A Key-Driven Framework for Identity-Preserving Face Anonymization

Miaomiao Wang， Guang Hua， Sheng Li*， Guorui Feng*

Shanghai University, Singapore Institute of Technology, Fudan University

#NDSSSymposium2025

➢ Background

❑ Mask-Based

Mosaic

Blur

How to balance
privacy and availability?

- **High level of privacy protection**
- **Cannot be used for visual tasks**

❑ DeepLearning-Based

Original **DeID**

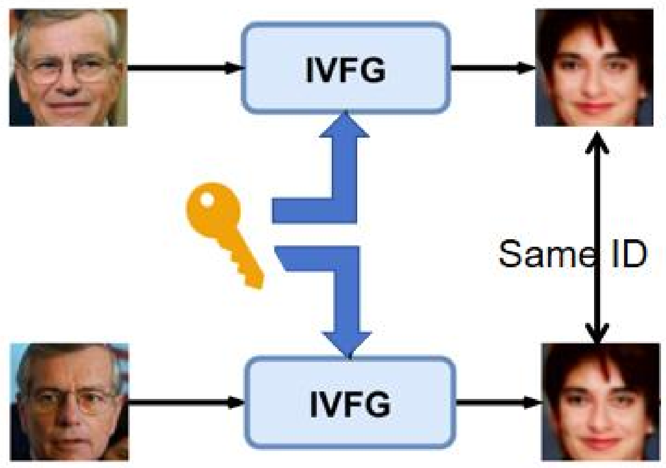
DeIDGAN (2021)

- **High level of privacy preserving**
- **Identity is permanently lost**

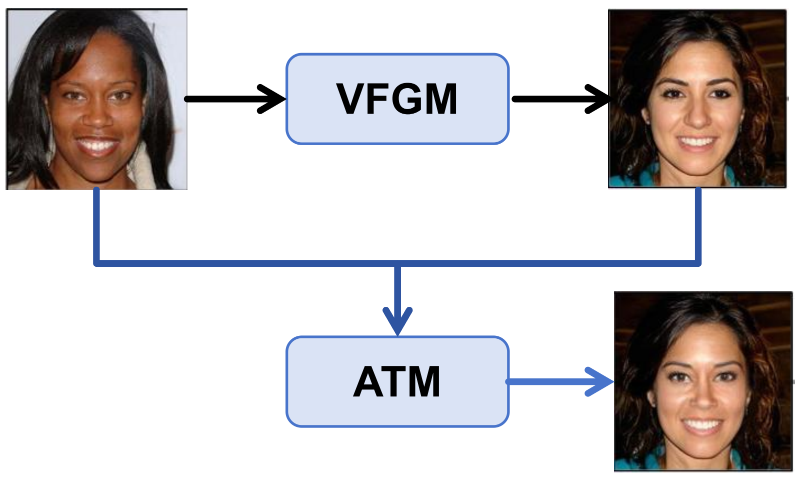
➢ Background

To balance privacy and recognizability

❑ Identifiable virtual faces(2022)



❑ Identity Hider(2024)



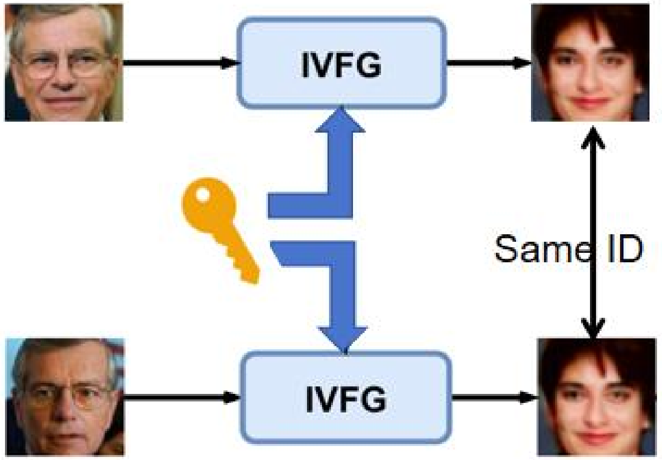
- Authentication in Virtual Domain
- The original face is not synchronized

- Facial feature synchronization
- Identity is easily leaked

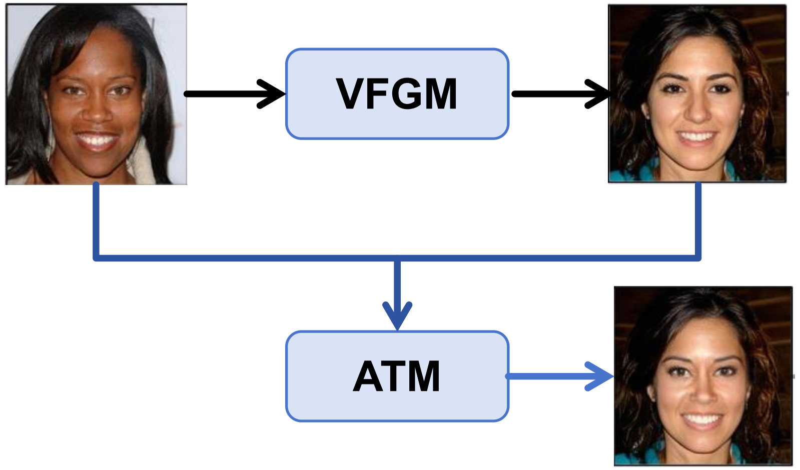
➢ Background

To balance privacy and recognizability

❑ Identifiable virtual faces(2022)



