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

CRIPAC
智能感知与计算研究中心
Center for Research on Intelligent
Perception and Computing

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

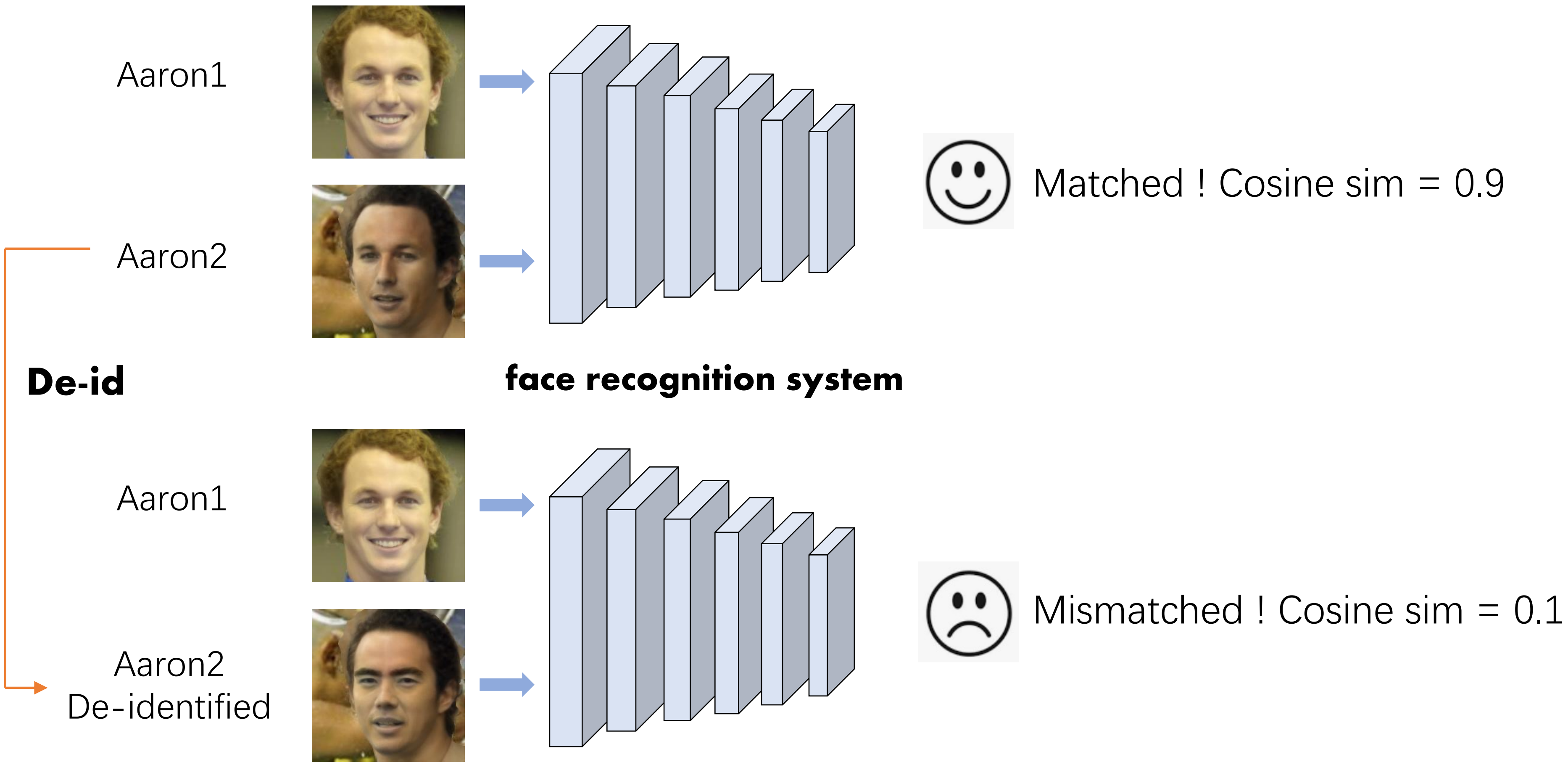
RiDDLE： Reversible and Diversified De-identification with Latent Encryptor

Dongze Li[1,2], Wei Wang[2*], Kang Zhao[3], Jing Dong[2] , Tieniu Tan[2,4]

[1] School of Artificial Intelligence, University of Chinese Academy of Sciences
[2] Center for Research on Intelligent Perception and Computing, CASIA
[3] Alibaba Group [4] Nanjing University

