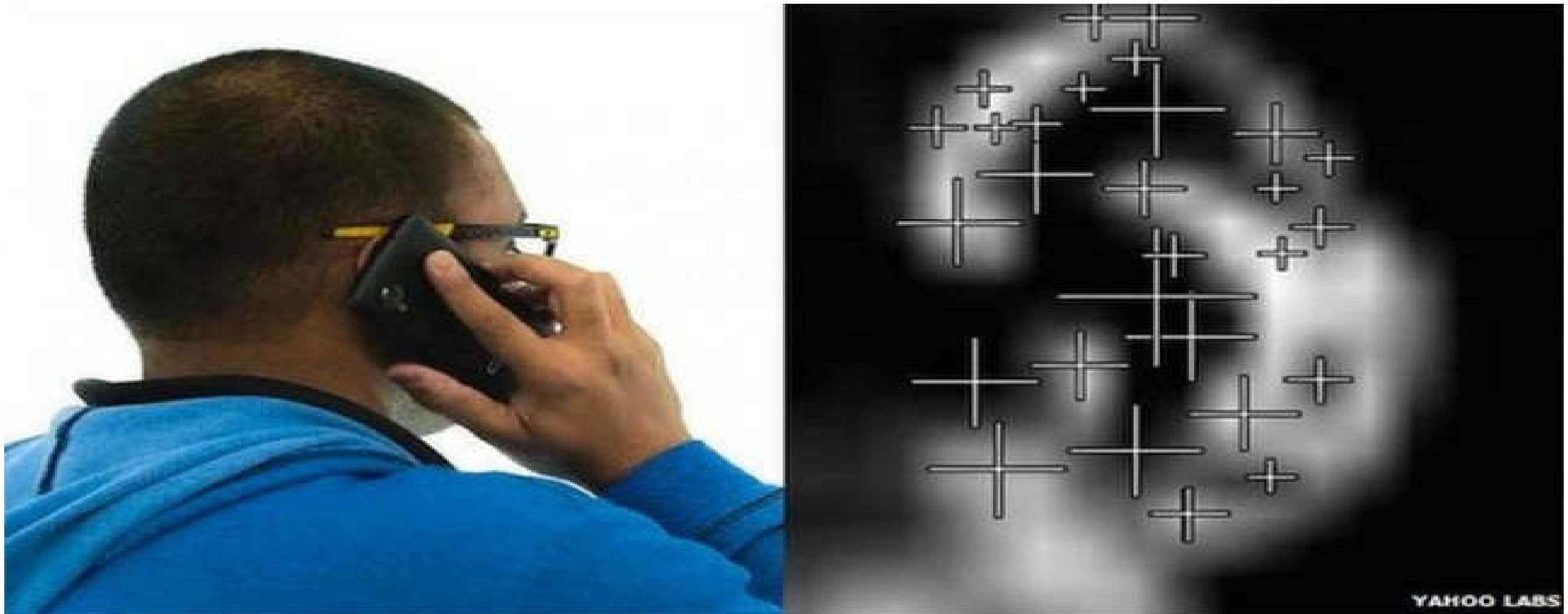
NEWS

Home | Video | World | Asia | UK | Business | Tech | Science | Magazine | Entertainment & Arts | Health | More