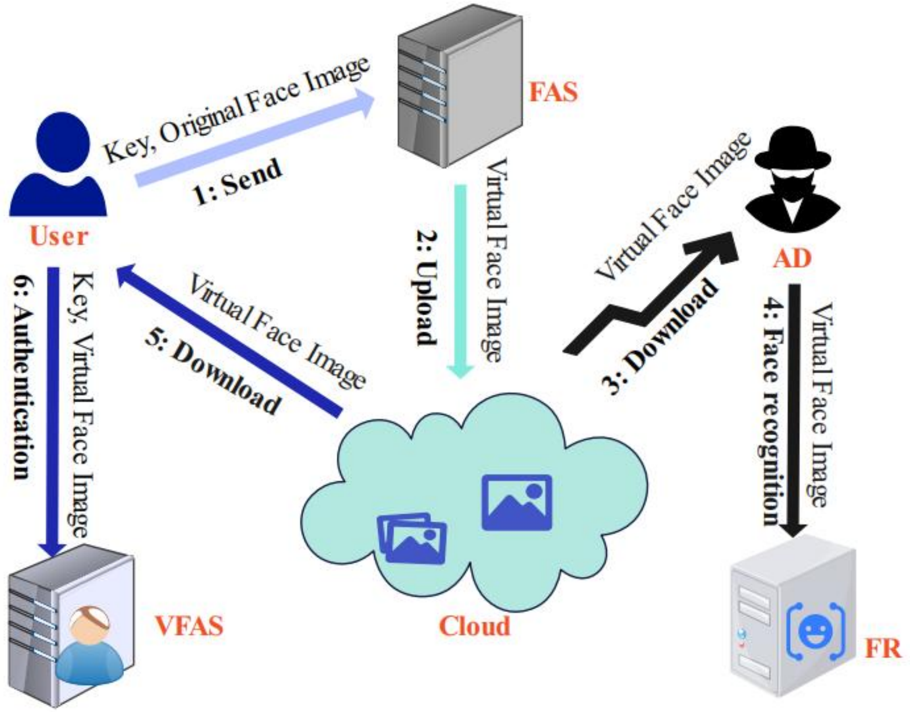
❑ Identity Hider(2024)



How to resolve the conflict between facial privacy and recognizability, while maintaining head posture and facial expressions?

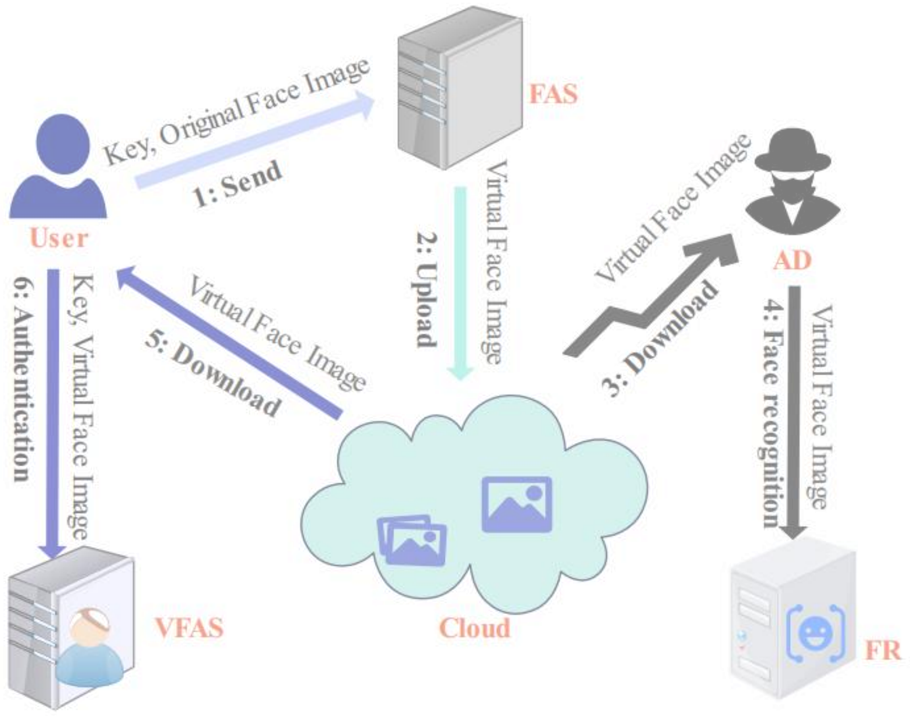
➢ Research Objective

❑ Overall Objective



➢ Research Objective

❑ Overall Objective



- Virtual faces anonymize human eyes and machines
- Anonymity, synchronism, diversity, differentiation, interactivity, realism, and recognizability
- Prevent unauthorized access

