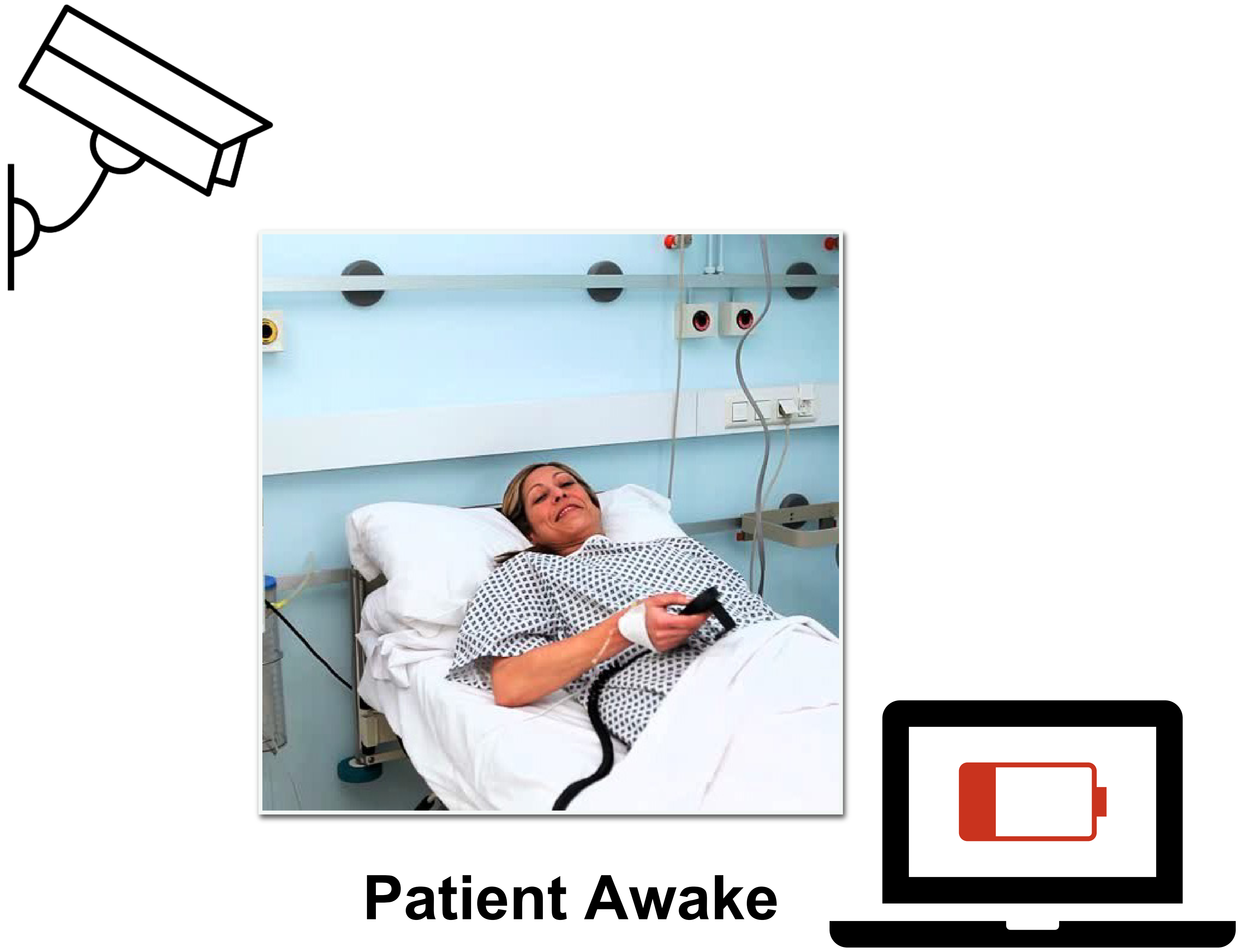
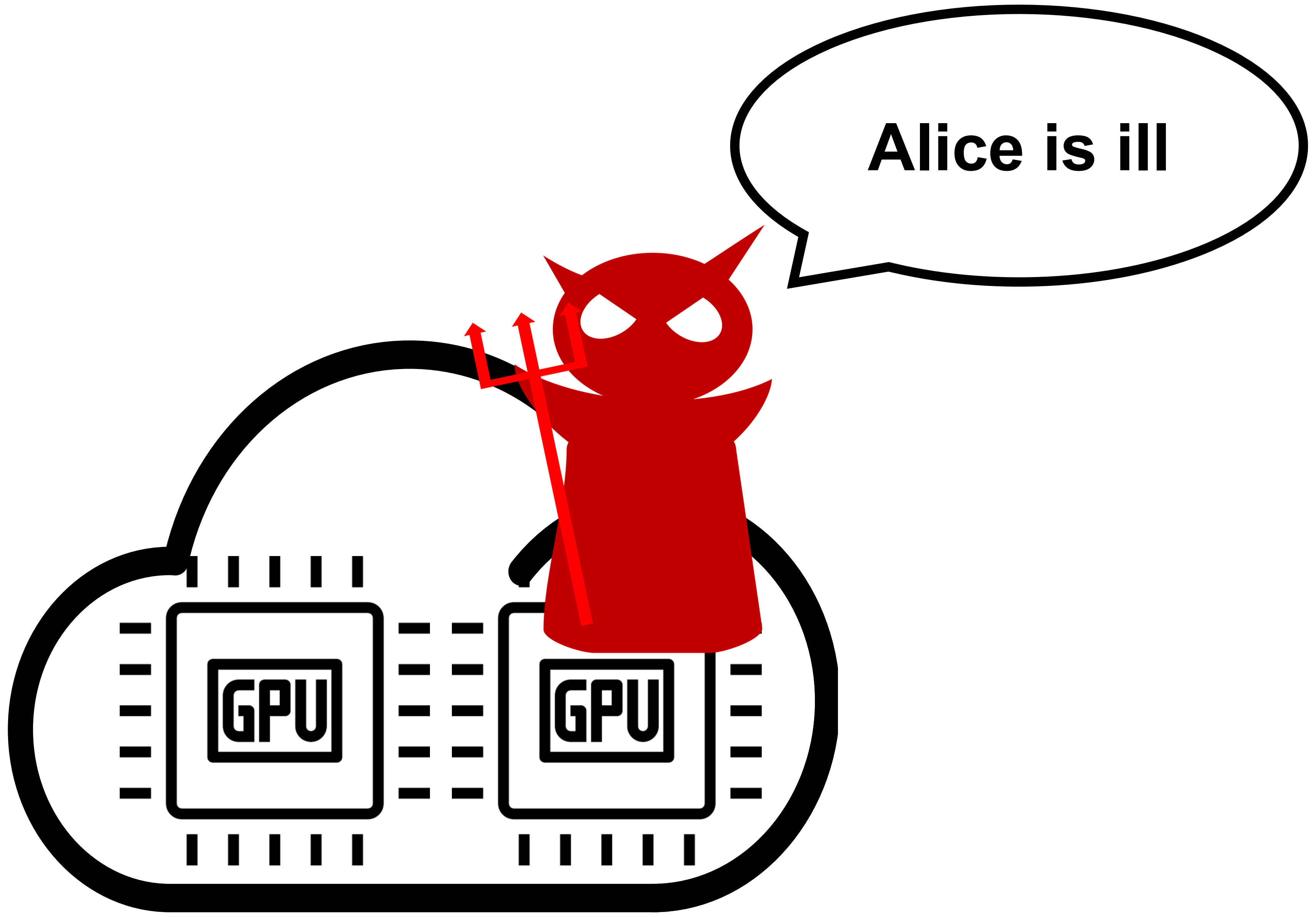
FaceMix: Privacy-Preserving Face Attribute Classification on the Cloud

