ScrambleMix: A Privacy-Preserving Image Processing for Edge-Cloud Machine Learning

Koki Madono [1], Masayuki Tanaka [2,3], Masaki Onishi [2]
Waseda University, AIST, Tokyo Institute of Technology

Edge Cloud Machine Learning

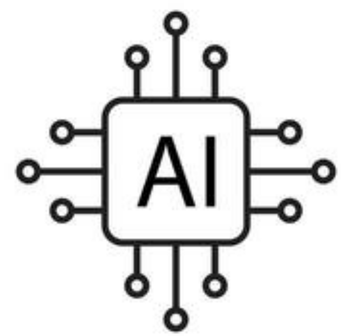
Use **Cloud AI model** for prediction

Edge side

Cloud side

Personal Image

Want to know tower name

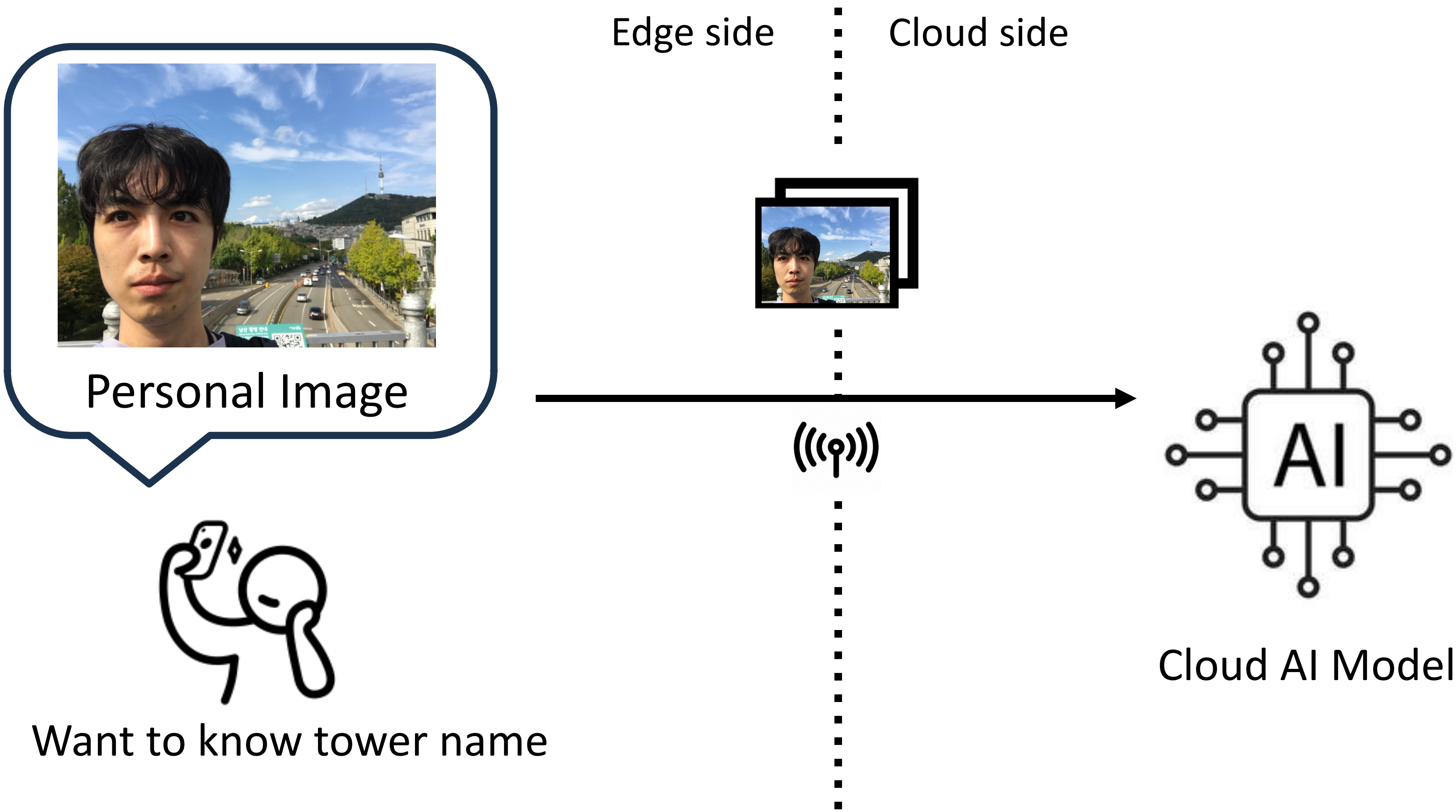


Cloud AI Model

Edge Cloud Machine Learning

Use Cloud AI model for prediction

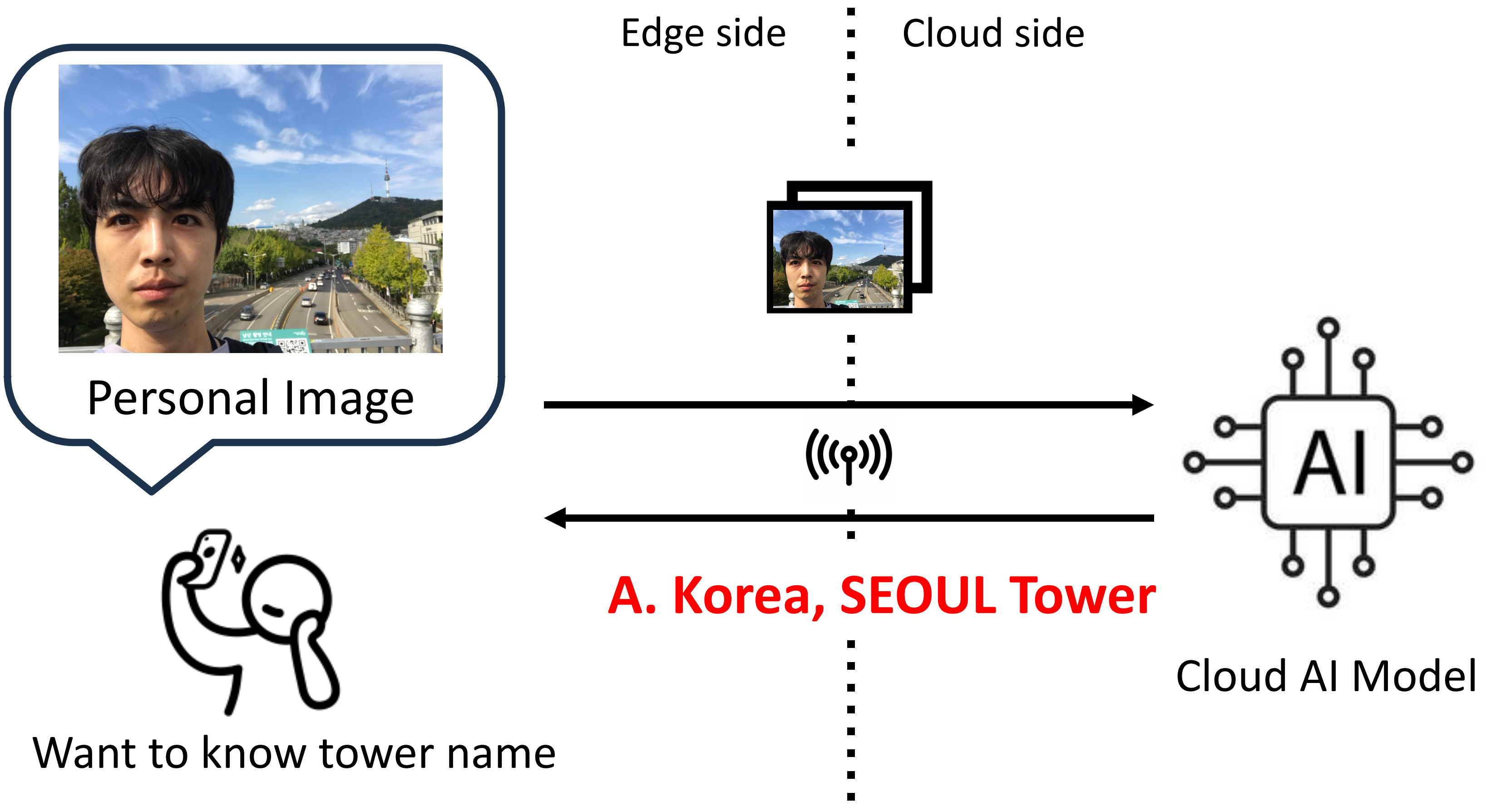
1. sending the data



Edge Cloud Machine Learning

Use Cloud AI model for prediction

1. sending the data
2. **receive inference results**