A key-driven face anonymization and authentication recognition (KFAAR) framework

➢ Methodology

Shanghai University

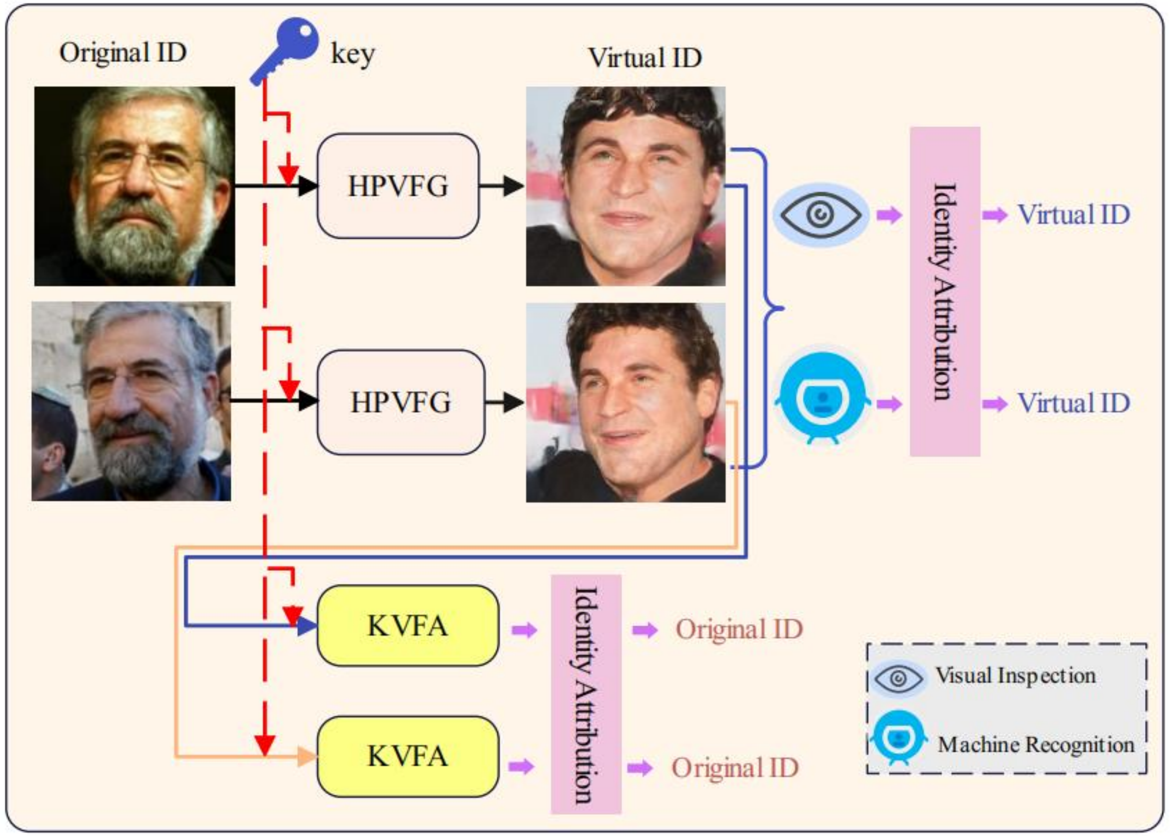


Illustration of the proposed framework

Head posture-preserving virtual face generation (HPVFG) module

Generate virtual faces that maintain head posture and facial expressions

The key-controllable virtual face authentication (KVFA) module

Identity authentication using the correct key

Methodology

Properties of virtual face

- **Anonymity**: The virtual face has a different identity from the original face.

$$R(\underbrace{G(x_1,k_1)}_{\text{The virtual face}}) \neq R(\underbrace{x_1}_{\text{The original face}})$$

R: Face recognizer
G: Virtual face generator

- **Synchronism**: Virtual faces generated from the same original identity belong to the same identity.

$$R(G(x_1,k_1)) = R(G(x_2,k_1))$$

x_1, x_2: Faces with the same identity collected at different times

- **Diversity**: The virtual identities generated by the same original face but using different keys should be different.