Zhijian Liu*, **Zhanghao Wu***, **Ligeng Zhu**, **Chuang Gan**, and **Song Han**

Massachusetts Institute of Technology

Motivation: DNNs on the Edge Devices





Solution I: **Inference on the Edge**

Privacy is well-preserved.

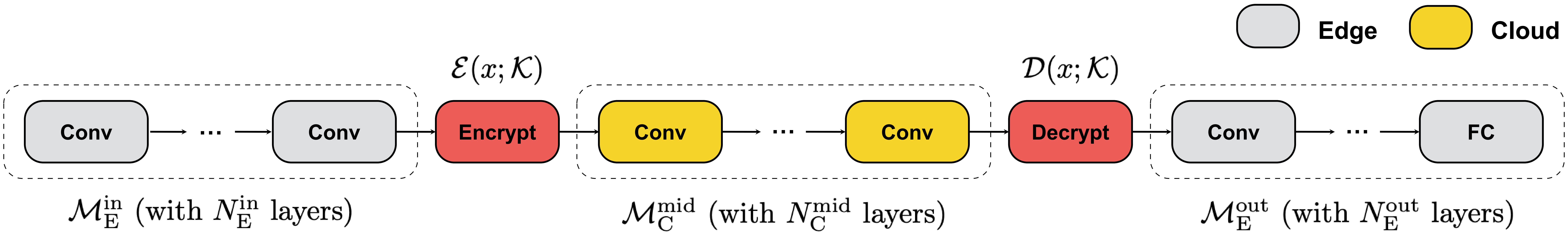
Computation is expensive.