Problem

Personal image is dangerous to send public network

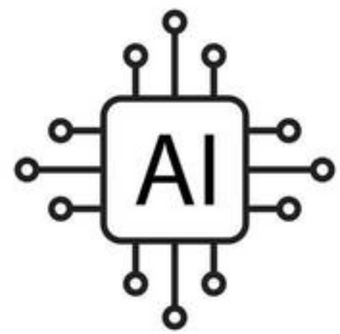
Edge side

Cloud side

Personal Image

Steal image

Anonymous Hacker



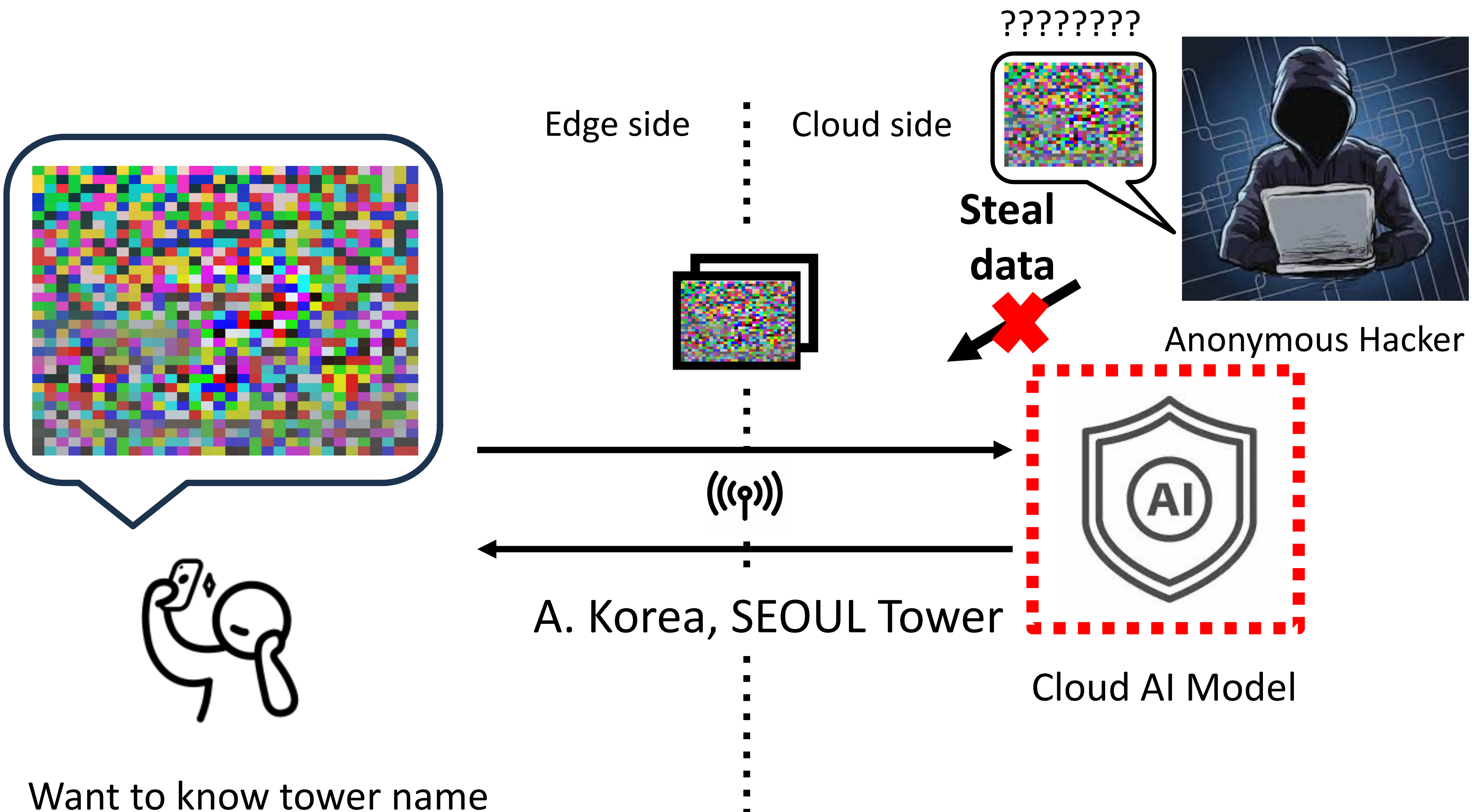
A. Korea, SEOUL Tower

Cloud AI Model

Want to know tower name

High Level Solution

Use AI understandable Image Encryption



High Level Solution

Use AI understandable Image Encryption

[approaches]

(1) DataMix[Liu et al]

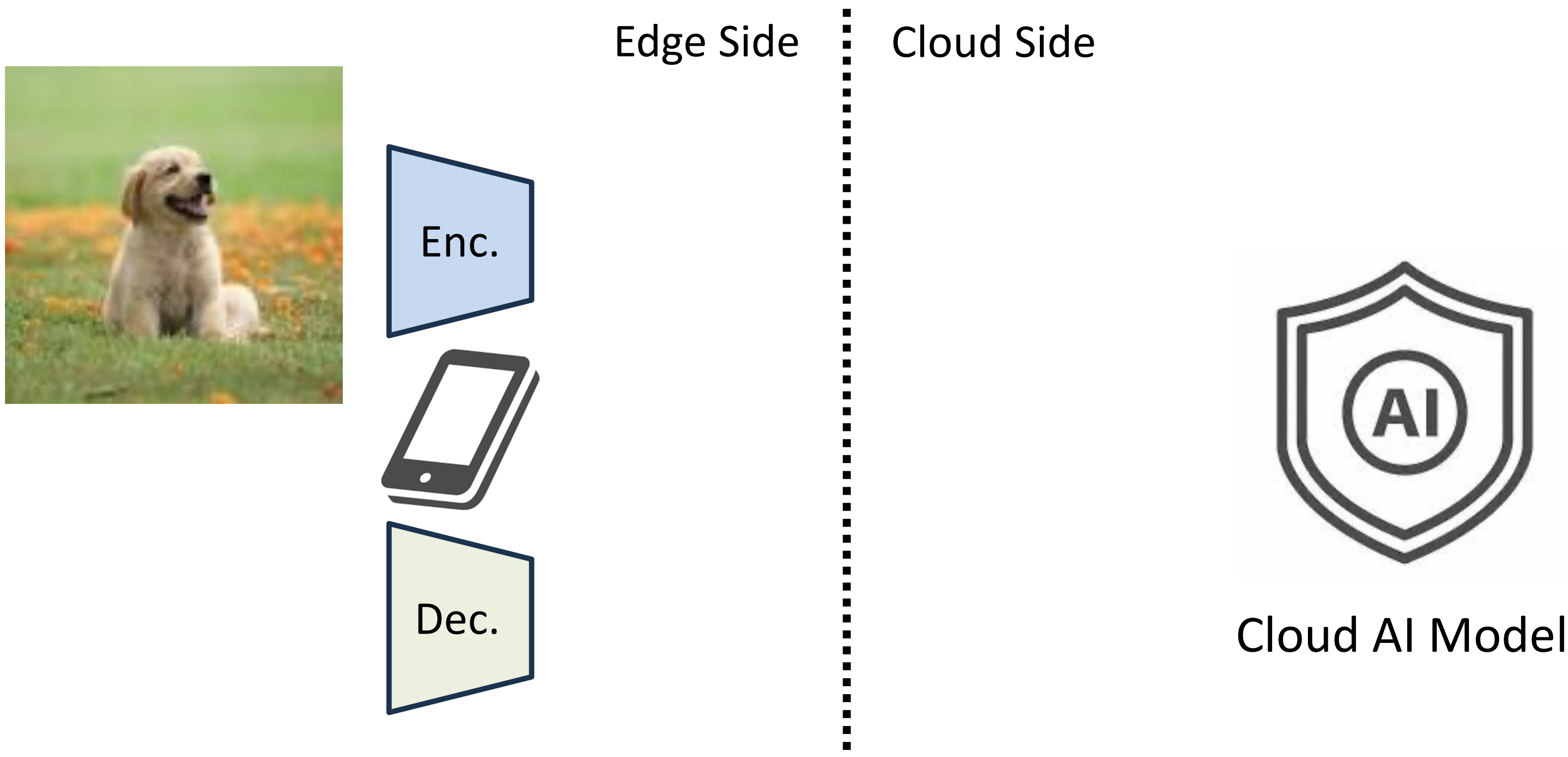
(2) InstaHide[Haung et al]

(3) Image Scrambling

(4) ScrambleMix (extension of Image Scrambling)
- Our approach

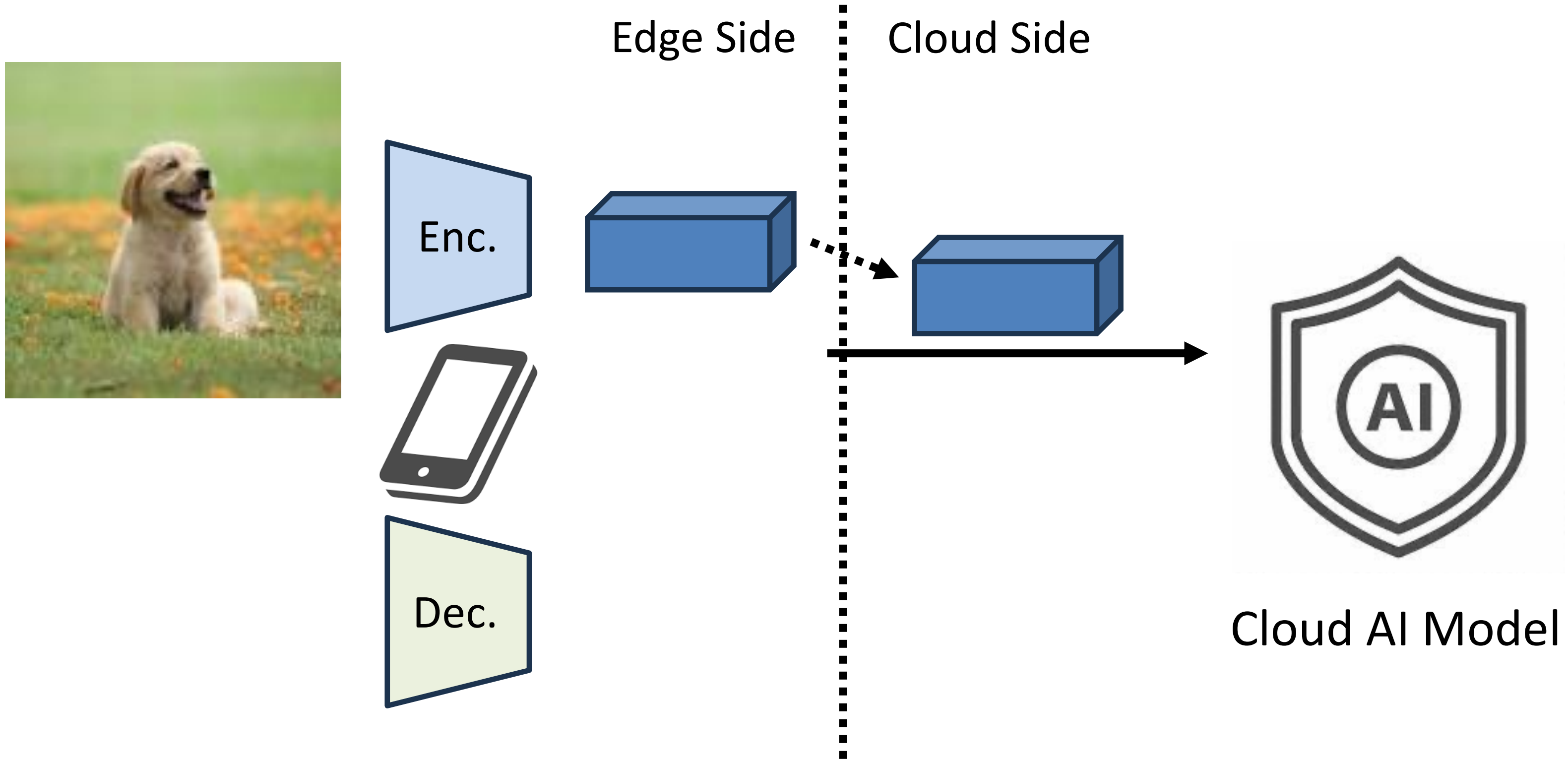
DataMix [Liu et al]