$$R(G(x_1,k_1)) \neq R(G(x_1,k_2))$$

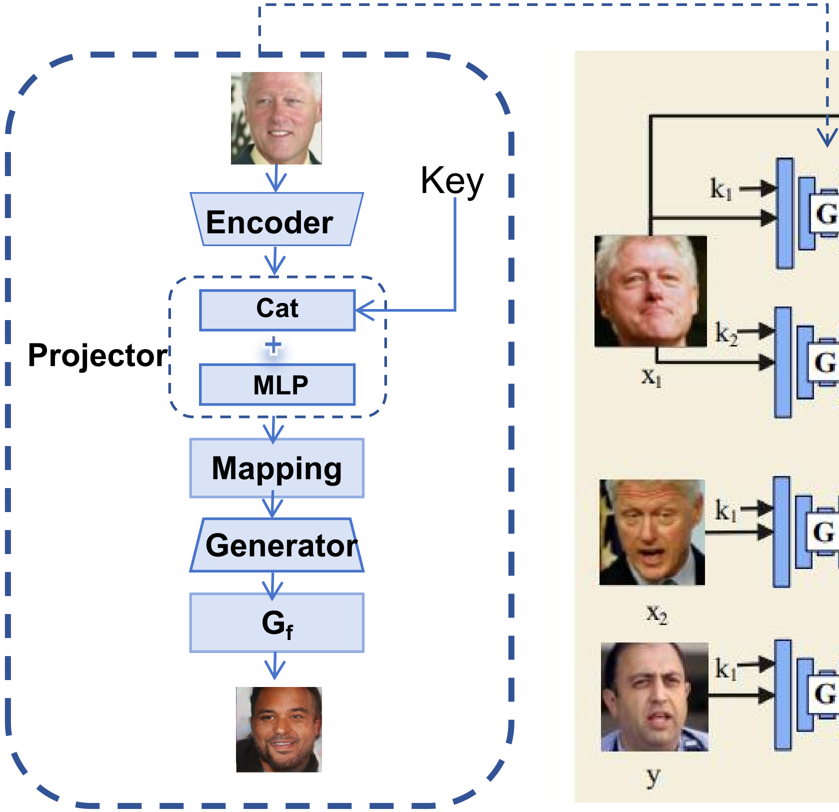
- **Differentiation**: For two original faces x and y which are from different identities, the generated virtual identities should be different even if they are derived from the same key.

$$R(G(x,k_1)) \neq R(G(y,k_1))$$

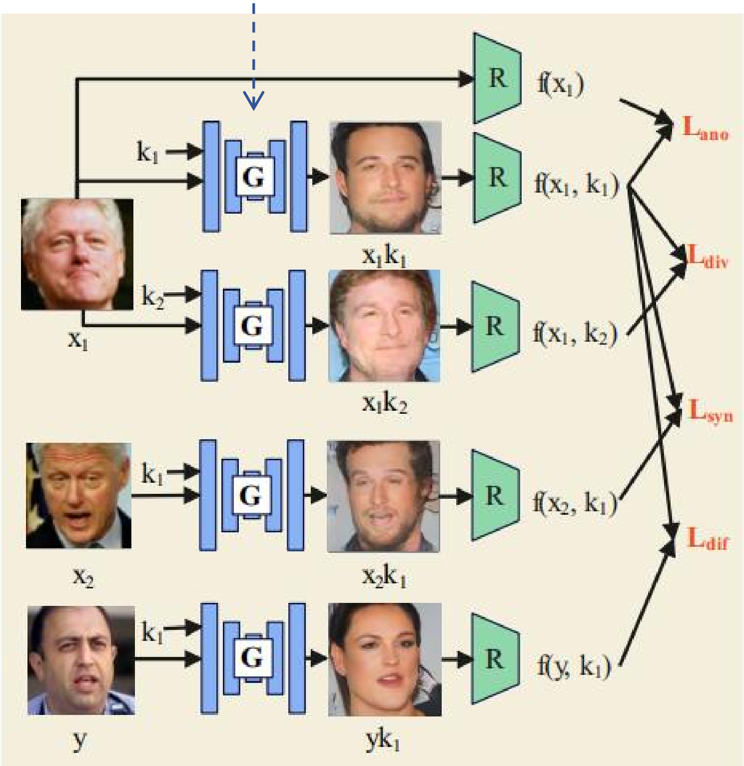
x, y: Faces are from different identities

➢ Methodology

❑ HPVFG



Network Architecture



Training Strategy

Anonymity Loss:

$$L_{ano} = L_{\cos}\left(R\left(G\left(x_1, k_1\right)\right), R\left(x_1\right), -1\right)$$

Synchronism Loss:

$$L_{syn} = L_{\cos}\left(R\left(G\left(x_1, k_1\right)\right), R\left(G\left(x_2, k_1\right)\right), 1\right)$$

Diversity Loss:

$$L_{div} = L_{\cos}\left(R\left(G\left(x_1, k_1\right)\right), R\left(G\left(x_1, k_2\right)\right), -1\right)$$