Solution II: **Inference on the Cloud**

Privacy is compromised.

Computation is affordable.

Privacy-Preserving Edge-Cloud Inference

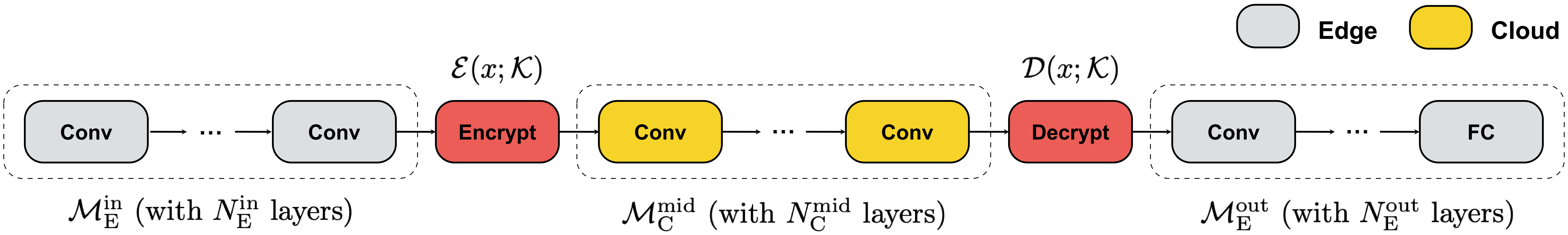


Properties required for encryption and decryption function:

Property I: They should not be **invertible** (without knowing the private key).

Property II: They should be **compatible** with the neural network model.

Privacy-Preserving Edge-Cloud Inference



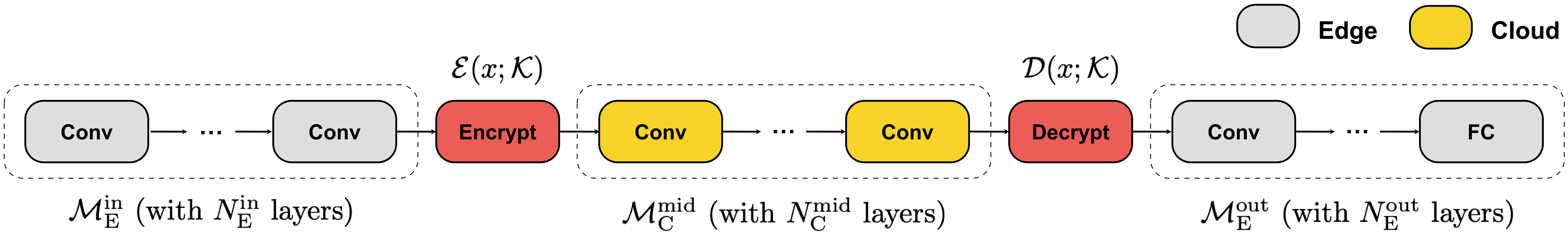
Given a batch of input data $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]$ and a linear model $\mathbf{M}$

$$\mathbf{M}(a\mathbf{x}_1 + b\mathbf{x}_2) = a\mathbf{M}(\mathbf{x}_1) + b\mathbf{M}(\mathbf{x}_2)$$

$$\mathbf{M}(c\mathbf{x}_1 + d\mathbf{x}_2) = c\mathbf{M}(\mathbf{x}_1) + d\mathbf{M}(\mathbf{x}_2)$$

$\mathbf{M}(\mathbf{x}_1)$ and $\mathbf{M}(\mathbf{x}_2)$ can be solved from $\mathbf{M}(a\mathbf{x}_1 + b\mathbf{x}_2)$ and $\mathbf{M}(c\mathbf{x}_1 + d\mathbf{x}_2)$