Technology

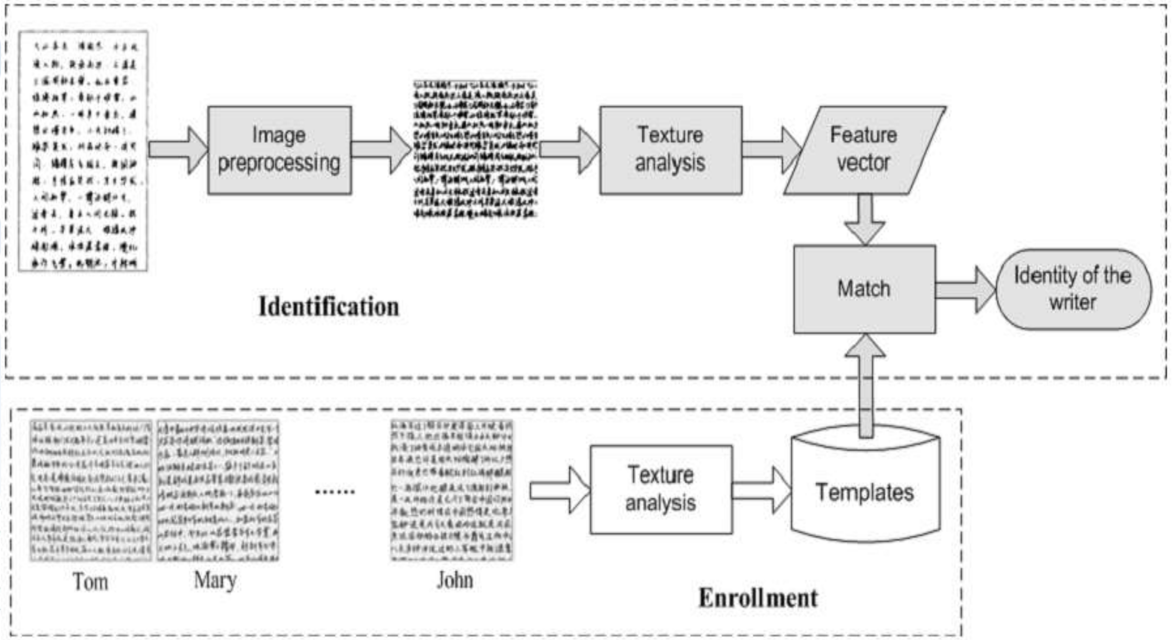
Yahoo tests ear-based smartphone identification system

28 April 2015 | Technology

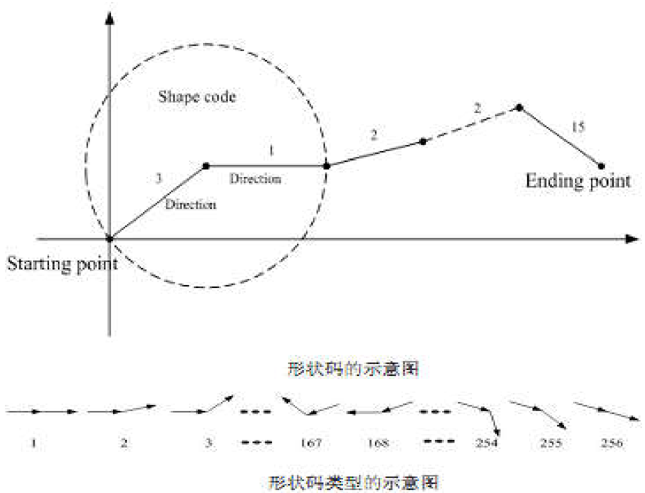


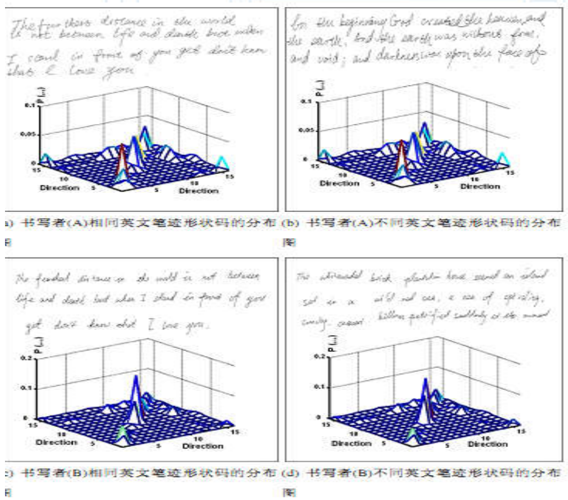
The system identifies users based on the shape of their ears

