Attribute-Preserving Face Dataset Anonymization via Latent Code Optimization

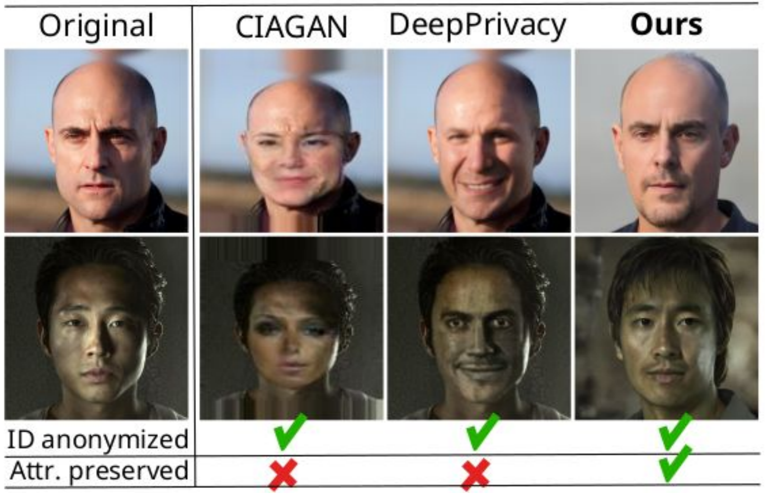
Simone Barattin*[1], Christos Tzelepis*[2], Ioannis Patras[1], and Nicu Sebe[1]
[1]University of Trento, [2]Queen Mary University of London

(* denotes equal contribution) - TUE-PM-371

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Goal of the work

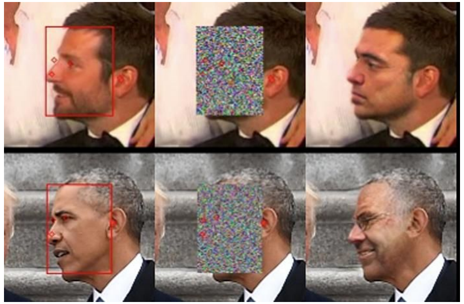
- Anonymize the identity of face images
- Maintain the original face attributes



Background

- Face obfuscation
  - Naive masking methods [1]
  - *k*-Same algorithm [2]

- Generative face anonymization
  - CIAGAN [3]
  - DeepPrivacy [4]



CIAGAN [3]

DeepPrivacy [4]

[1] Datong Chen, Yi Chang, Rong Yan, and Jie Yang. "Tools for protecting the privacy of specific individuals in video.", EURASIP 2007
[2] Elaine M Newton, Latanya Sweeney, and Bradley Malin. "Preserving privacy by de-identifying face images.", IEEE TKDE 2005
[3] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. “CIAGAN: Conditional identity anonymization generative adversarial networks”, CVPR 2020
[4] Hukkelås, Håkon, Rudolf Mester, and Frank Lindseth. "DeepPrivacy: A generative adversarial network for face anonymization.", ISVC 2019