Differentiation Loss:

$$L_{dif} = L_{\cos}\left(R\left(G\left(x, k_1\right)\right), R\left(G\left(y, k_1\right)\right), -1\right)$$

➢ Methodology

上海大学 Shanghai University

Properties of KVFA

Prevent Misidentification

- The original face and the virtual face belong to different identities

$$I\left(x\right) \neq I\left(G\left(x, k_1\right)\right)$$

The original face   The virtual face

I : Face Identifier
G: Virtual face generator

- When using the wrong key

$$I\left(x\right) \neq I\left(G\left(x, k_1\right), k_2\right)$$

k_1, k_2 : Different keys

- Authenticating virtual faces from different identities with the correct key

$$I\left(G\left(x, k_1\right), k_1\right) \neq I\left(G\left(y, k_1\right), k_1\right)$$

x, y : Faces are from different identities

Methodology

Properties of KVFA

Prevent False Rejection

- When using the correct key for authentication

$$I(x) = I(G(x, k), k)$$

- When using the correct key for authentication, two different virtual faces from the same original identity

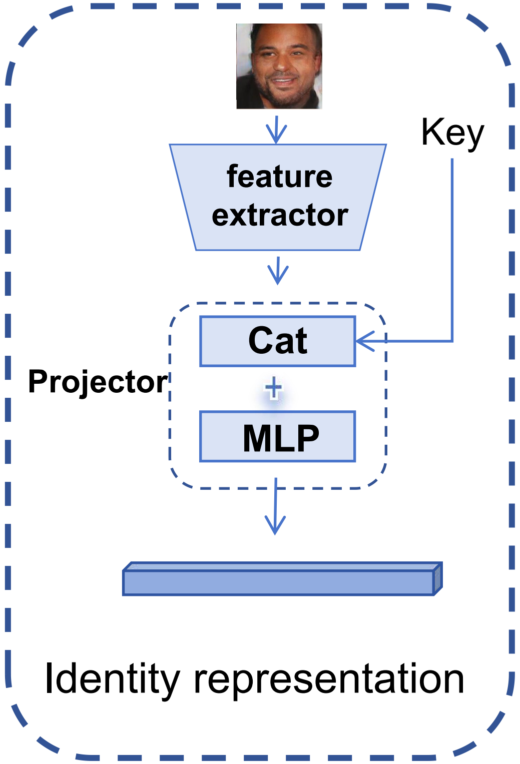
$$I(G(x_1, k_1), k_1) = I(G(x_1, k_2), k_2)$$

$$I(G(x_1, k_1), k_1) = I(G(x_2, k_2), k_2)$$

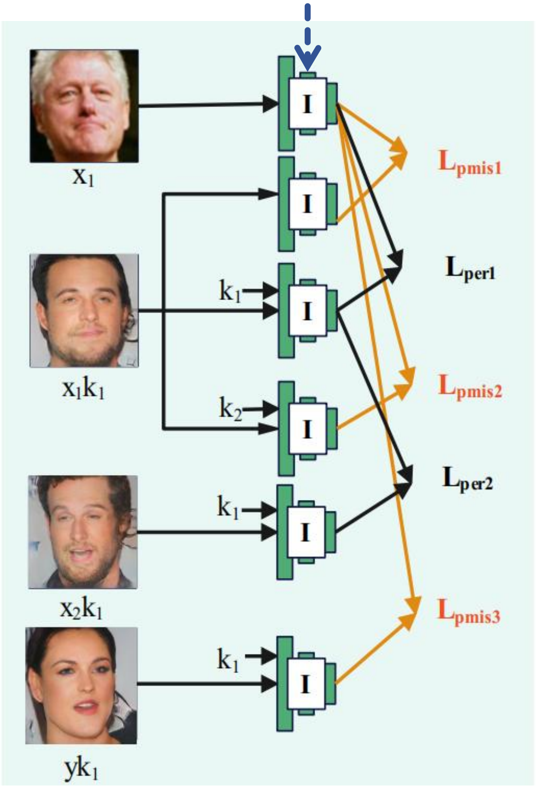
➢ Methodology

上海大学 Shanghai University

❑ KVFA



Network Architecture



Training Strategy

Prevent Misidentification Loss:

$$L_{pmis1} = L_{\cos}\left(I\left(G\left(x_1, k_1\right)\right), I\left(x_1\right), -1\right)$$

$$L_{pmis2} = L_{\cos}\left(I\left(G\left(x_1, k_1\right), k_2\right), I\left(x_1\right), -1\right)$$

$$L_{pmis3} = L_{\cos}\left(I\left(G\left(x, k_1\right), k_1\right), I\left(G\left(y, k_1\right), k_1\right), -1\right)$$