Handwriting Biometrics



Handwriting texture analysis for writer identification





Statistical analysis of stroke shape features for writer identification

Outline

- Preamble
- Overview of Recent Progress on Biometrics
  - ✓ Fingerprint Recognition
  - ✓ Iris Recognition
  - ✓ Face Recognition
  - ✓ Gait Recognition
  - ✓ Person Re-Identification
  - ✓ Hand Vein Recognition
  - ✓ Speaker Recognition
  - ✓ Others
- Future Directions and Conclusions

Challenges of Biometric Recognition

50 Years of Biometric Research:

Almost

~~The~~ Solved, The Unsolved, and The Unexplored

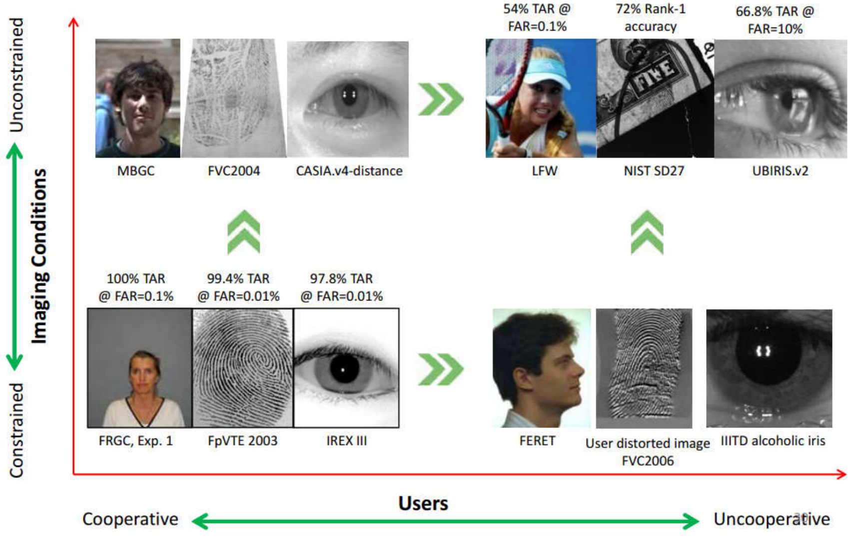
Anil Jain

Michigan State University

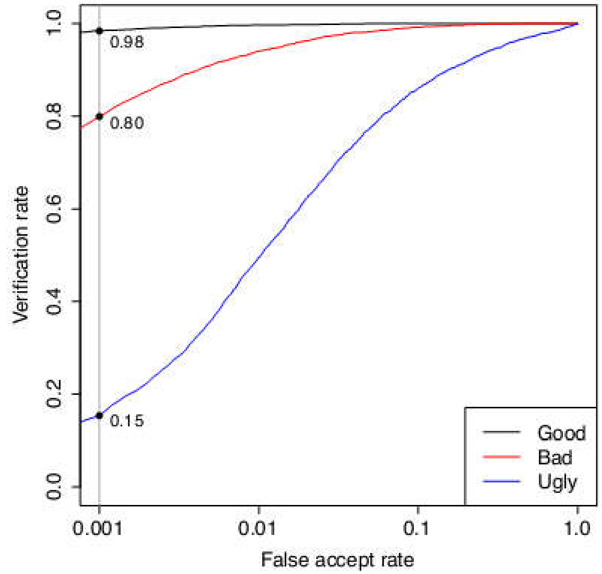
June 5, 2013

Keynote Talk Delivered at the International Conf. on Biometrics, Madrid, Spain, June 5, 2013