Background: Face De-identification

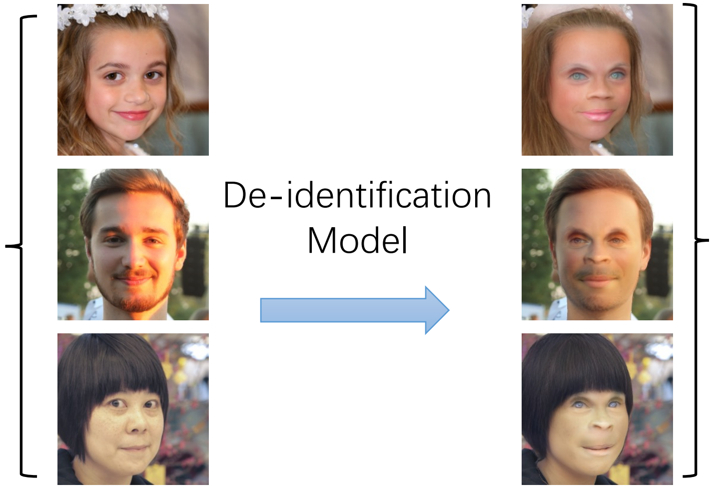
To hide the identity information of face images for privacy protection



Motivation

Drawbacks of Current Works

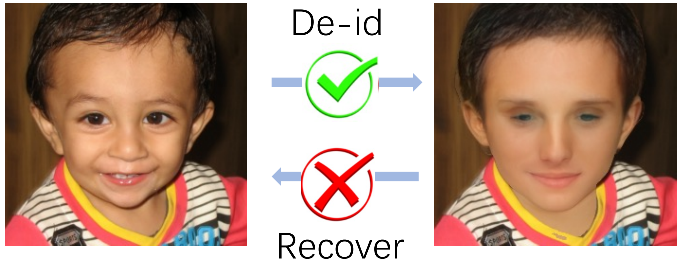
Lack of diversity



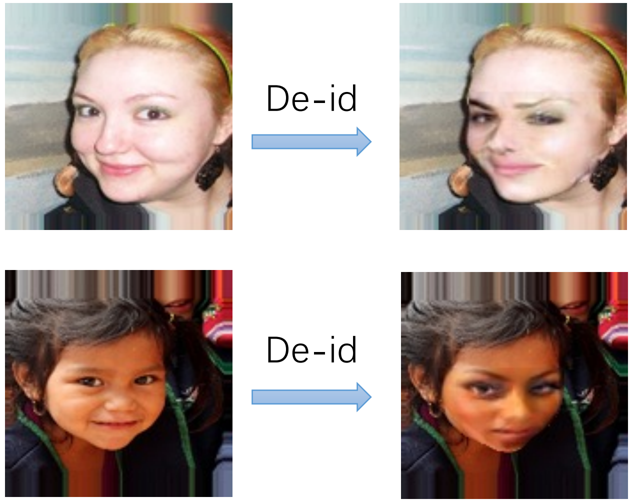
Different People

Anonymous faces
with similar appearance

Irreversible



Poor Quality



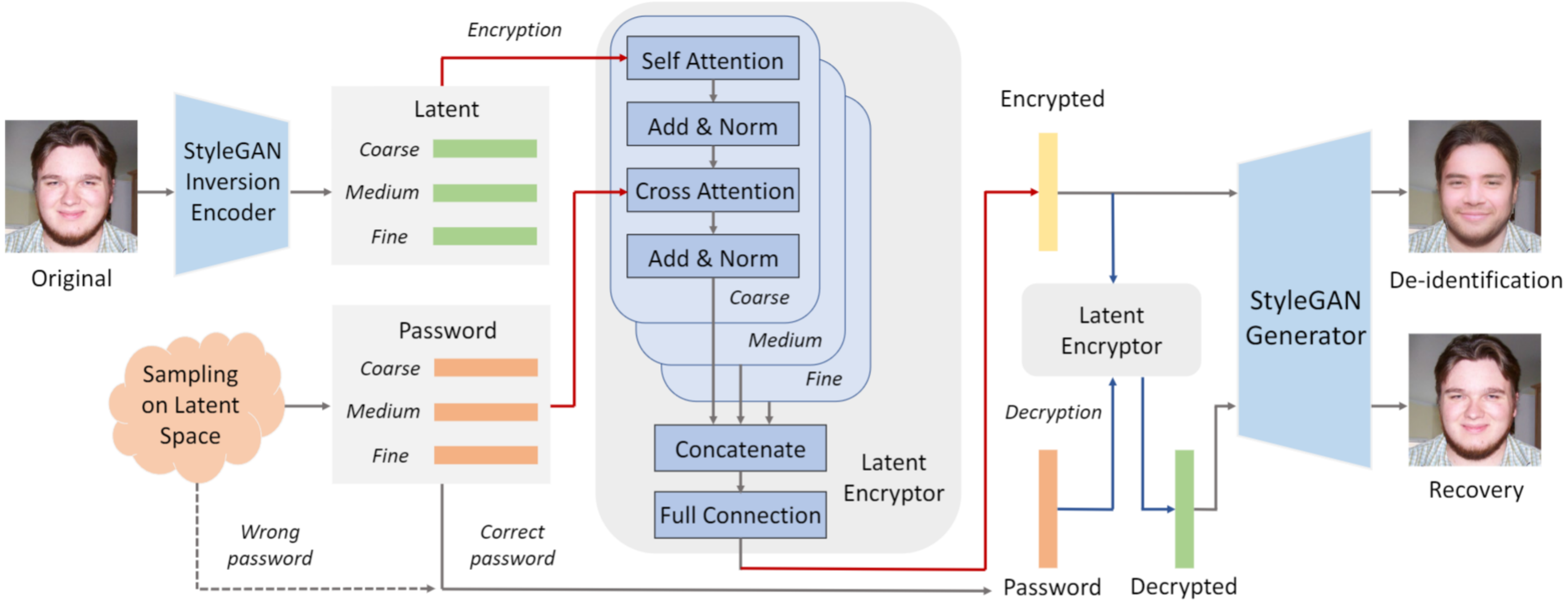
Fix: Identity diversity loss term

Fix: Password Scheme

Fix: StyleGAN

Our Solution:
Reversible and **D**iversified **De**-identification with **L**atent **E**ncryptor （**RiDDLE**)

Latent Space Identity Encryption and Decryption



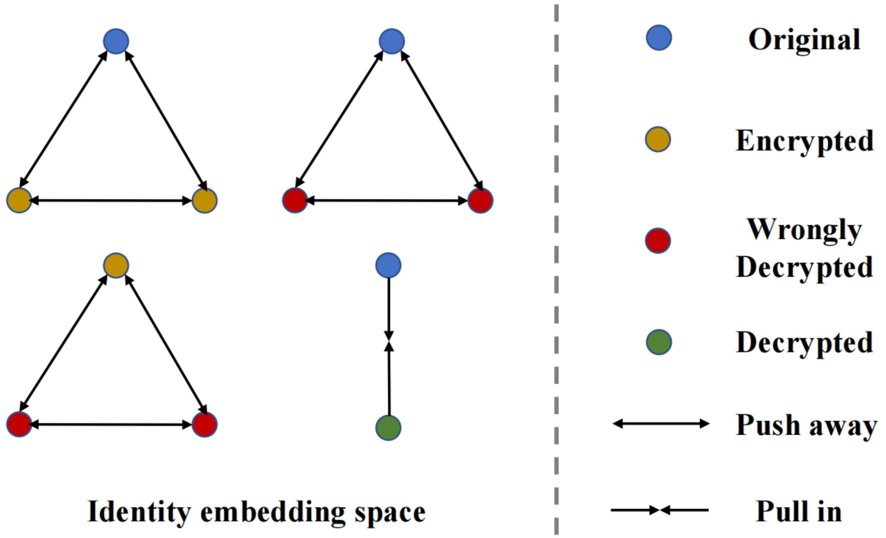
During encryption
Each password is associated with a unique identity.

During decryption
Password is correct ? -> Recover the original identity.
Password is incorrect ? -> Return a wrong identity with realism.

RiDDLE also supports data-free training by randomly sampling latents with StyleGAN

Loss Functions

De-identification and Recovery



$$\mathcal{L}_{div} = \frac{1}{m^2(n+1)^2} \cdot \mathbf{sum}((\mathbf{1} - \mathbf{I}) \cdot \mathcal{M}),$$

$$\mathcal{L}_{deid} = \sum_{i}^{m(n+1)} \max\left(\epsilon, \cos\left(F_e(\mathbf{x}), F_e(\mathbf{x}^i)\right)\right)$$

$$\mathcal{L}_{rec} = \sum_{i}^{m} \left(1 - \cos\left(F_e(\mathbf{x}), F_e(\mathbf{x}^i_{corr})\right)\right)$$

$$\mathcal{L}_{id} = \mathcal{L}_{div} + \mathcal{L}_{deid} + \mathcal{L}_{rec}.$$