Prevent False Rejection Loss:

$$L_{per1} = L_{\cos}\left(I\left(G\left(x_1, k_1\right), k_1\right), I\left(x_1\right), 1\right)$$

$$L_{per2} = L_{\cos}\left(I\left(G\left(x_1, k_1\right), k_1\right), I\left(G\left(x_2, k_1\right), k_1\right), 1\right)$$

➢ Experiment

上海大学 Shanghai University

Evaluation of Virtual Faces

Anonymity and Diversity

Methods	Anonymity ↑		Diversity↑	
	LFW	CelebA	LFW	CelebA
CIAGAN [22]	0.674	0.0628	0.000	0.000
IVFG [36]	**0.988**	0.889	0.750	**0.787**
VFGM [33]	0.524	0.561	0.550	0.532
Ours	0.962	**0.922**	**0.783**	0.728

Synchronism and Detection Rate

Methods	AUC ↑		EER ↓		Detection Rate ↑	
	LFW	CelebA	LFW	CelebA	LFW	CelebA
CIAGAN [22]	-	-	-	-	0.986	0.998
IVFG [36]	0.929	0.933	0.103	0.122	1.0	1.0
VFGM [33]	0.889	0.920	0.182	0.139	1.0	1.0
Ours	**0.973**	**0.992**	**0.092**	**0.089**	**1.0**	**1.0**

Head Posture and Facial Expression

Methods	Yaw ↓		Pitch ↓		Roll ↓		Emotion ↑	
	LFW	CelebA	LFW	CelebA	LFW	CelebA	LFW	CelebA
CIAGAN [22]	3.913	3.152	4.387	4.424	3.037	3.291	0.584	0.622
VFGM [33]	3.557	3.623	**2.519**	2.333	4.011	3.997	0.726	0.689
IVFG [36]	25.663	28.537	19.334	18.261	17.991	17.309	0.433	0.412
Ours	**2.018**	**1.992**	3.102	**2.119**	**1.998**	**2.006**	**0.805**	**0.833**

Visual Quality

Methods	LFW	CelebA
CIAGAN [22]	7.64	6.90
VFGM [33]	6.99	6.91
IVFG [36]	**6.19**	**6.78**
Ours	7.29	6.82