From Solved to Unsolved



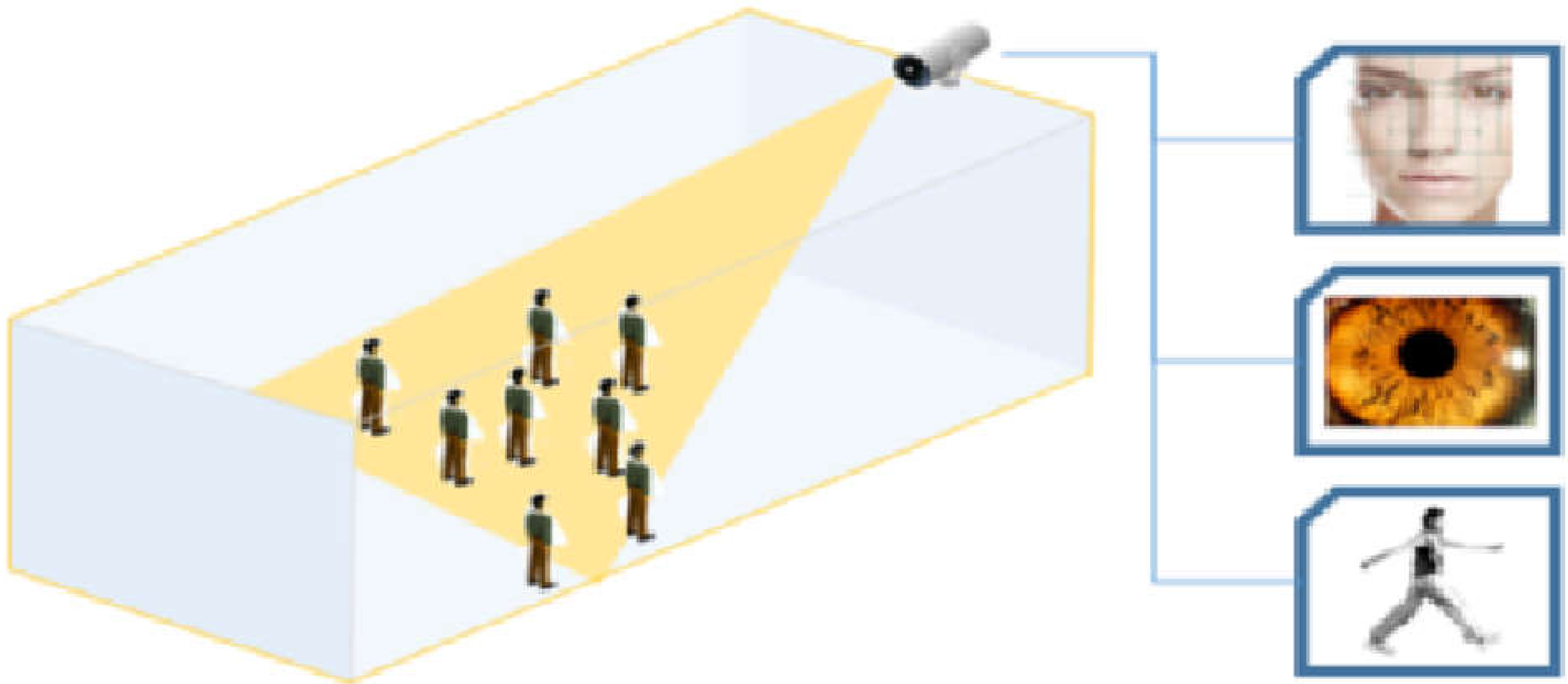
Jonathon Phillips
NIST



An Introduction to the Good, the Bad, & the Ugly Face Recognition Challenge Problem (FG2011)