Maintaining Image Quality and Utility

$$\mathcal{L}_{pix} = \|\mathbf{x} - \mathbf{x}^*\|_1 .$$

Image level reconstruction

$$\mathcal{L}_{LPIPS} = \|F_p(\mathbf{x}) - F_p(\mathbf{x}^*)\|_2$$

Feature level reconstruction

$$\mathcal{L}_{parse} = \|F_s(\mathbf{x}) - F_s(\mathbf{x}^*)\|_2,$$

Avoid unreal face features

$$\mathcal{L}_{latent} = \|\mathbf{w} - \mathbf{w}^*\|_2 .$$

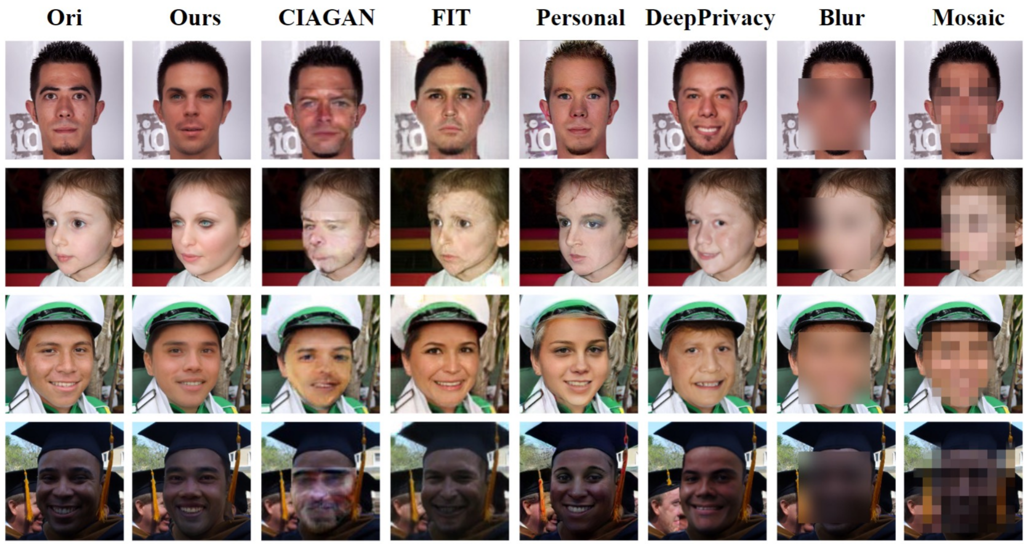
Latent space regularization

Final Loss Term

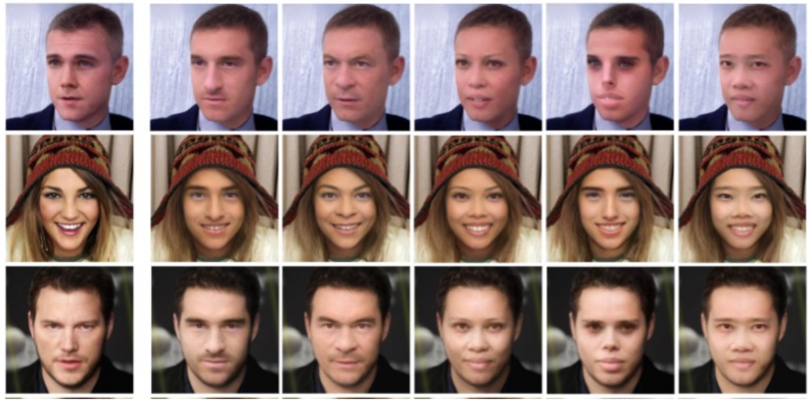
$$\mathcal{L}_{total} = \mathcal{L}_{id} + \lambda_{pix}\mathcal{L}_{pix} + \lambda_{LPIPS}\mathcal{L}_{LPIPS} + \lambda_{parse}\mathcal{L}_{parse} + \lambda_{latent}\mathcal{L}_{latent}.$$