Superior to existing SOTA methods in multiple metrics

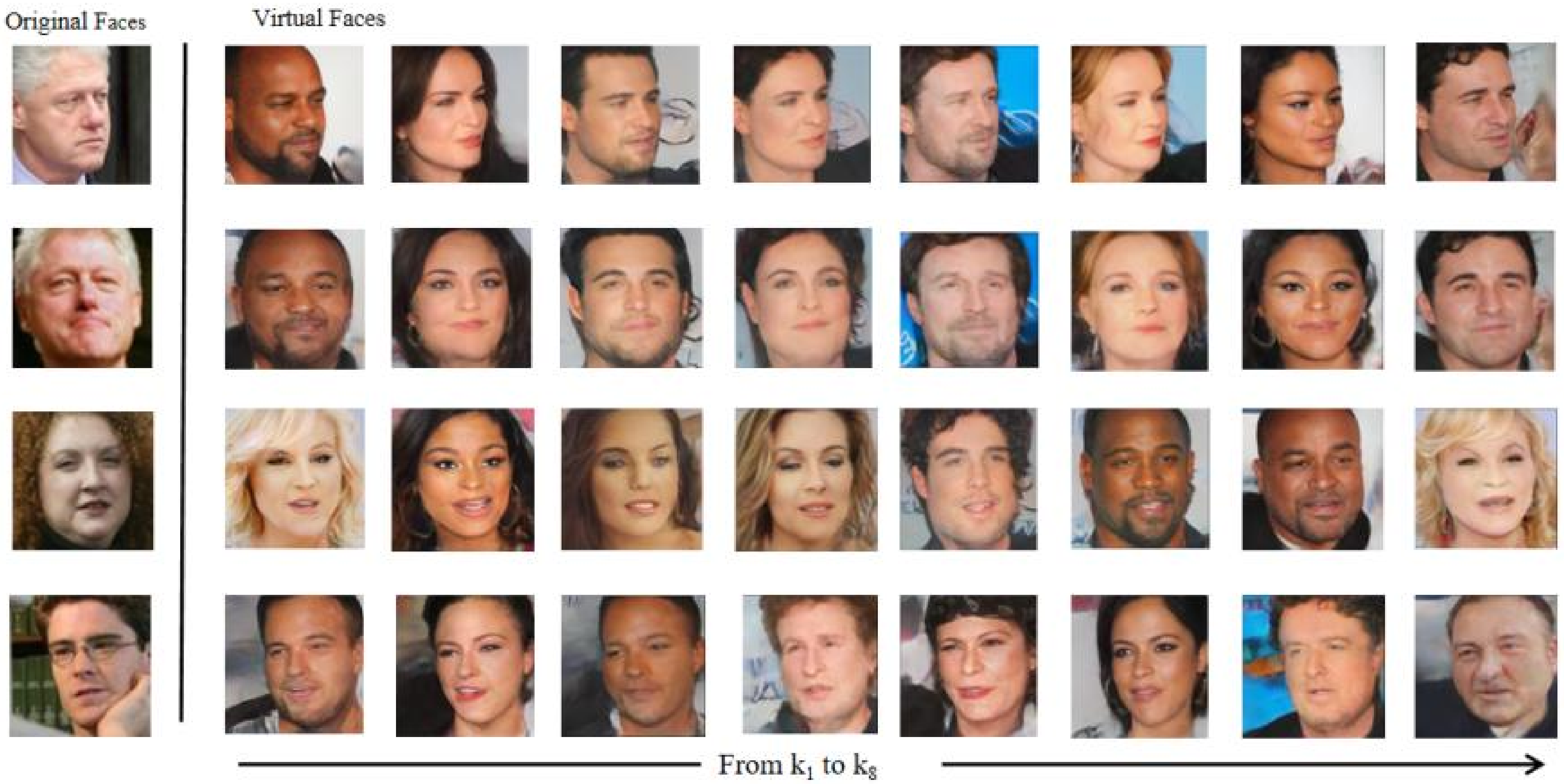
➢ Experiment

❑ Examples of the virtual faces.

➢ Experiment

上海大学
Shanghai University

❑ Examples of the virtual faces.



➢ Experiment

Evaluation of KVFA

Recognition Accuracy

Dataset	CRR↑	FAR↓	AUC ↑
LFW	0.927	0.078	0.984
CelebA	0.956	0.064	0.989

Different Authentication Scenarios

	LFW	CelebA
Scenario 1	0.498	0.516
Scenario 2	0.326	0.441
Scenario 3	0.522	0.398
Scenario 4	**0.852**	**0.799**

Cannot access KVFA

Bypasses the key

Use the wrong key

"In-the-wild" experiment

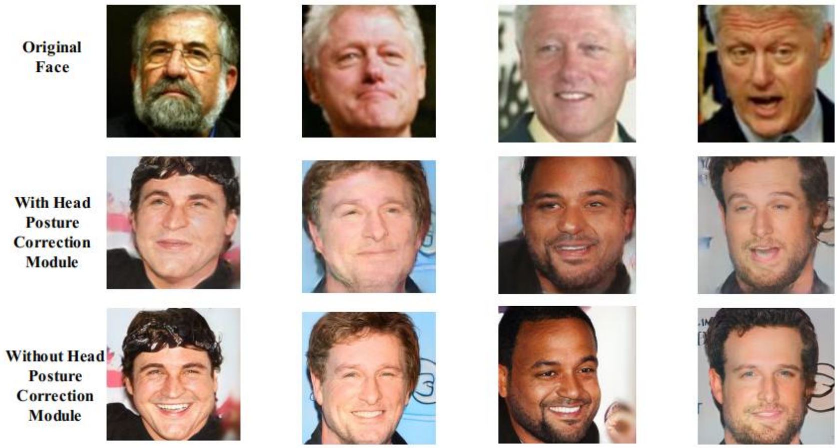
Datasets	Performance of the virtual faces				Recognition accuracy of KVFA	
	EER	Anonymity	Diversity	FID	CRR	FAR
LFW	0.092	0.962	0.783	7.29	0.927	0.078
FFHQ	0.089	0.931	0.707	6.36	0.887	0.081

- **Accurately authenticate the identity of a virtual face with the correct key**
- **Performed well on additional datasets**

Experiment

Ablation Studies

Head Posture Correction Module



Able to meet synchronization requirements

Each Loss of HPVFG

Methods	AUC	EER	Detection	Anonymity	Diversity	FID
w/o L_{ano}	0.914	0.091	1.0	0.908	0.779	6.86
w/o L_{syn}	0.897	0.104	1.0	0.974	0.787	7.73
w/o L_{div}	0.942	0.091	1.0	0.962	0	7.91
w/o L_{dif}	0.917	0.095	1.0	0.953	0.790	8.04
Ours	0.973	0.092	1.0	0.962	0.783	7.29