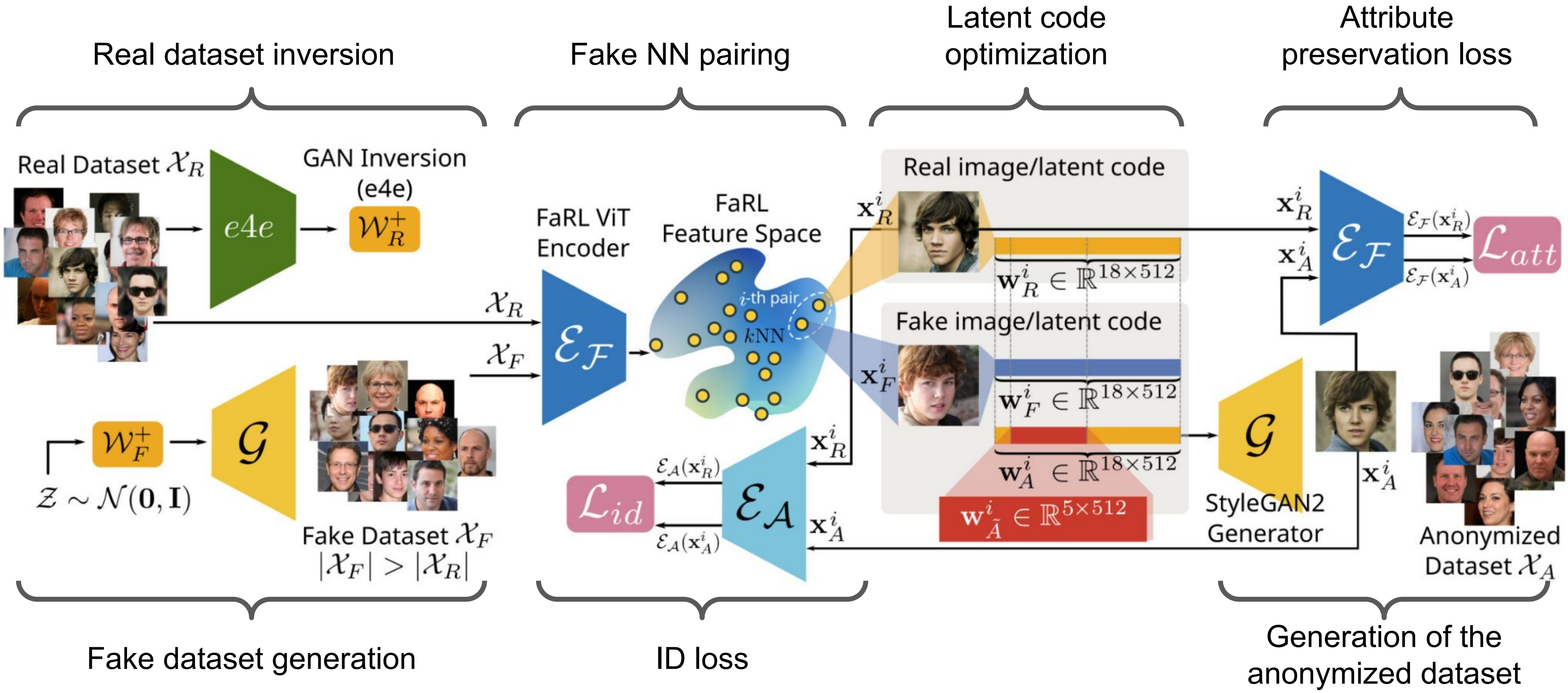
Background

Challenges and proposed solution

- Costly and unstable training of additional neural networks
- Facial attributes and expression are not preserved

- Use only pre-trained models
  - Greatly reduces the computational cost
- Use a novel loss to retain fine-grained facial details
  - Meanwhile the identity is changed