Privacy-Preserving Edge-Cloud Inference

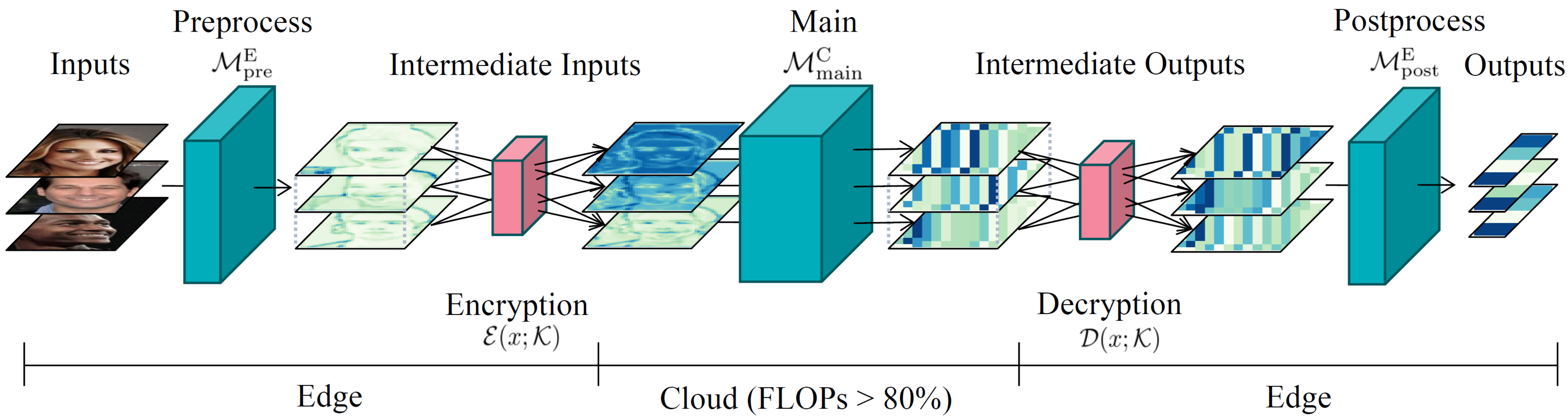


Private Key: a random coefficient matrix A

Encryption: linearly combine the batched inputs: $\mathbf{y} = A\mathbf{x}$

Decryption: solve the linear equations to obtain the outputs: $\mathbf{y} = A^{-1}\mathbf{x}$

Privacy-Preserving Edge-Cloud Inference



Compared with other edge-cloud frameworks

	Baseline (cloud)	Osia *et al.* [41, 42]	Tramer *et al.* [50]	Ours	
Computation (on edge)	0%	93%	$\sim$1%*	1%	13%
Transmission size	0.6 MB	0.4 MB	86.5 MB	12.3 MB	3.0 MB
Transmission time	0.4 sec	0.3 sec	33.1 sec	6.6 sec	1.7 sec
GPU utilization (cloud)	100%	100%	$\sim$10%*	100%	100%
Input privacy	×	✓	✓	✓	✓
Output privacy	×	×	✓	✓	✓

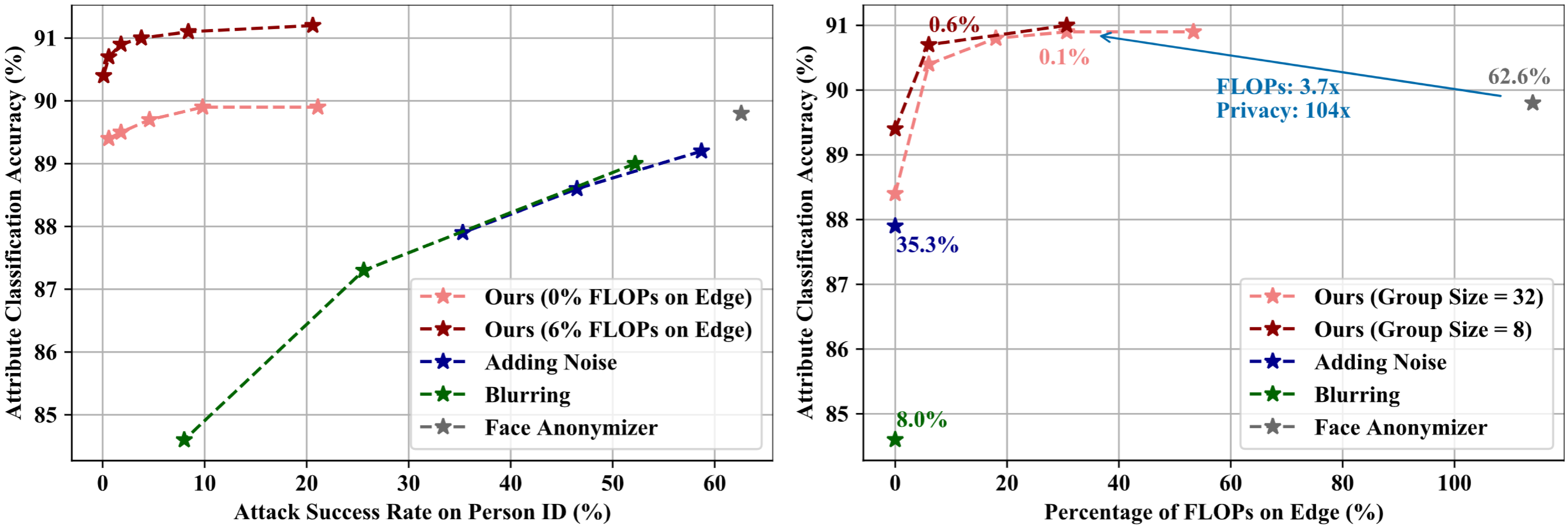
Table 1: Our framework achieves high *efficiency* on the edge, introduces small *network communication overhead*, attains full *resource utilization* on the cloud, and protects both *input and output privacy*. All benchmarks are conducted on VGG-16 [48] with input image size of 224×224, and the transmissions are over the 4G LTE network with upload speed of 15 Mbps, download speed of 30 Mbps, and delay of 25 ms. As for our framework, we send the output activation of the first or second convolution layer to the cloud (the last two columns). In this table, the red entries are unsatisfactory.