The performance is the best when all losses participate in training

Each Loss of KVFA

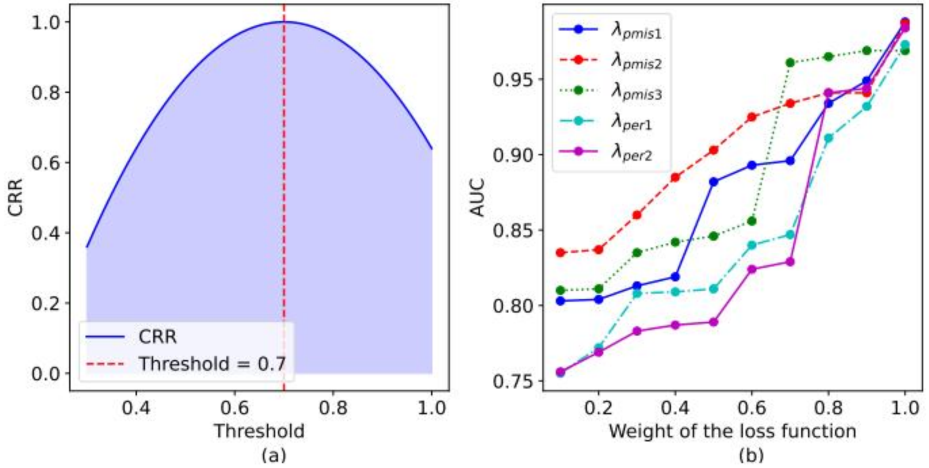
Methods	CRR	FAR	AUC
w/o L_{tot1}	0.756	0.298	0.833
w/o L_{tot2}	0.636	0.270	0.744
Ours	0.945	0.068	0.987

The performance is the best when all losses participate in training

➢ Experiment

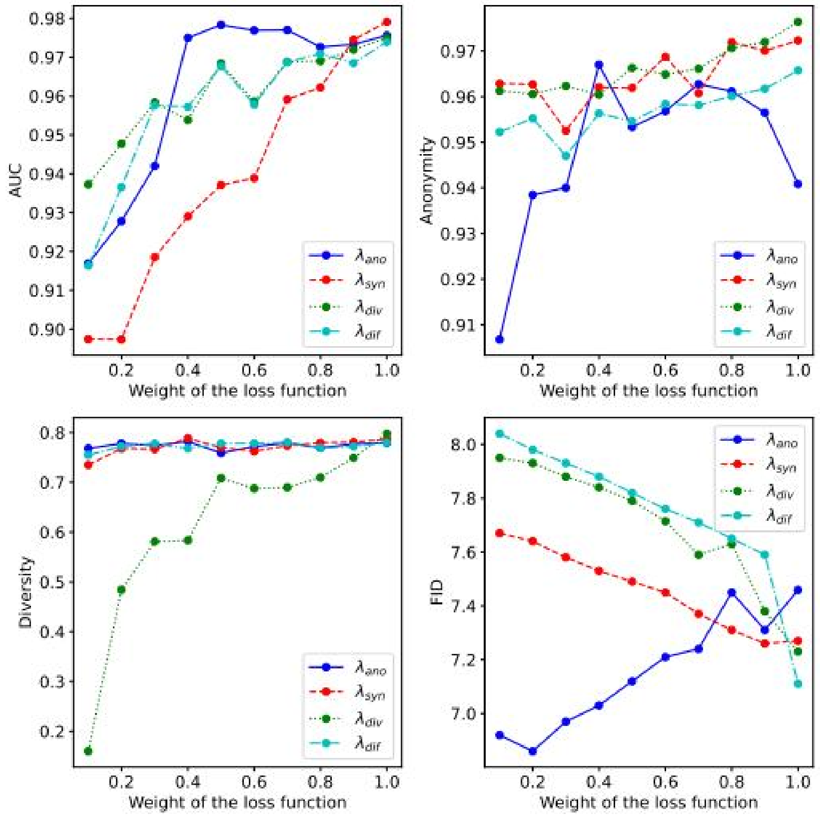
Ablation Studies

Threshold Setting for KVFA



- The best performance is achieved when the threshold is set to 0.7
- The performance of KVFA loss function is best when the weight is all set to 1

Each Loss of HPVFG



When λ_{ano}=0.4 and other weights are 1, the performance of virtual faces reaches a good balance

➢ Experiment

Performance Analysis of Key

Key Generation

Use the secure random number generator KeyGen to generate the key on the client side, ensuring that the key is unique and not easily obtained by adversaries

Key Storage

The key is stored only on the client side, and FAS deletes the key after generating the virtual face

Fault tolerance of key

key length	Key error bit				
	0 bit	1 bit	3 bits	5 bits	16 bits
8 bits	0.811	0.708	0.622	0.492	–
128 bits	0.803	0.691	0.640	0.523	0.522
256 bits	0.862	0.702	0.603	0.595	0.485

The impact of key length

key length	Anonymization	FID
8 bits	0.962	7.29
128 bits	0.956	**6.95**
256 bits	**0.964**	7.02

Conclusion

Performance Analysis of Key

Contributions

- Propose HPVFG to generate anonymous, synchronized, diverse, and high-quality virtual faces.
- Develop the KVFA module for authenticating virtual faces' original identities with the correct key.
- Successfully harmonize privacy with recognizability in virtual faces.

Future direction

- Enhance security by fortifying protection mechanisms, researching detection and prevention strategies against adversary model training, and bolstering system security.

- Streamline key strategies to fix random key issues and stabilize privacy.