Pipeline overview



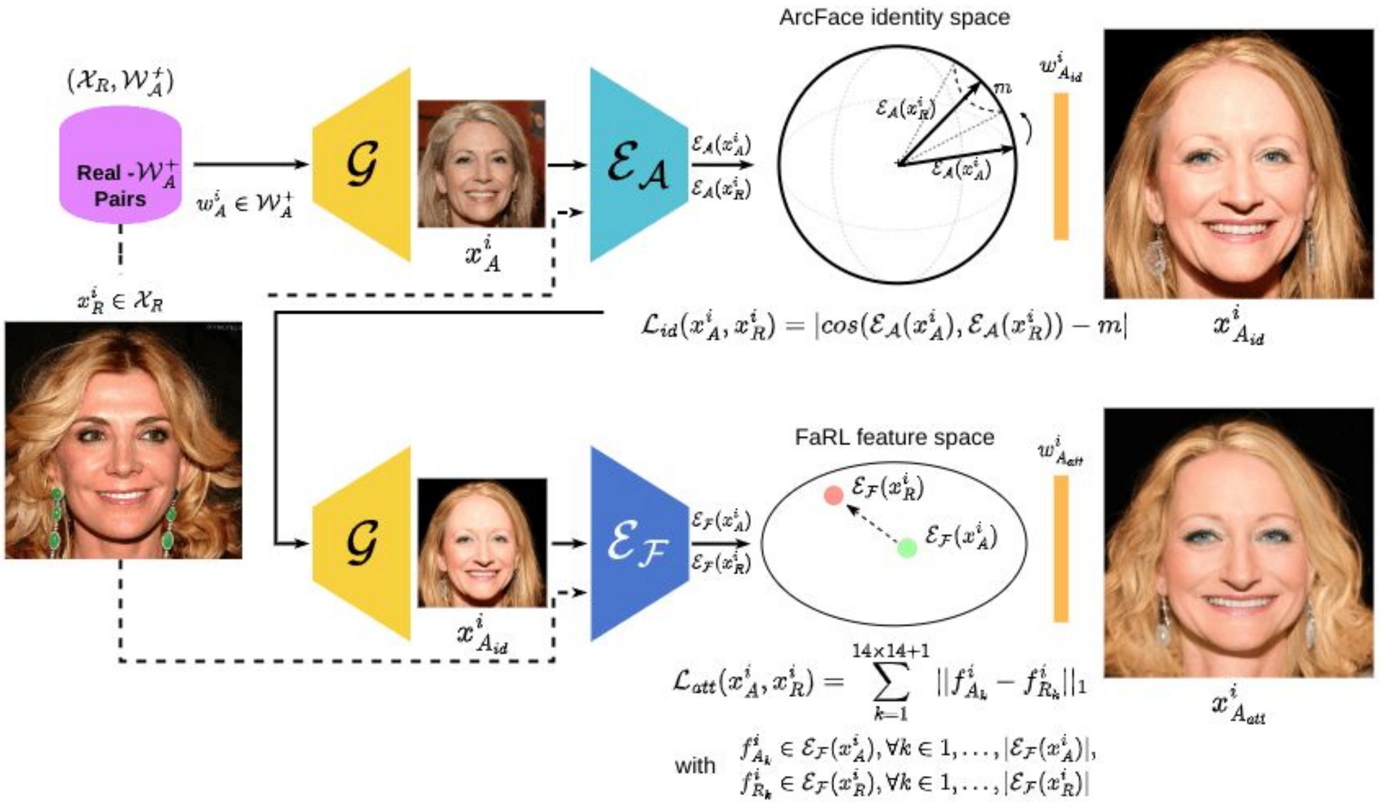
Anonymization process

- Proposed **identity loss** $\mathcal{L}_{id}(\mathbf{x}_A^i, \mathbf{x}_R^i) = \left|\cos\left(\mathcal{E}_\mathcal{A}(\mathbf{x}_A^i), \mathcal{E}_\mathcal{A}(\mathbf{x}_R^i)\right) - m\right|$
    - $\mathcal{E}_A$ denotes the pre-trained ArcFace [1] encoder
    - Controls the similarity between the real and the anonymized faces via the hyperparameter *m*

- Proposed **attribute preservation loss** $\mathcal{L}_{att}(\mathbf{x}_A^i, \mathbf{x}_R^i) = \left\|\mathcal{E}_\mathcal{F}(\mathbf{x}_A^i) - \mathcal{E}_\mathcal{F}(\mathbf{x}_R^i)\right\|_1$
    - $\mathcal{E}_F$ denotes the pre-trained FaRL [2] visual encoder (ViT-based)
    - Imposes the preservation of the real images' facial features on the anonymized ones

[1] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Cotsia, and Stefanos P Zafeiriou. "ArcFace: Additive angular margin loss for deep face recognition.", PAMI 2021
[2] Yinglin Zheng, Hao Yang, Ting Zhang, Jianmin Bao, Dong-dong Chen, Yangyu Huang, Lu Yuan, Dong Chen, Ming Zeng, and Fang Wen. "General facial representation learning in a visual-linguistic manner", CVPR 2021

Anonymization process



Experiments

Datasets

- CelebA-HQ [1]
  - 30000 frontal-face images
  - 40 facial attribute annotations
  - Test the ability of the method to anonymize high quality images

- Labelled Faces in the Wild (LFW) [2]
  - 13000 in-the-wild images
  - No facial attribute annotation is provided
  - Test the ability of the method to anonymize images in-the-wild

[1] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. “Deep learning face attributes in the wild.”, ICCV 2015
[2] Huang, Gary B., et al. "Labeled faces in the wild: A database for studying face recognition in unconstrained environments." Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition. 2008.

Results

- Image quality evaluation
  - Fréchet Inception Distance (FID)
  - Face detection rate (MTCNN, dlib)

	FID↓	Detection↑		Face re-ID↓	
		dlib(%)	MTCNN(%)	CASIA(%)	VGG(%)
Randomly generated	**18,09**	100	99.99	3.61	1.08
CIAGAN [35]	37,94	95.10	99.82	**2.19**	**0.37**
DeepPrivacy [21]	32.99	98.82	99.85	3.61	1.05
Our (ID)	44.12	98,58	97.99	3.28	0.58
Our (ID+attributes)	44.11	**100**	**100**	3.06	2.06
Our	**29.93**	**100**	**100**	2.80	1.67