Results on Facial Attribute Classification

	Efficiency (↓ is better)		Accuracy (↑ is better)		Privacy (↓ is better)	
	Edge Params	Edge FLOPs	Valid Acc.	Test Acc.	Person ID	Face Attrs.
Baseline (all on Edge)	11.21M	1.50G	91.6%	91.0%	0.1%	50.0%
Baseline (all on Cloud)	0	0	91.6%	91.0%	85.5%	79.3%
Adding Noise $\mathcal{N}(0,4)$ [43]	0	0	89.2%	88.6%	46.5%	73.1%
Adding Noise $\mathcal{N}(0,8)$ [43]	0	0	88.5%	87.9%	35.3%	70.8%
Blurring (16×16) [43, 45]	0	0	89.6%	89.0%	52.2%	73.1%
Blurring (8×8) [43, 45]	0	0	87.9%	87.3%	25.6%	68.7%
Face Anonymizer [43]	11.38M	47.13G	90.5%	89.8%	62.6%	76.3%
FaceMix (Ours) ($S_G = 8$, $N_{pre} = 1$)	**0.05M**	**0.09G**	91.2%	90.7%	**0.6%**	**51.5%**
FaceMix (Ours) ($S_G = 8$, $N_{pre} = 2$)	0.12M	0.28G	91.2%	90.7%	**0.6%**	**51.6%**
FaceMix (Ours) ($S_G = 8$, $N_{pre} = 3$)	0.20M	0.46G	**91.4%**	**91.0%**	**0.6%**	**51.5%**

Table 2: Privacy-preserving facial attribute classification on CelebA. The red entries are unsatisfactory (efficiency: the fewer FLOPs the better; privacy: the lower attack success rate the better). We require fewer computations on the edge, while maintaining higher accuracy and lower attack success rate.