Results: Qualitative

De-identification

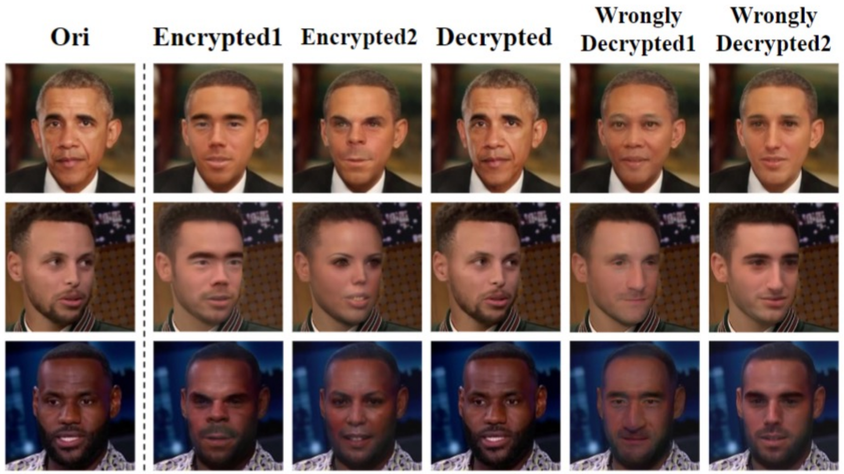


Diverse Identity Generation



Recovery

Results in the wild

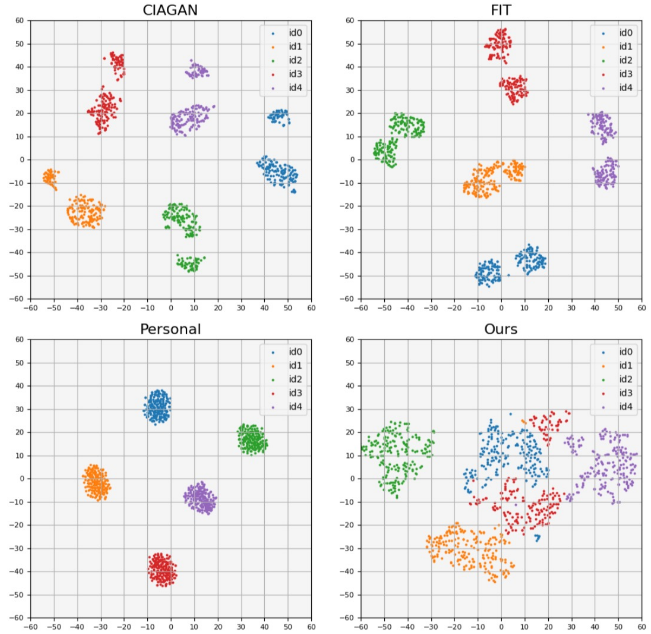


Results: Quantitative

De-identification / Recovery

Type	Method	FaceNet CASIA	FaceNet VGGFace2	SphereFace
De-id ↓	Ours	**0.016**	**0.032**	0.025
	Ours-DF	0.034	0.037	0.025
	CIAGAN [17]	0.019	0.034	**0.010**
	FIT [7]	0.042	0.072	0.065
	Personal [4]	0.020	0.042	0.017
	DeepPrivacy [10]	0.266	0.184	0.120
Recovery ↑	Ours	**0.996**	**0.998**	**1.000**
	Ours-DF	0.953	0.949	1.000
	FIT [7]	0.967	0.974	**1.000**
	Personal [4]	0.965	0.965	0.998

Diversity



Recovered Image Quality

	MSE↓	LPIPS↓	SSIM↑	PSNR↑
FIT [7]	0.005	0.186	0.934	23.130
Personal [4]	0.003	0.220	0.846	26.391
Ours	**0.002**	**0.043**	**0.966**	**26.499**
Ours-DF	0.004	0.277	0.760	25.483

Image Utility

Method		Ours	Ours-DF	CIAGAN [17]	FIT [7]	Personal [4]	DeepPrivacy [10]
FID ↓		**15.389**	26.802	32.611	30.331	25.715	23.713
Face detection ↑	MtCNN	**1.000**	1.000	0.992	**1.000**	**1.000**	**1.000**
	Dlib	0.991	0.975	0.937	0.984	**0.992**	0.980
Bounding box distance ↓	MtCNN	**3.824**	5.720	20.387	7.879	4.213	4.654
	Dlib	**1.700**	3.109	15.476	4.218	2.726	2.685
Landmark distance ↓	MtCNN	**1.674**	3.252	8.042	3.572	2.358	3.280
	Dlib	**1.512**	2.973	8.930	4.047	2.459	2.896

Ablation Study

Qualitative

Quantitative

	De-id ↓	Recovery ↑	FID ↓
w/o data	0.034	0.953	26.802
w/o transformer	0.018	0.985	22.704
w/o identity diversity loss	0.025	0.993	25.816
full	**0.016**	**0.996**	**15.389**

- w/o data -> Degradation in quality, higher privacy level
- w/o identity diversity loss -> Naïve De-identification
- w/o transformer -> Degradation in quality

Summary

- Expose the drawbacks of the current face de-identification methods.
- Propose a de-identification method based on a novel latent encryptor and a password scheme.
- Our method achieves better quality, higher diversity and stronger reversibility on various face datasets and in the wild images.

Thanks