- Face de-identification evaluation
  - Face re-identification

	FID↓	FID (C-HQ)↓	Detection↑		Face re-ID↓	
			dlib(%)	MTCNN(%)	CASIA(%)	VGG(%)
CIAGAN [35]	**22.07**	85.23	98.14	99.89	**0.17**	**0.91**
DeepPrivacy [21]	23.46	123.67	96.7	99.57	2.74	1.52
Our	27.45	**68.88**	**100**	**100**	2.07	1.58

Results

- Attribute preservation evaluation
  - Attribute classification approach
  - Accuracy of the trained classifier

	Inner face	Outer face	Combined
Original	**0.8409**	**0.8683**	**0.8539**
CIAGAN[35]	0.7277	0.8372	0.7852
DeepPrivacy[21]	0.7658	0.8511	0.8135
Our	**0.7817**	**0.8518**	**0.8181**

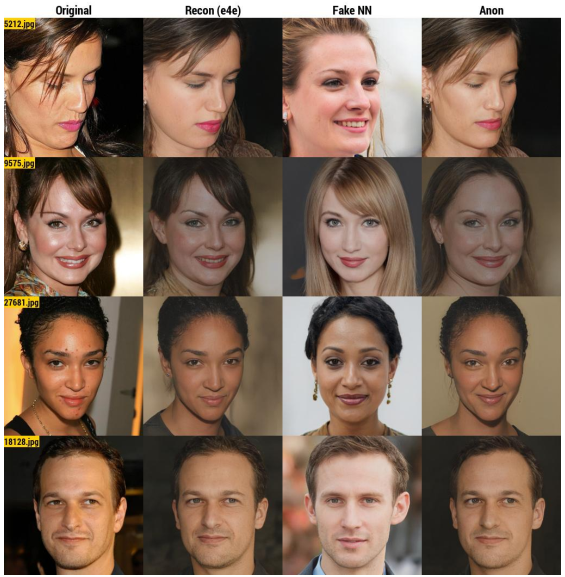
- Use pseudo-labels for LFW
  - Two pre-trained attribute classifiers
  - Lin et al. [30] predicts CelebA-HQ's attributes
  - Jiang et al. [22] predicts 5 facial attributes

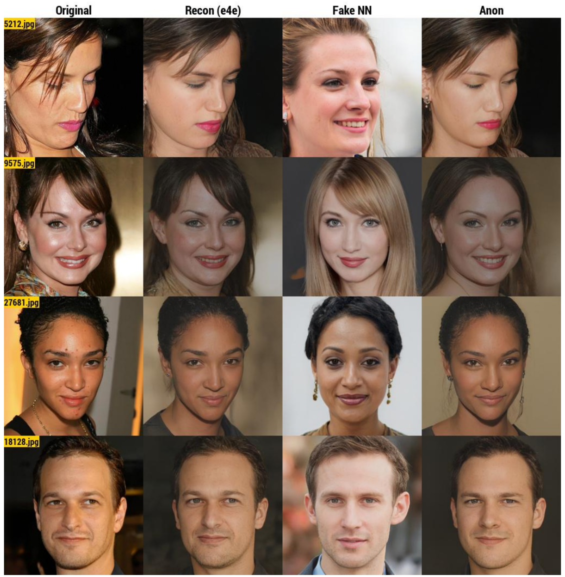
	CelebA-HQ (labels from [30])	LFW (labels from [30])	LFW (labels from [22])
CIAGAN [35]	0.7721	0.9143	0.7045
DeepPrivacy [21]	0.7902	0.9133	0.7019
Our	**0.8215**	**0.9157**	**0.7209**