(0) Deploy Encoder/decoder on edge device and AI on cloud side



DataMix [Liu et al]

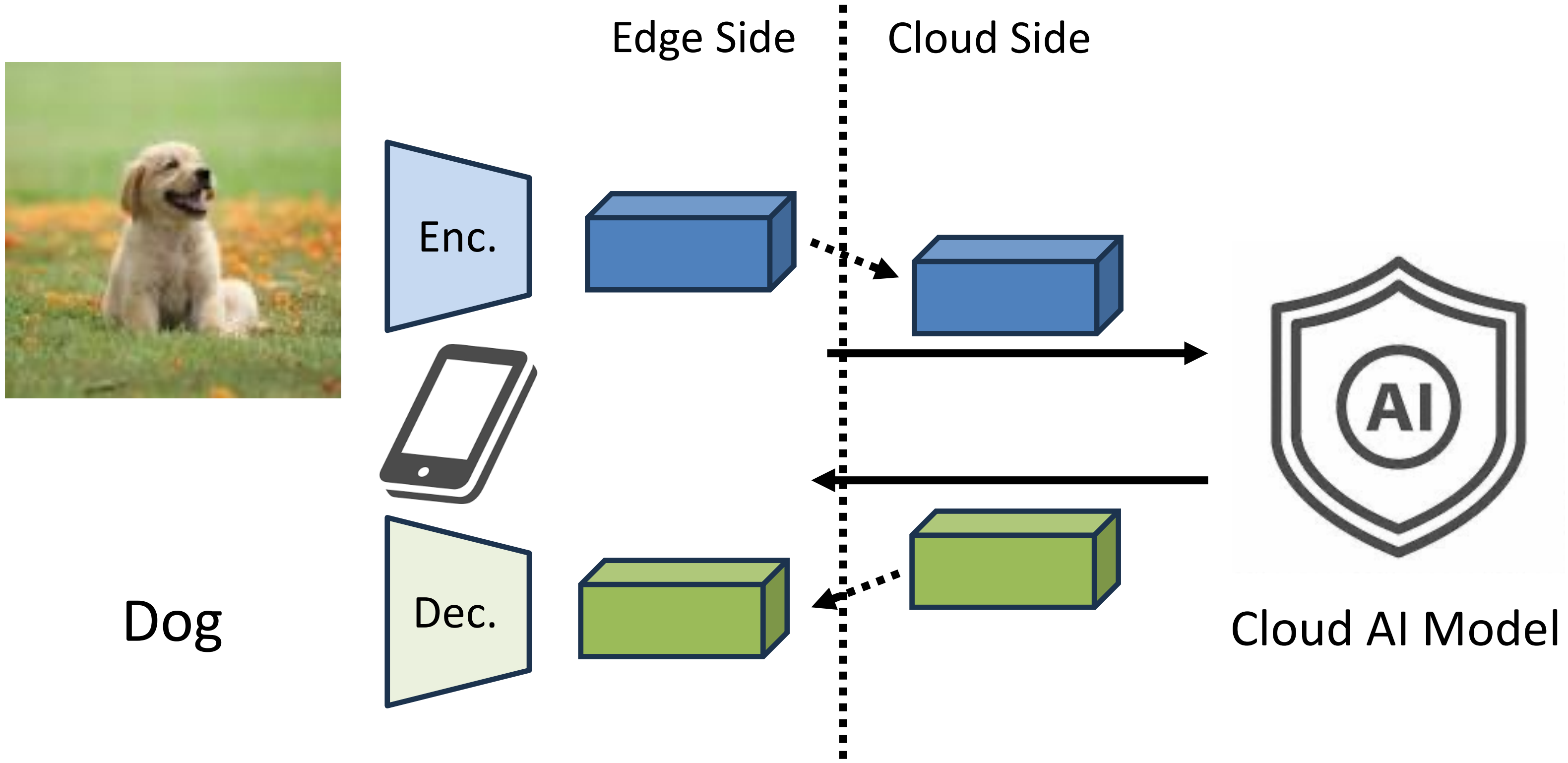
(1) Send encoded feature to the server



DataMix [Liu et al]

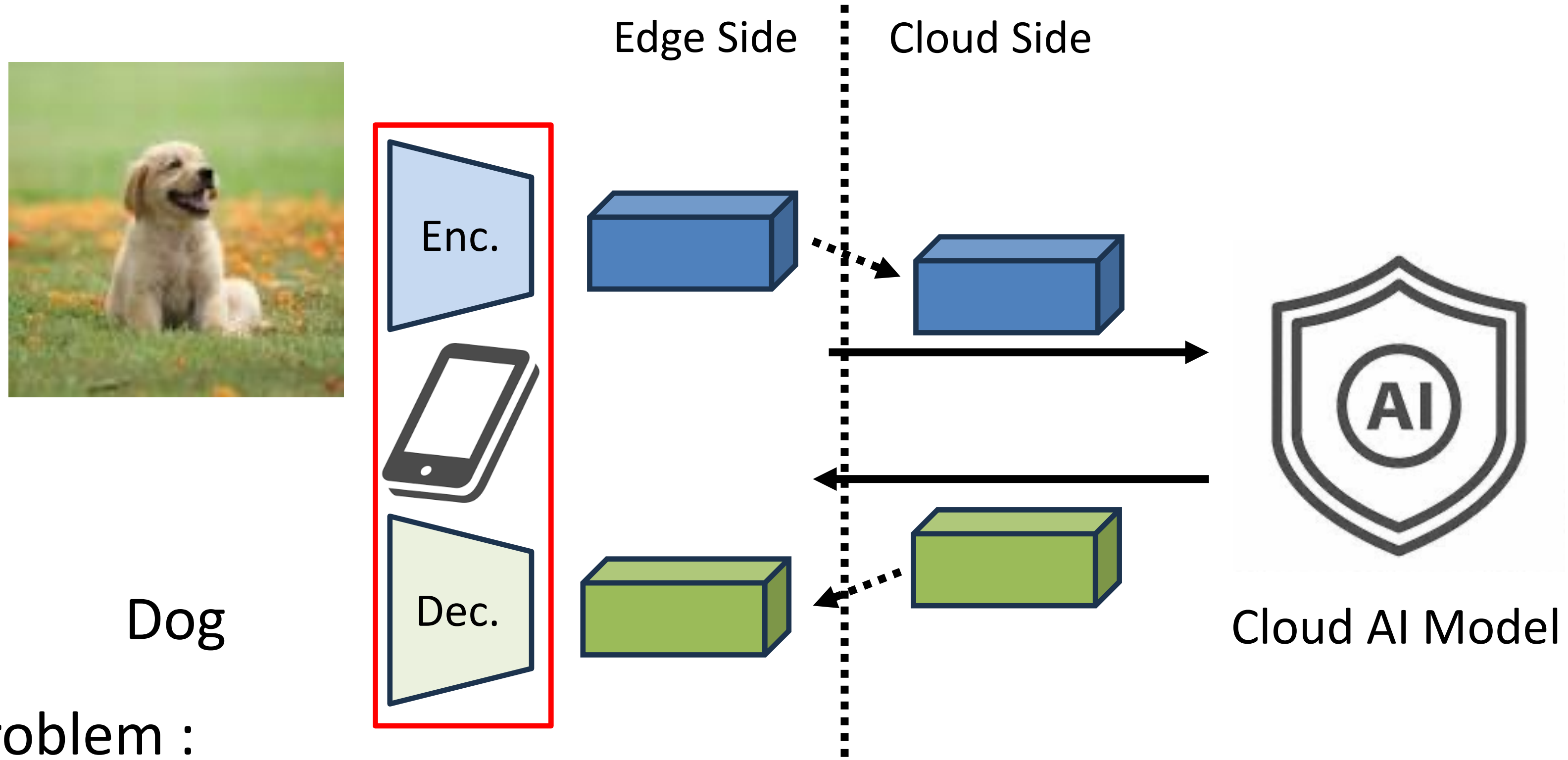
(1) Send encoded feature to the server

(2) Received feature and decode message.



DataMix [Liu et al]

(1) Send encoded feature to the server
(2) Received feature and decode message.