Results on Facial Attribute Classification

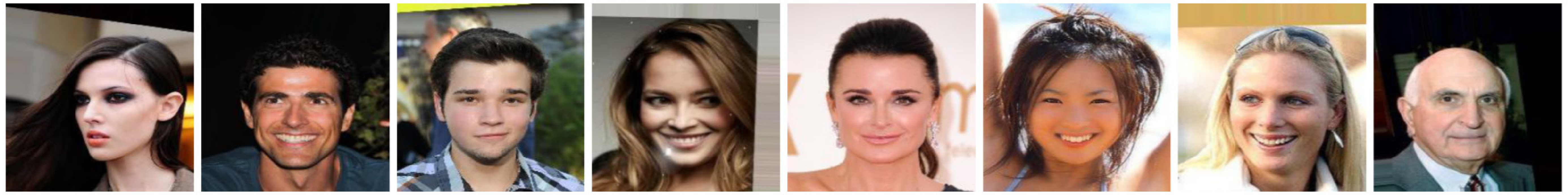


(a) Accuracy *vs*. Privacy

(b) Accuracy *vs*. Efficiency

Attack Results on Faces from Encryptions

Raw

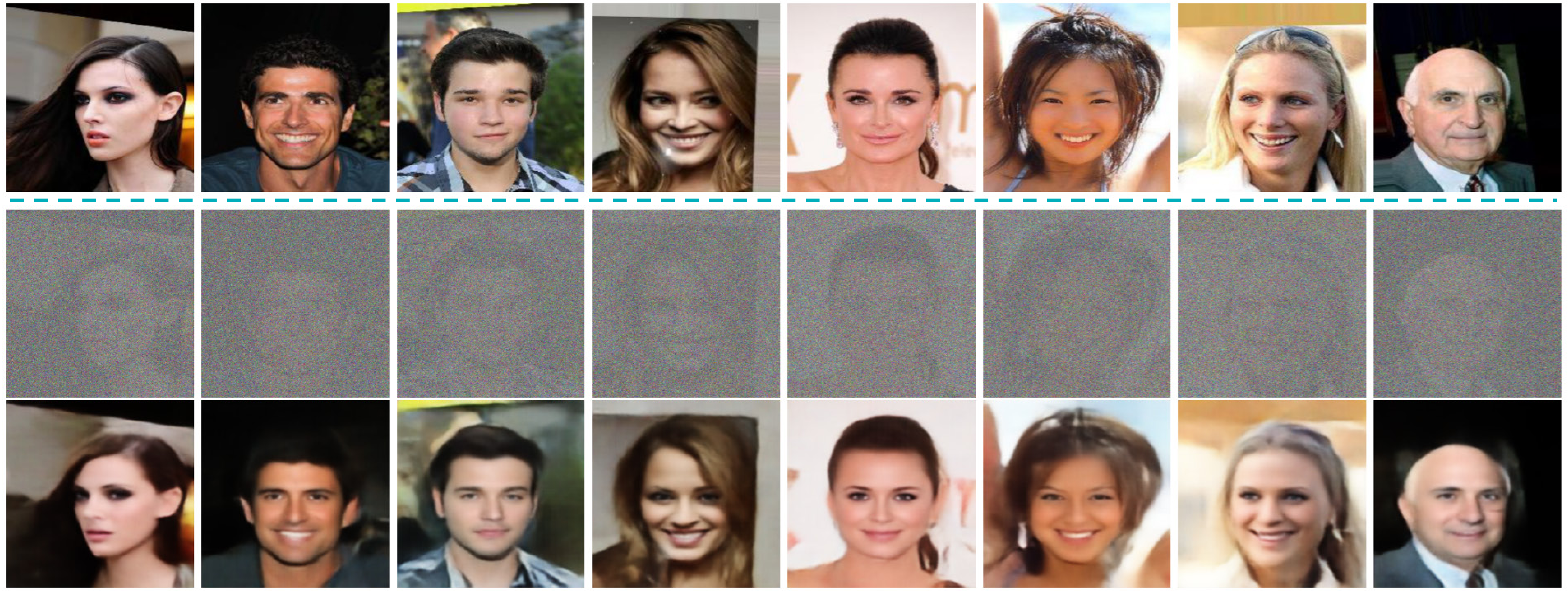


Attack Results on Faces from Encryptions

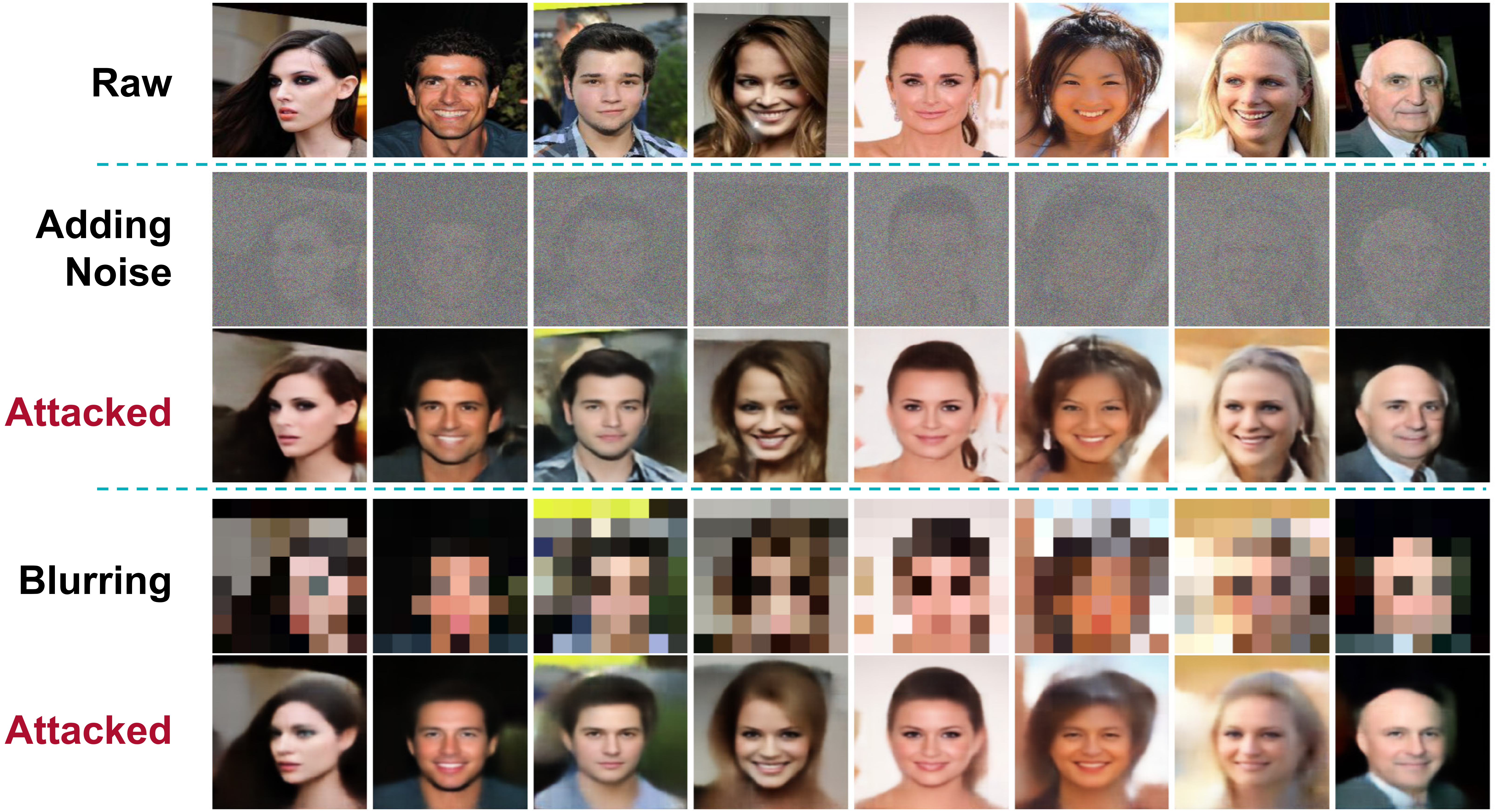
Raw

Adding Noise

Attacked



Attack Results on Faces from Encryptions



Attack Results on Faces from Encryptions

Raw

Adding Noise

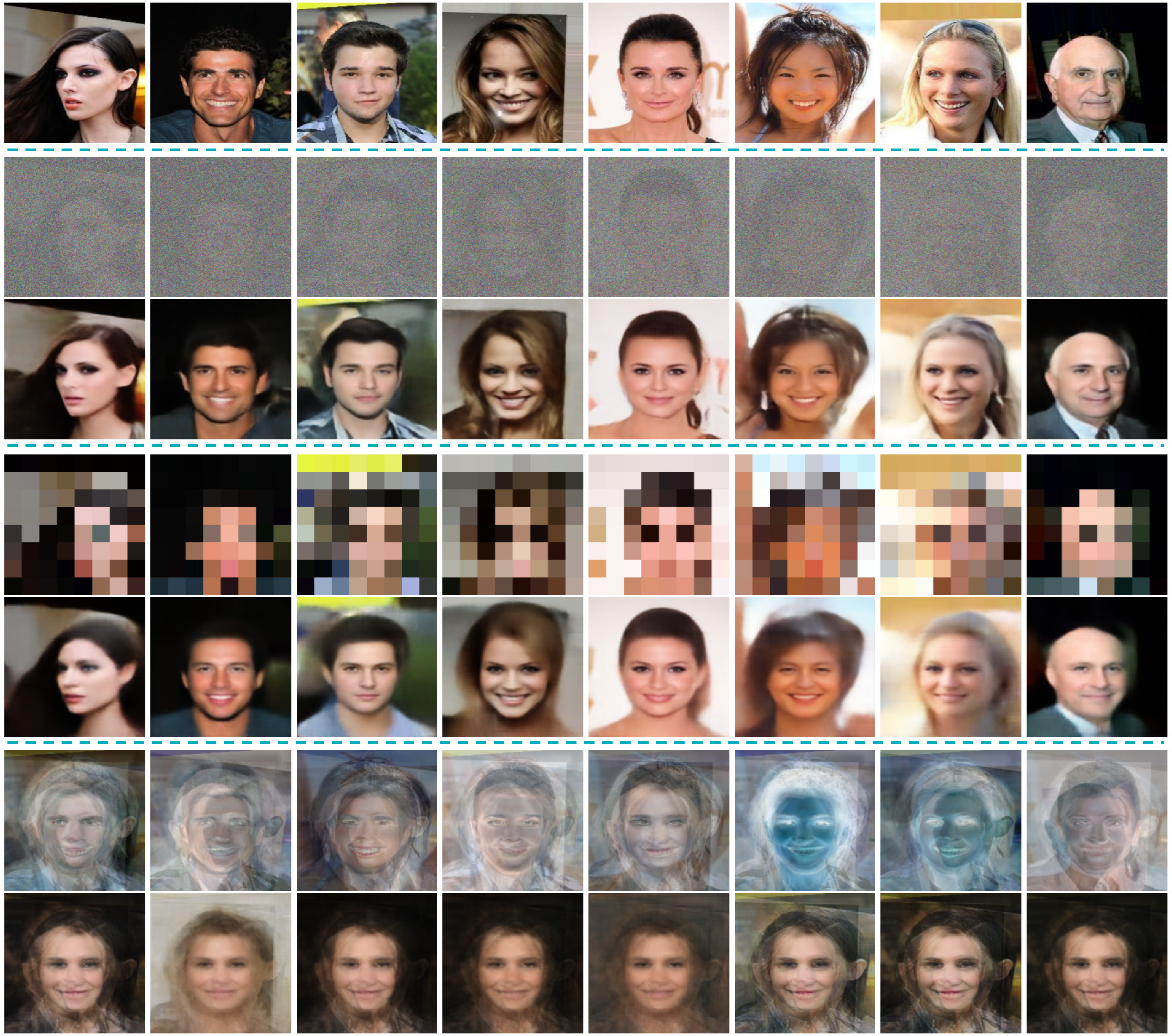
Attacked

Blurring

Attacked

Ours

Defended

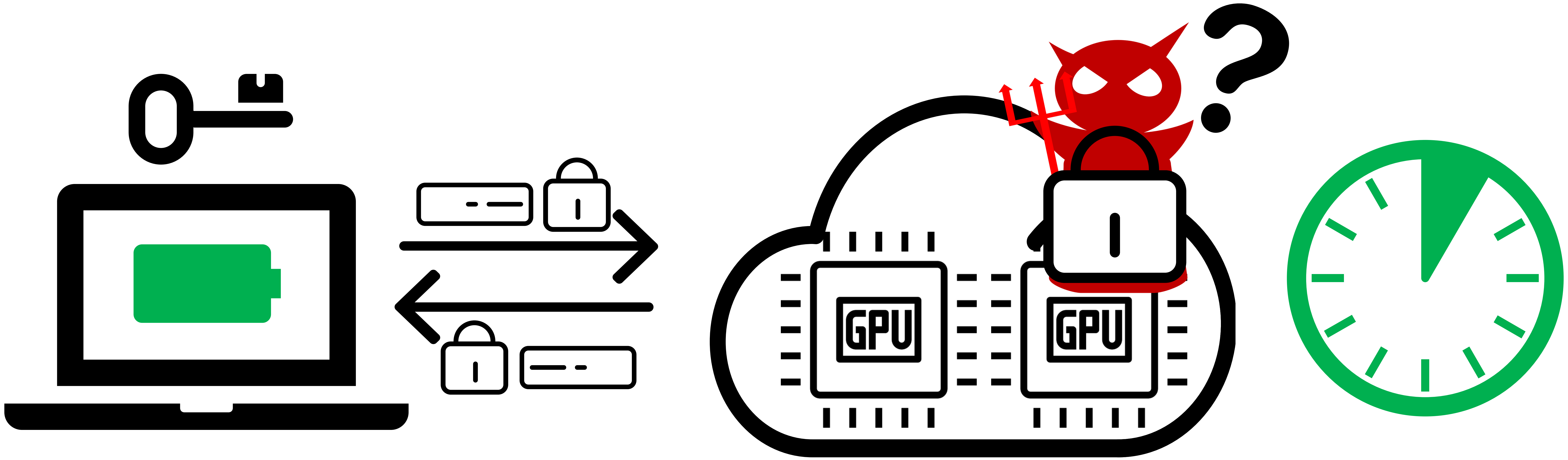