Results

ID preserved — ID changed

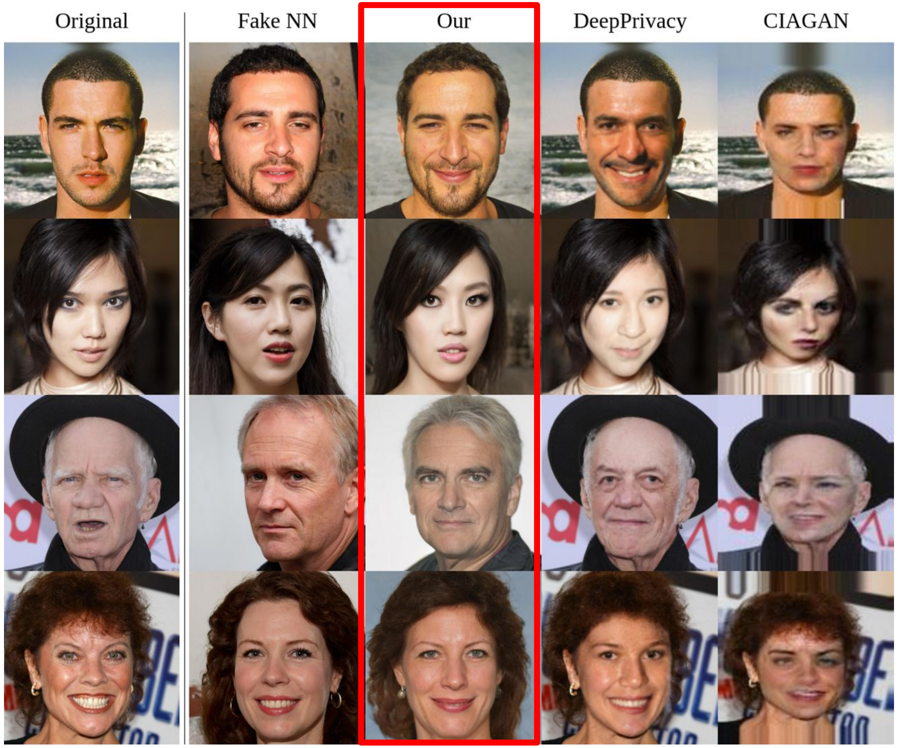
m=1.0 ← m=0.5 → m=0.0

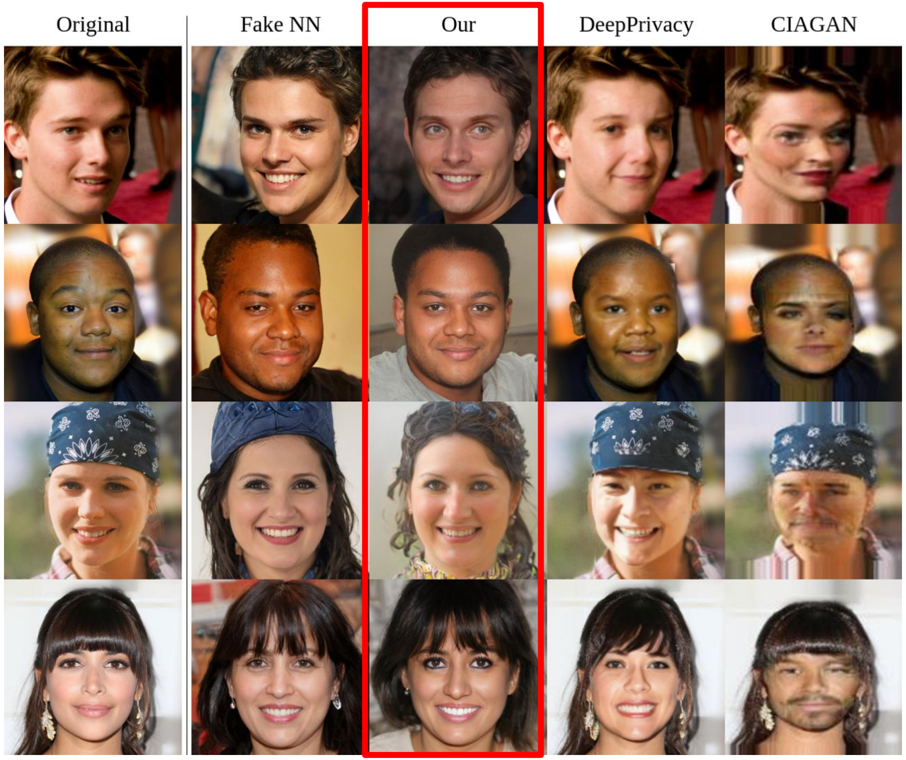




	FID	Detection	Face re-ID		Accuracy
		MTCNN(%)	CASIA(%)	VGG(%)	
Our (m=0.0)	29.93	**100**	**2.80**	**1.67**	0.8181
Our (m=0.9)	**27.58**	**100**	3.41	1.76	**0.83**

Results


Original
Fake NN
Our
DeepPrivacy
CIAGAN


Original
Fake NN
Our
DeepPrivacy
CIAGAN

Results


Original
Fake NN
Our
DeepPrivacy
CIAGAN
Original
Fake NN
Our
DeepPrivacy
CIAGAN

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Attribute-Preserving Face Dataset Anonymization via Latent Code Optimization

Simone Barattin*[1], Christos Tzelepis*[2], Ioannis Patras[1], and Nicu Sebe[1]
[1]University of Trento, [2]Queen Mary University of London

(* denotes equal contribution)

Code:
https://github.com/chi0tzp/FALCO