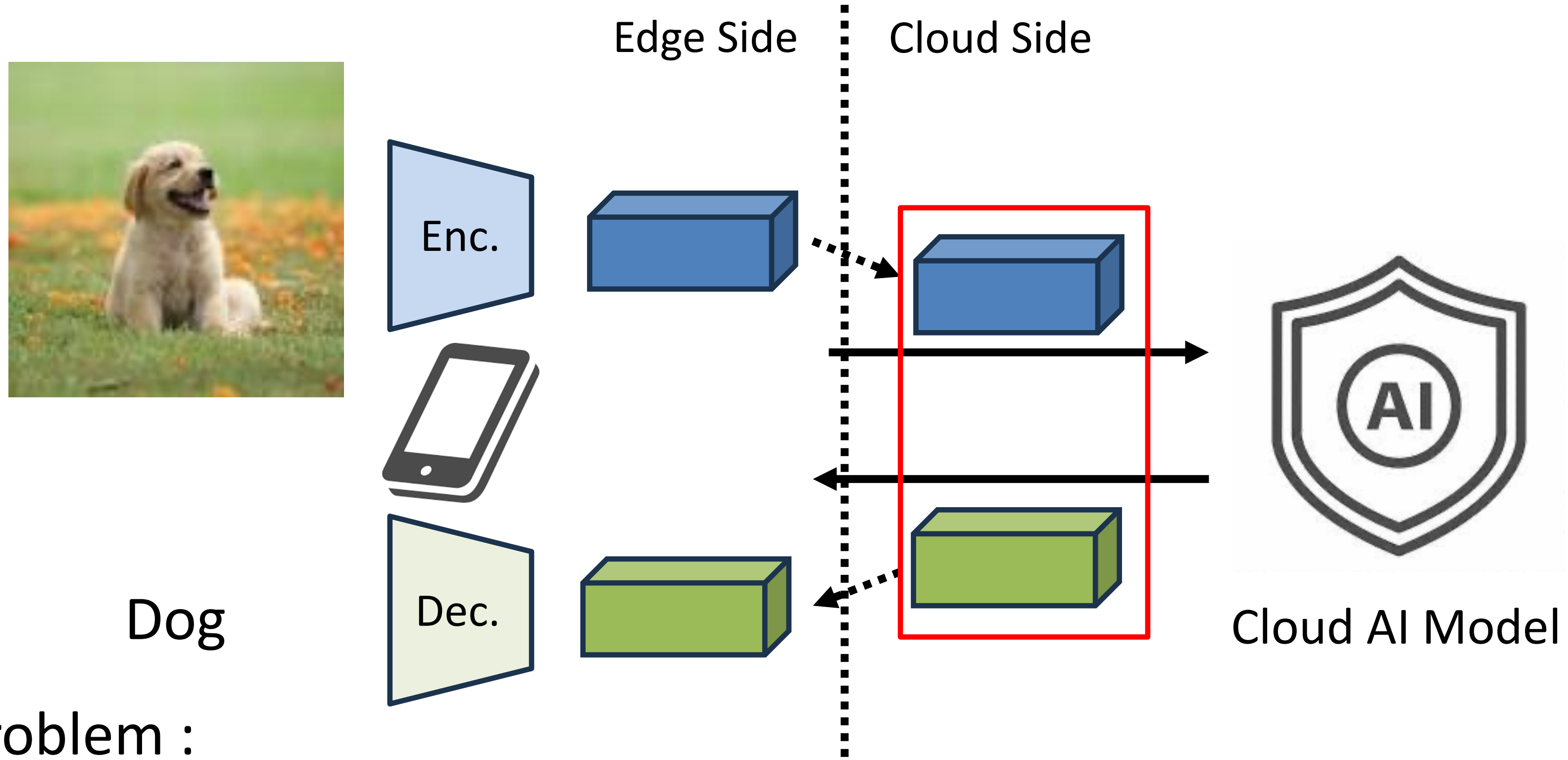


Problem :

(1) Encoder/Decoder are necessary

DataMix [Liu et al]

(1) Send encoded feature to the server
(2) Received feature and decode message.



Problem :
(1) Encoder/Decoder are necessary
(2) Feature limits accuracy.

InstaHide [Haung et al]

(0) Prepare two imaegs & Cloud AI model

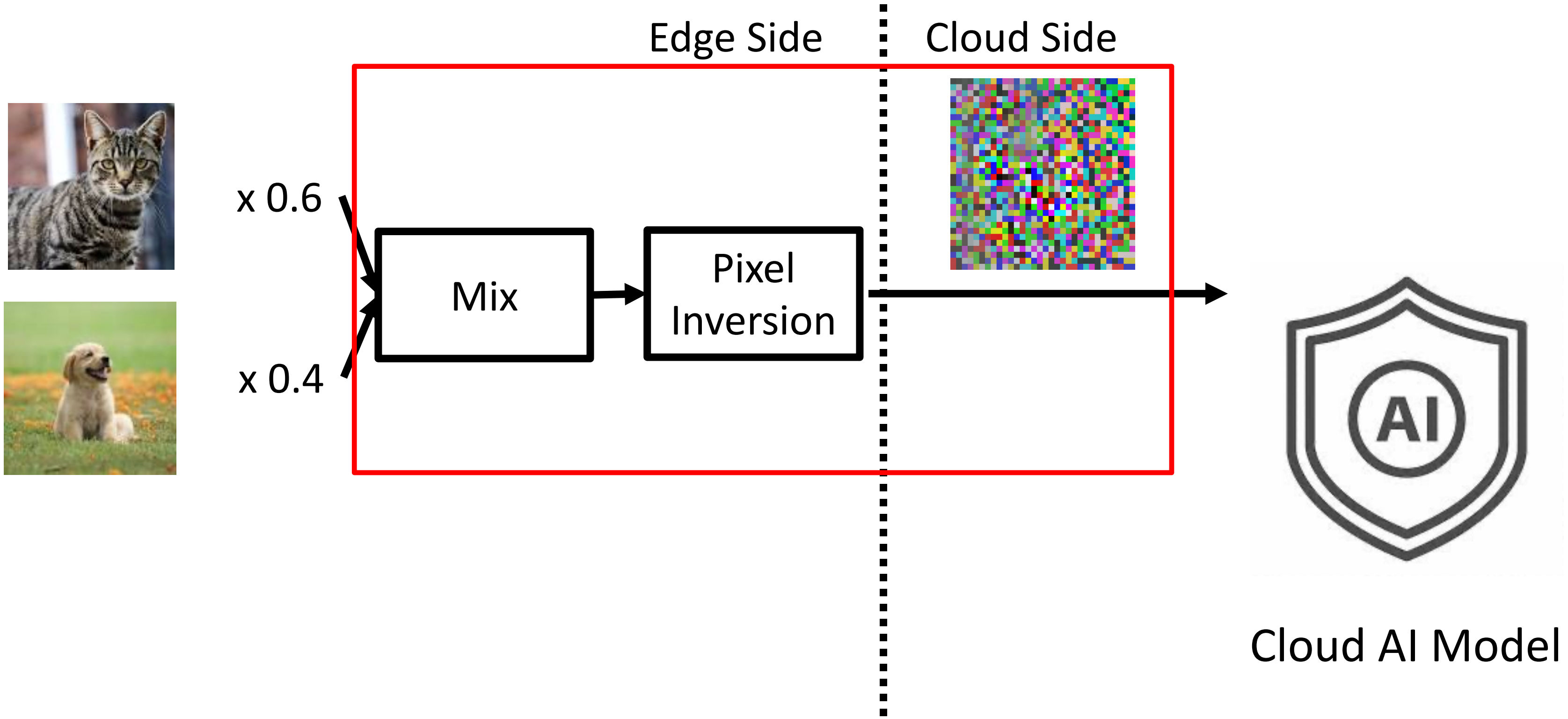
Edge Side

Cloud Side

Cloud AI Model

InstaHide [Haung et al]

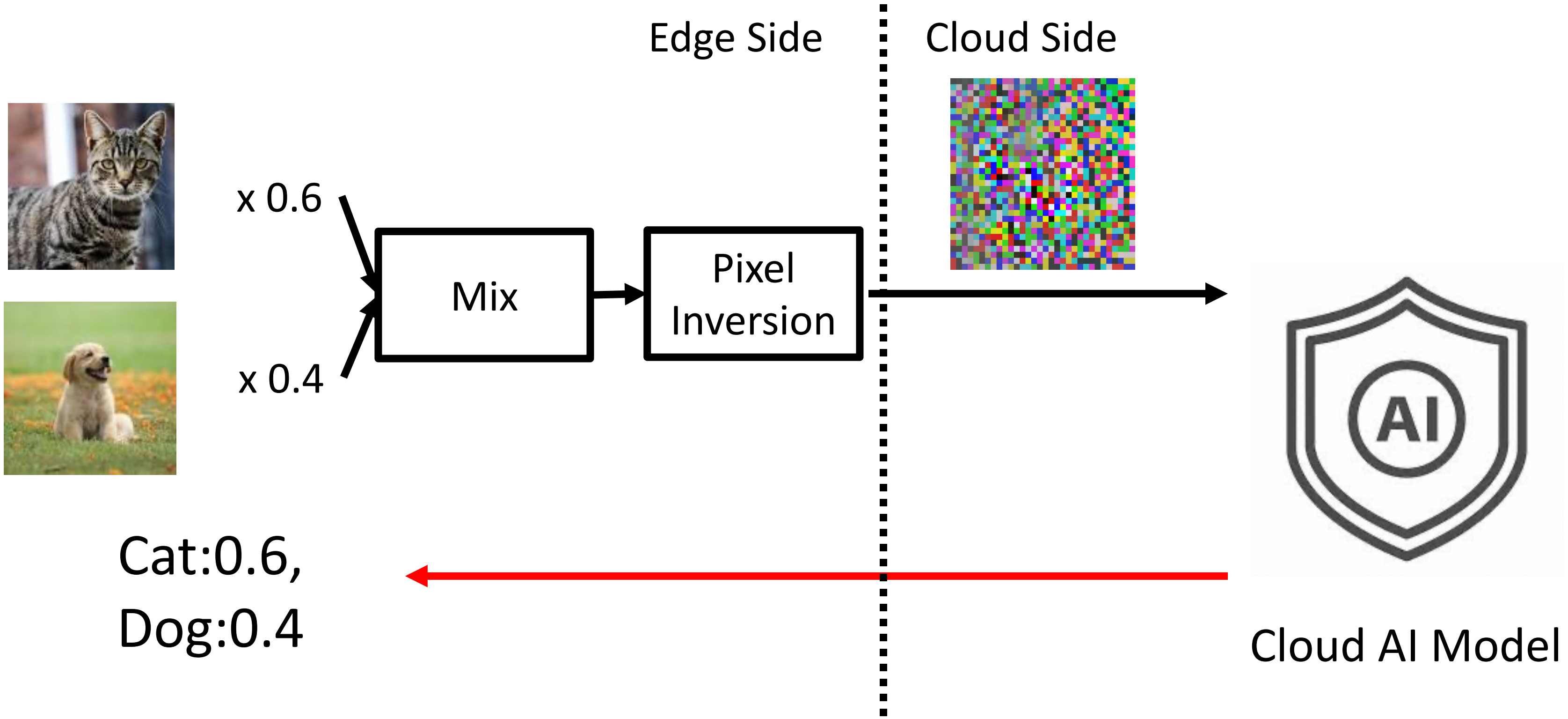
(1) Mix Images and Encrypt images



InstaHide [Haung et al]