Additional Results on Facial Attribute Classification

	Efficiency (↓ is better)		**Accuracy** (↑ is better)		**Privacy** (↓ is better)	
	Edge Params	Edge FLOPs	Test Acc.	Bal. Acc.	Face Recon.	Face Attrs.
Baseline (all on Edge)	11.21M	1.50G	91.1%	87.1%	-0.56	50.0%
Baseline (all on Cloud)	0	0	91.1%	87.1%	-0.00	87.1%
Adding Noise $\mathcal{N}(0,4)$ [43]	0	0	88.5%	82.5%	-0.03	82.7%
Adding Noise $\mathcal{N}(0,8)$ [43]	0	0	87.7%	81.3%	-0.02	81.4%
Blurring (16×16) [43, 45]	0	0	88.8%	83.4%	-0.03	83.6%
Blurring (8×8) [43, 45]	0	0	87.0%	80.6%	-0.07	77.8%
FaceMix (Ours) ($S_G = 8$, $N_{pre} = 1$)	**0.05M**	**0.09G**	90.5%	86.8%	**-0.37***	**50.6%**
FaceMix (Ours) ($S_G = 8$, $N_{pre} = 2$)	0.12M	0.28G	90.7%	86.9%	**-0.37***	**50.7%**
FaceMix (Ours) ($S_G = 8$, $N_{pre} = 3$)	0.20M	0.46G	**90.7%**	**87.1%**	**-0.37***	**50.6%**

Table 3: Privacy-preserving facial attribute classification on LFWA. Bal. Acc. denotes the balanced accuracy on the test set and Face Recon. represents the inverse mean square error of the reconstructed images (GAN-based) with the raw inputs. *The GAN-based attack model is applied on the encrypted input image without the *preprocessing* model for fair comparison.

Future Works




Language


Speech

Thank You!