Future Directions

- Multi-biometrics at a distance

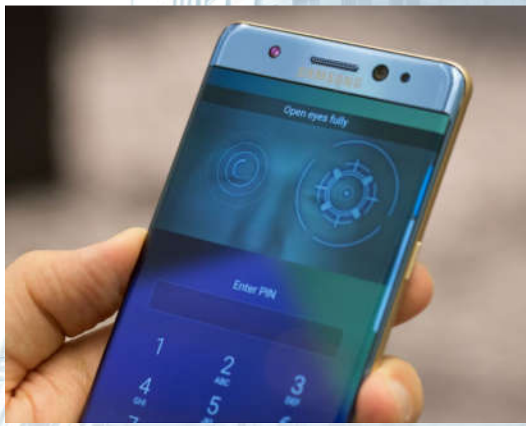


Future Directions

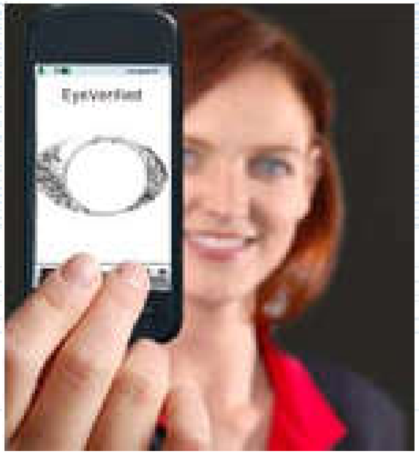
- Multi-biometrics for mobile devices

Fingerprint

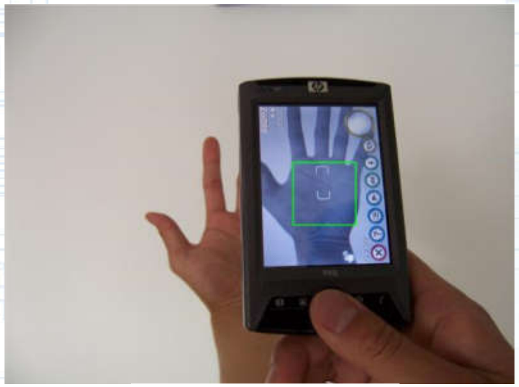
Face



Iris



Eyeprint



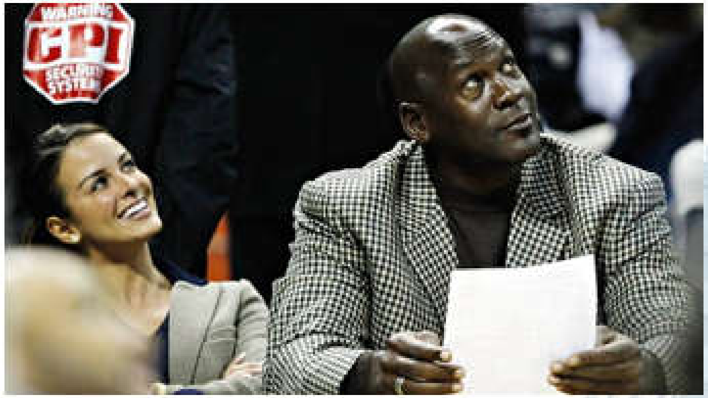
Palmprint

Voiceprint

Future Directions

- Demographic Analysis from Biometric Data

What demographic and affective information can be derived from this face image?



Identity	Rose	Jordan
Gender	Female	Male
Ethnicity	White	Black
Age	27	45
Affect	Happy	Surprised

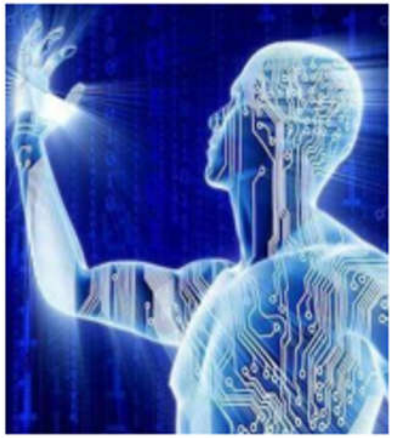
How to determine such information from biometric data?

Yunlin Sun, Man Zhang, Zhenan Sun, Tieniu Tan, Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018.