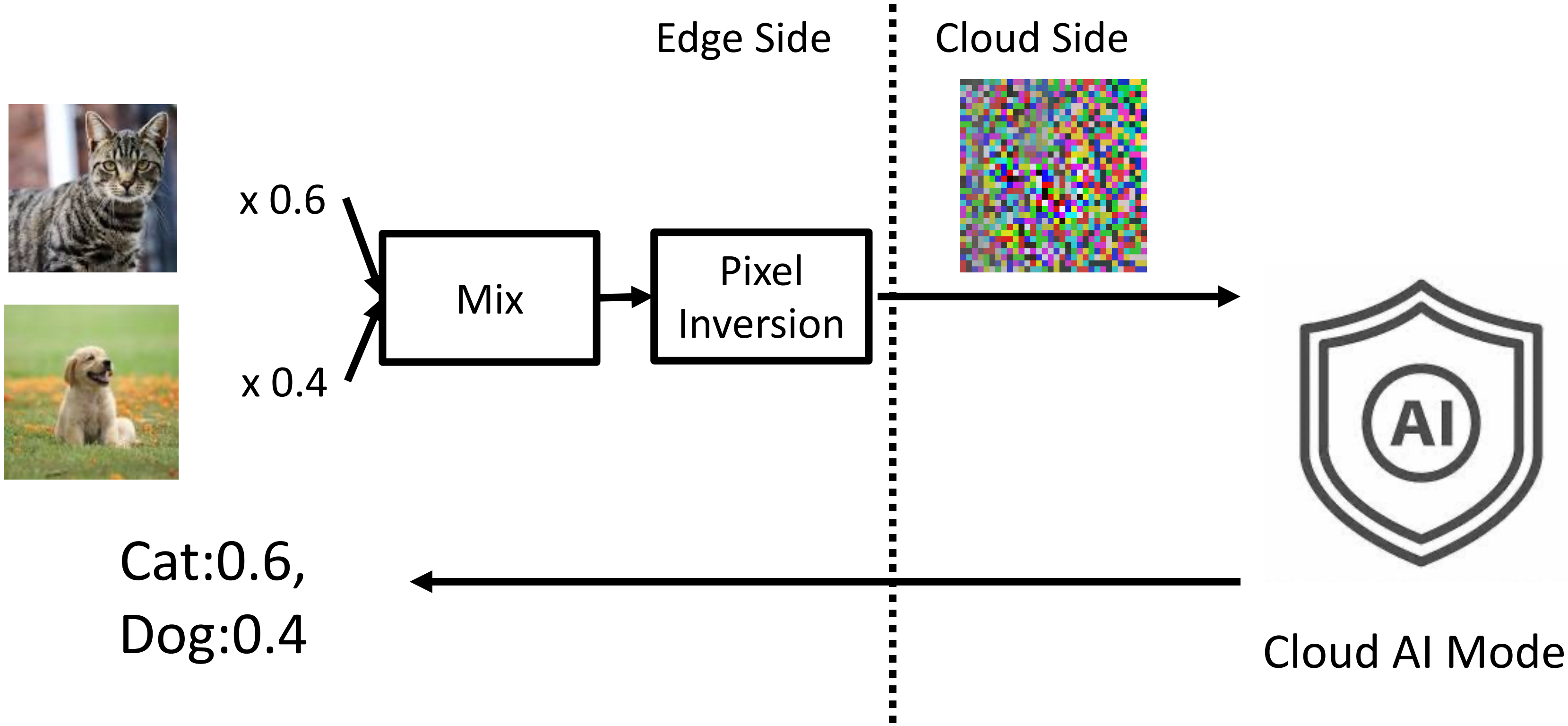
(1) Mix Images and Encrypt images

(2) Received Inference results.



InstaHide [Haung et al]

(1) Mix Images and Encrypt images
(2) Received Inference results.



Problem : Two images are necessary for prediction

Image Scrambling [Tanaka, Sirichoptedumrong et al]

Approach to make scrambled image using a key

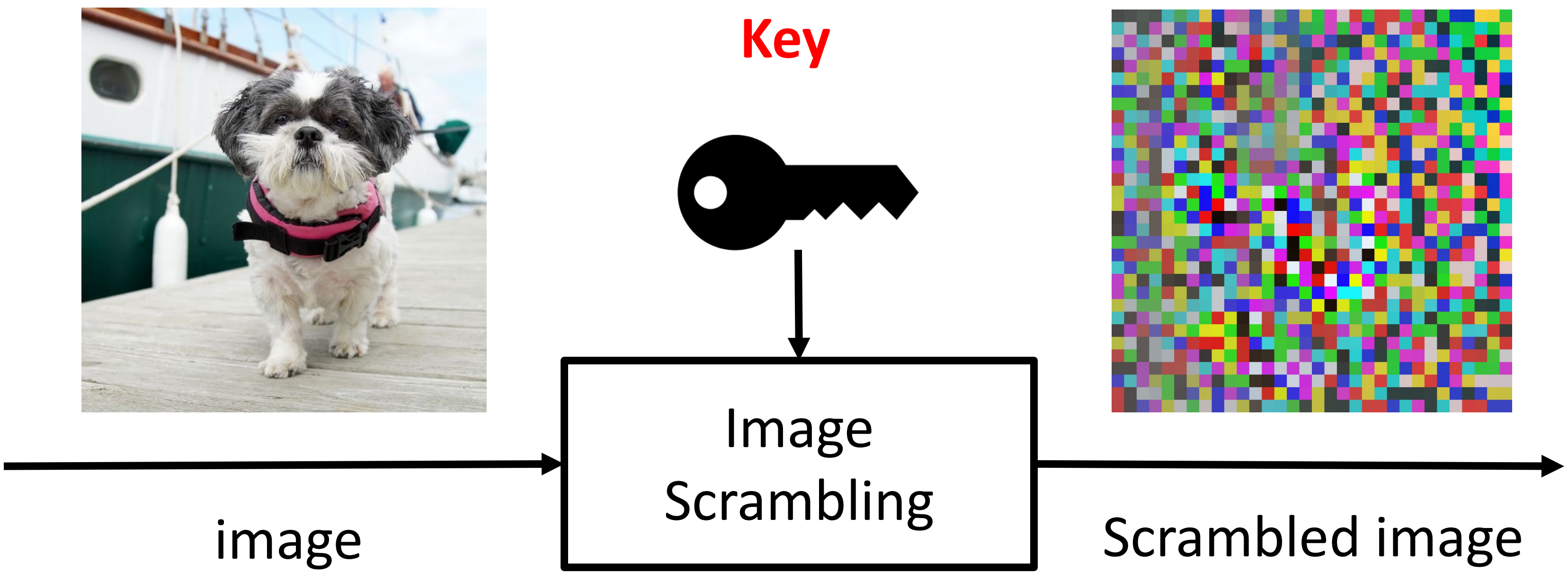


Image Scrambling [Tanaka, Sirichoptedumrong et al]

1. Send scrambled image

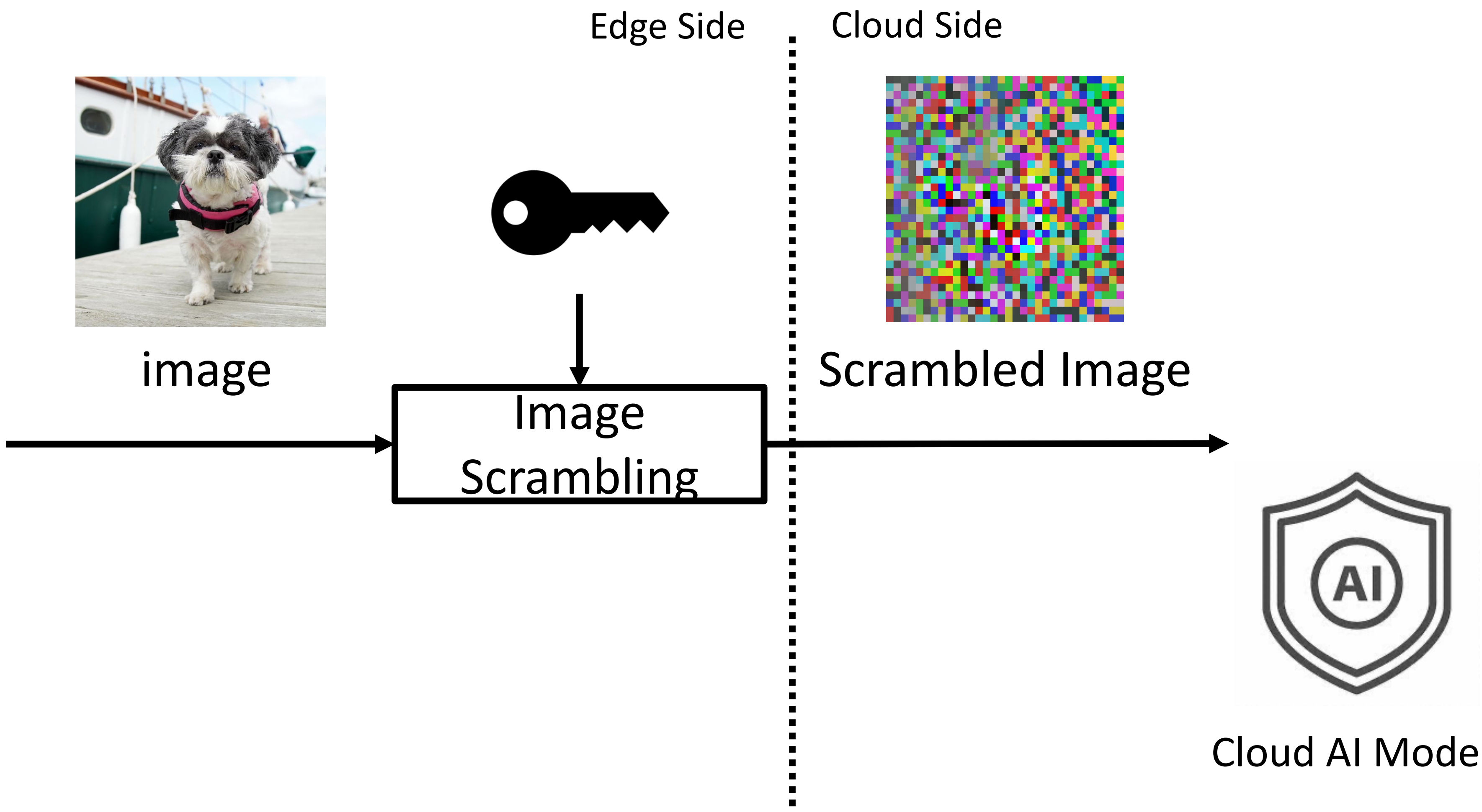


Image Scrambling [Tanaka, Sirichoptedumrong et al]

2. Get inference results

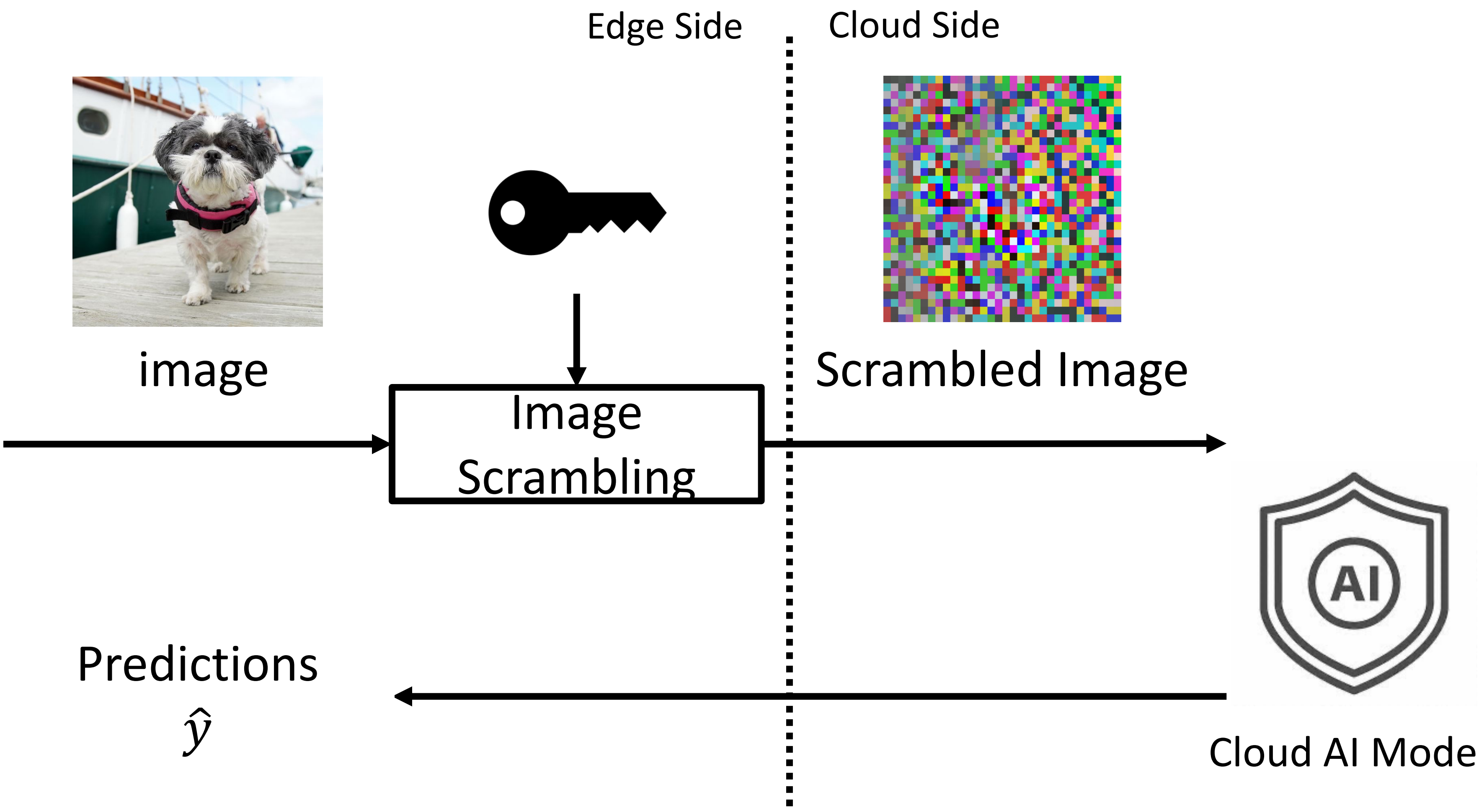


Image Scrambling [Tanaka, Sirichoptedumrong et al]

Situation : Chihuahua is dog or cat?

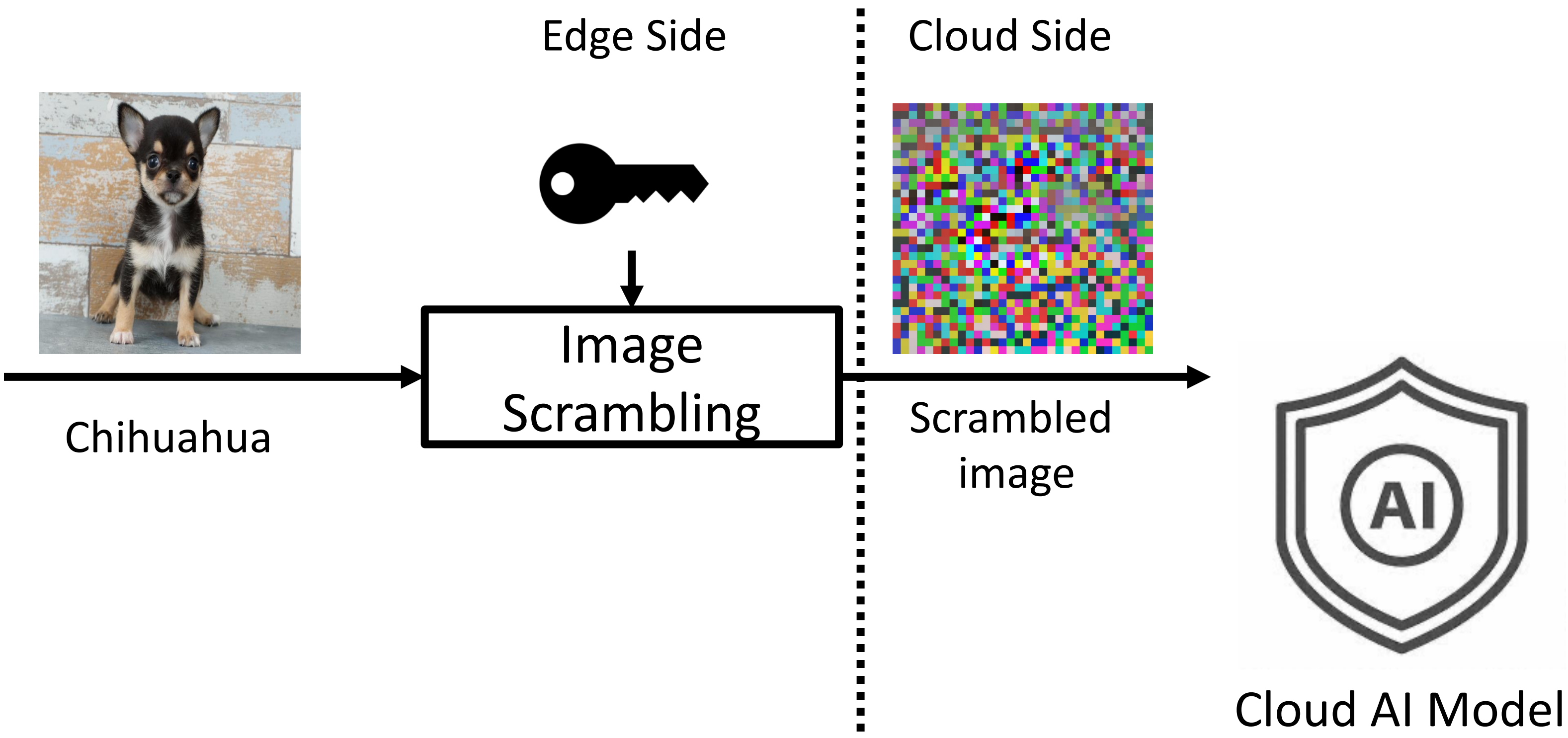


Image Scrambling [Tanaka, Sirichoptedumrong et al]