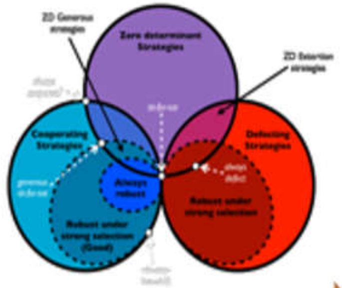
Future Directions

- Deepfake and Anti-Deepfake

Deepfake



Game Theory



Endless War

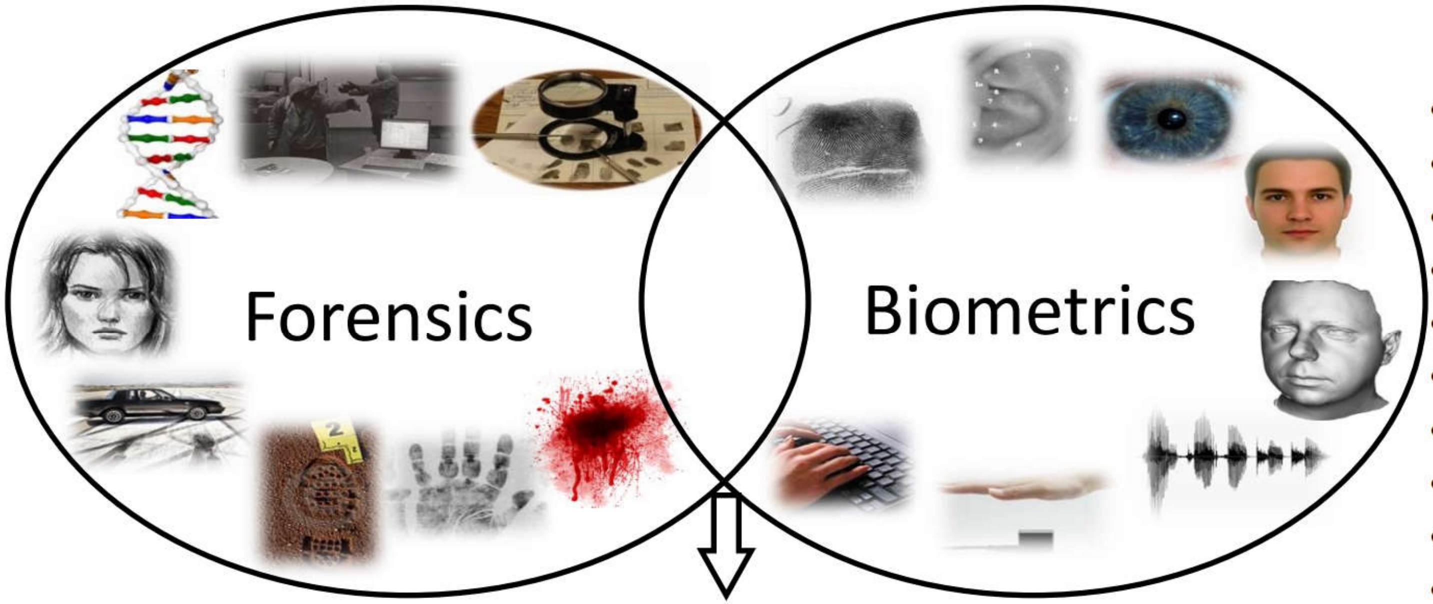
Anti-Deepfake

Future Directions

- Biometrics for forensic applications

Forensics & Biometrics: Shared Goals

- Latent prints
- Fibers
- Explosives
- Paint chips
- DNA
- Tire marks
- Shoe prints
- Bite marks
- SMT



- 2D Face
- 3D Face
- Fingerprint
- Iris
- Speech
- Signature
- Gait
- Ear
- Palmprint
- Keystroke

Forensics: Identify suspects from crime scene evidence

Biometrics: Automated person recognition from *body traits*

Anil K. Jain, Forensics: The Next Frontier for Biometrics, Iowa State University, Ames, Iowa, October 27, 2015.

Conclusions

- Great progress on biometric recognition has been achieved using novel sensors (biometrics-on-the-fly, light field camera) and algorithms (CNN, GAN).
- State-of-the-art biometric methods are accurate and fast enough for many practical applications.
- Many open problems remain to be resolved to make biometric recognition more user-friendly, robust and secure.

Acknowledgement

- Ran He, Liang Wang, Yunlong Wang, Boqiang Xu, CASIA
- Jianjiang Feng, Fang Zheng, Tsinghua Univ.
- Meina Kan，Jie Zhang，Hu Han，Shiguang Shan，ICT CAS
- Weihong Deng, Beijing University of Posts and Telecommunications
- Wenxiong Kang, South China University of Technology
- Wei-Shi Zheng, Sun Yat-sen University
- Wangmeng Zuo, Harbin Institute of Technology

Thank you!

Q & A