Limitation

① Lower Accuracy

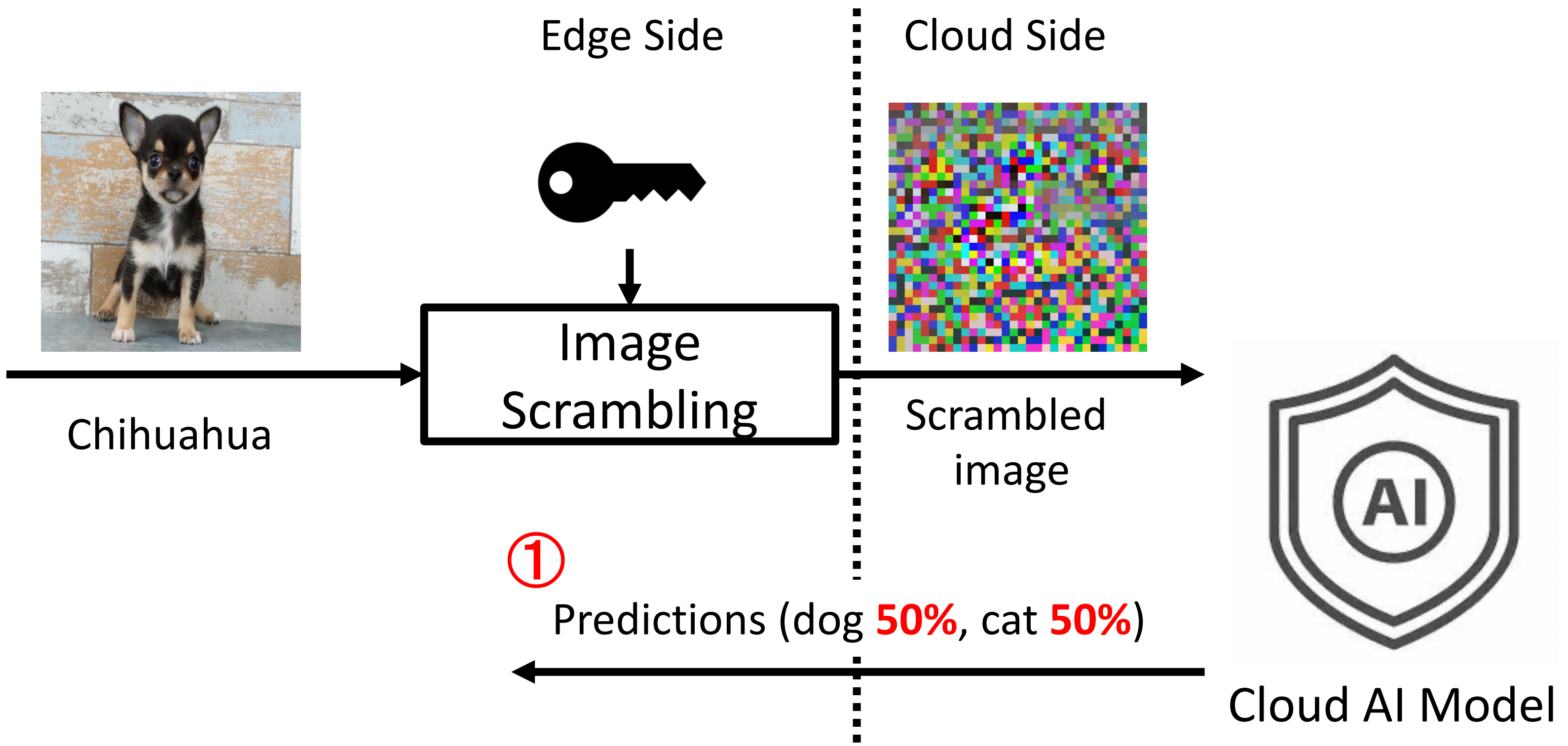
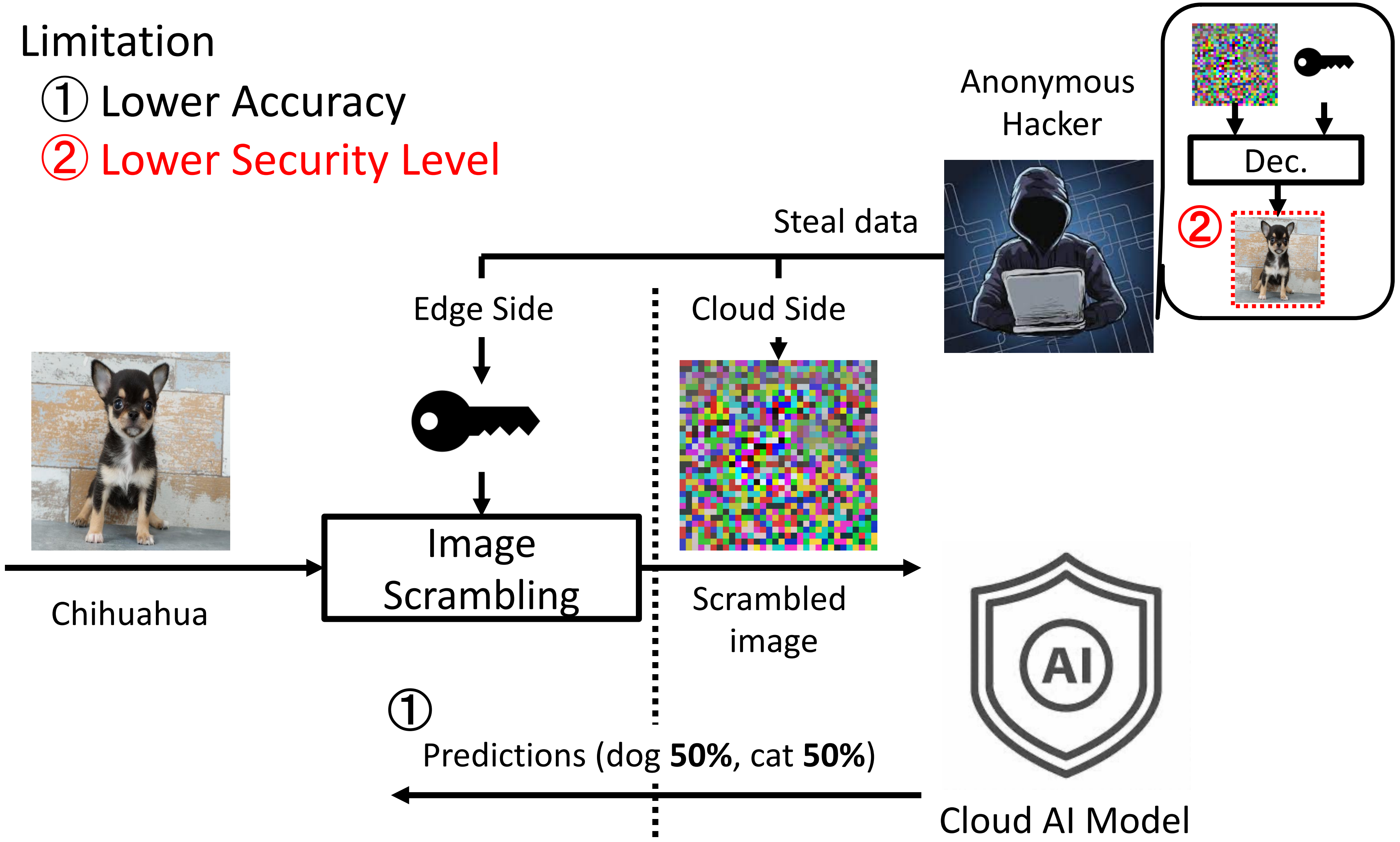


Image Scrambling [Tanaka, Sirichoptedumrong et al]

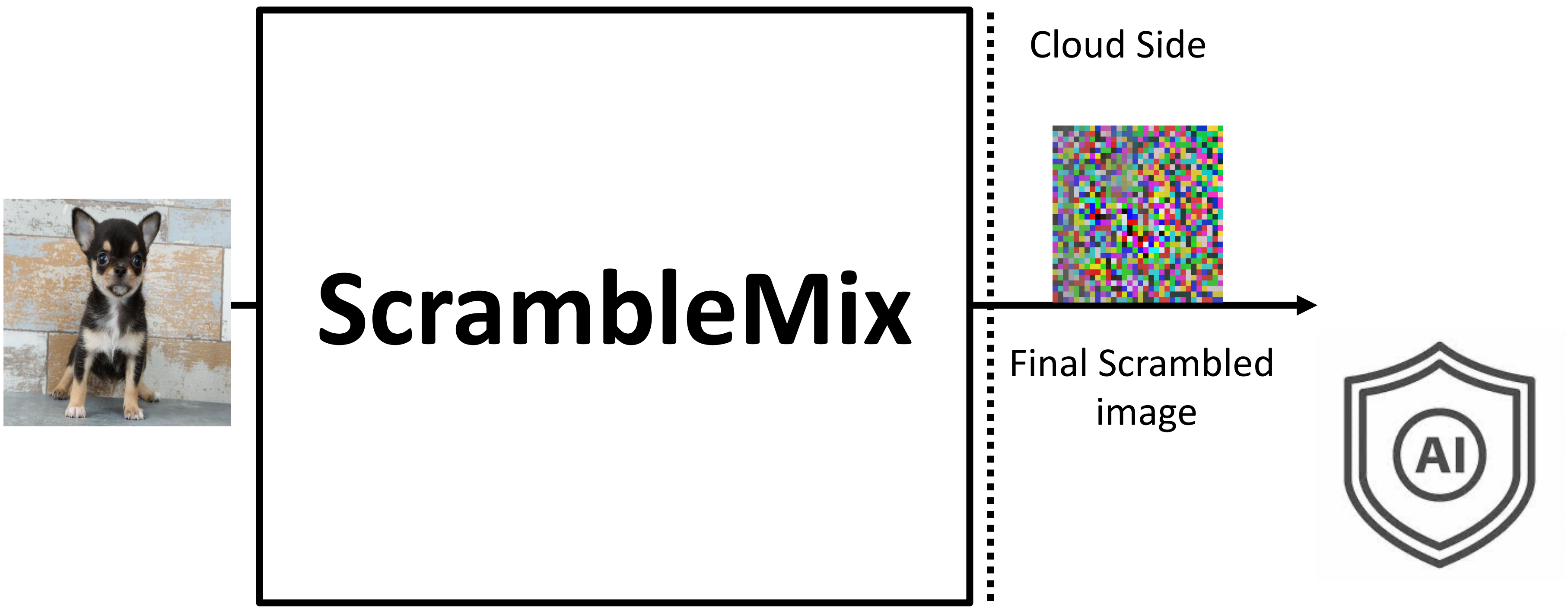
Limitation

① Lower Accuracy

② Lower Security Level



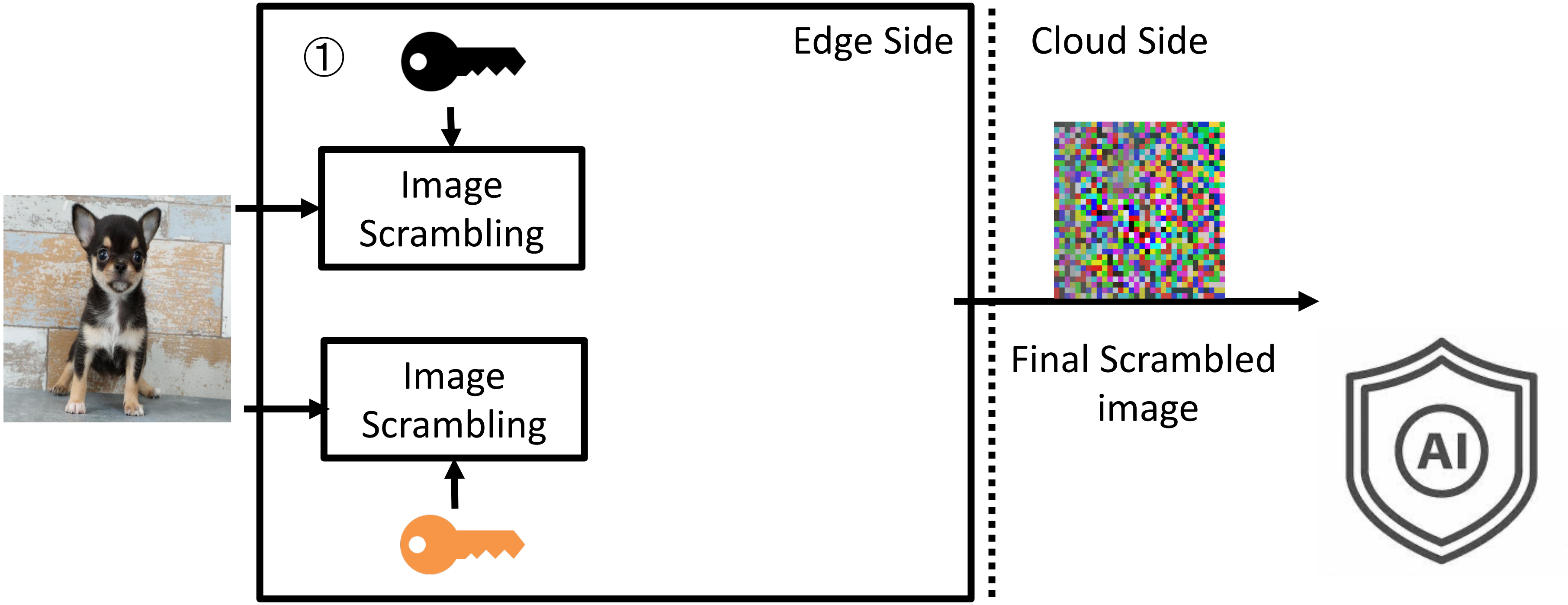
ScrambleMix (Proposed)



ScrambleMix (Proposed)

Differences from Image scrambling

① One key pairs (,) for scrambling

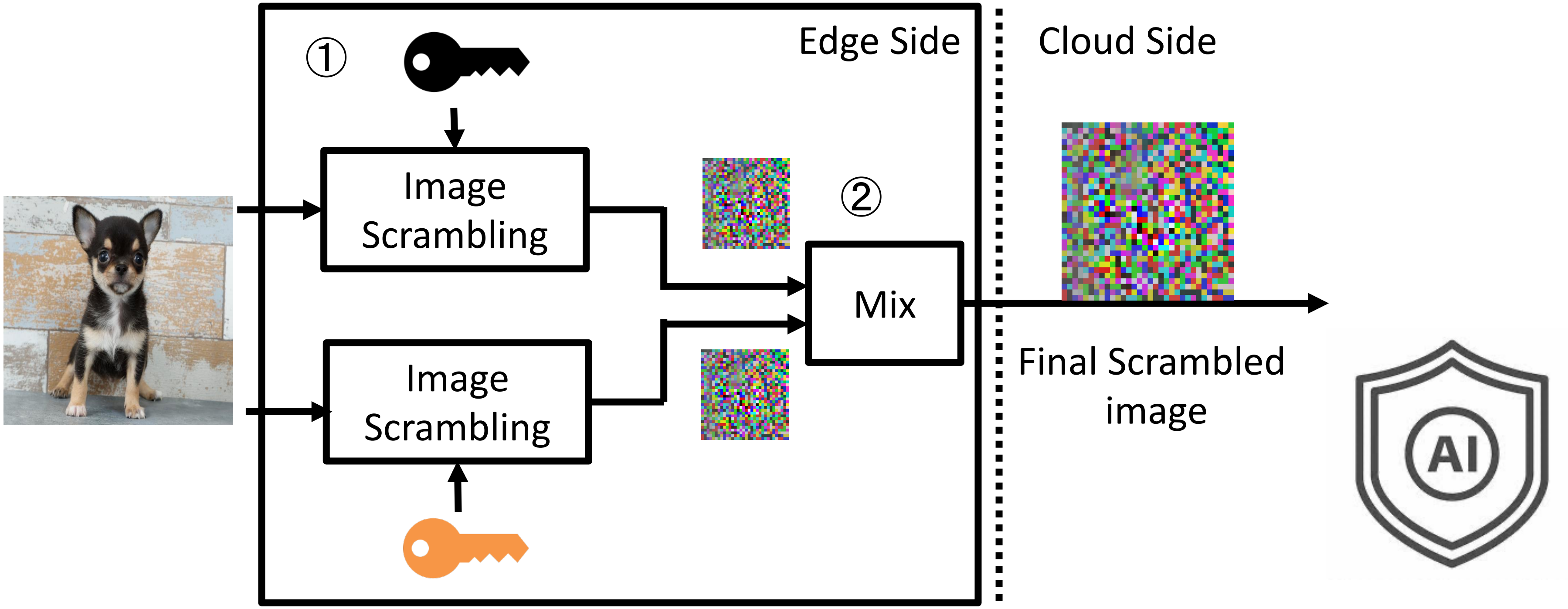


ScrambleMix (Proposed)

Differences from Image scrambling

① One key pairs (,) for scrambling

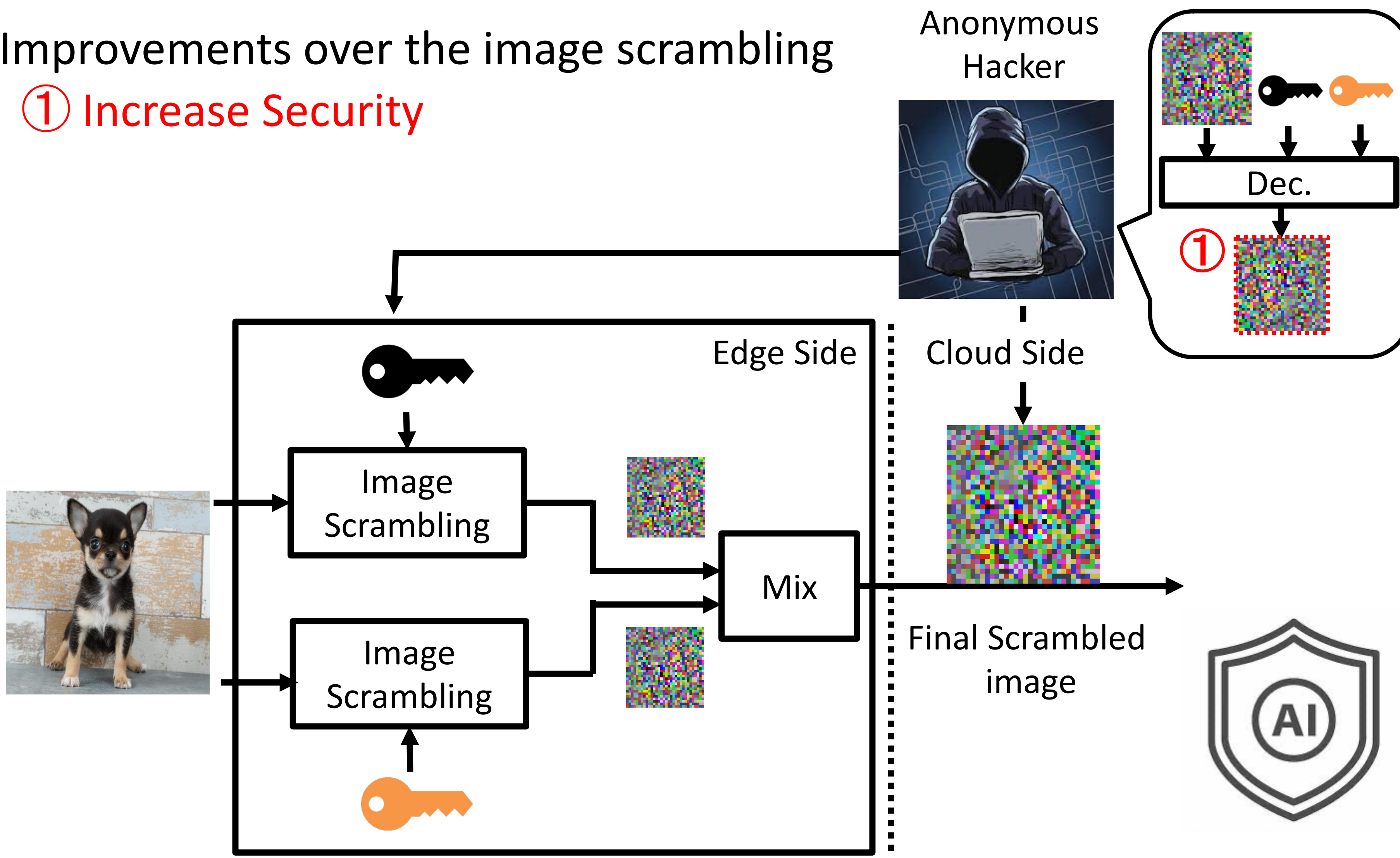
② Mix two scrambled images



ScrambleMix (Proposed)

Improvements over the image scrambling

① Increase Security

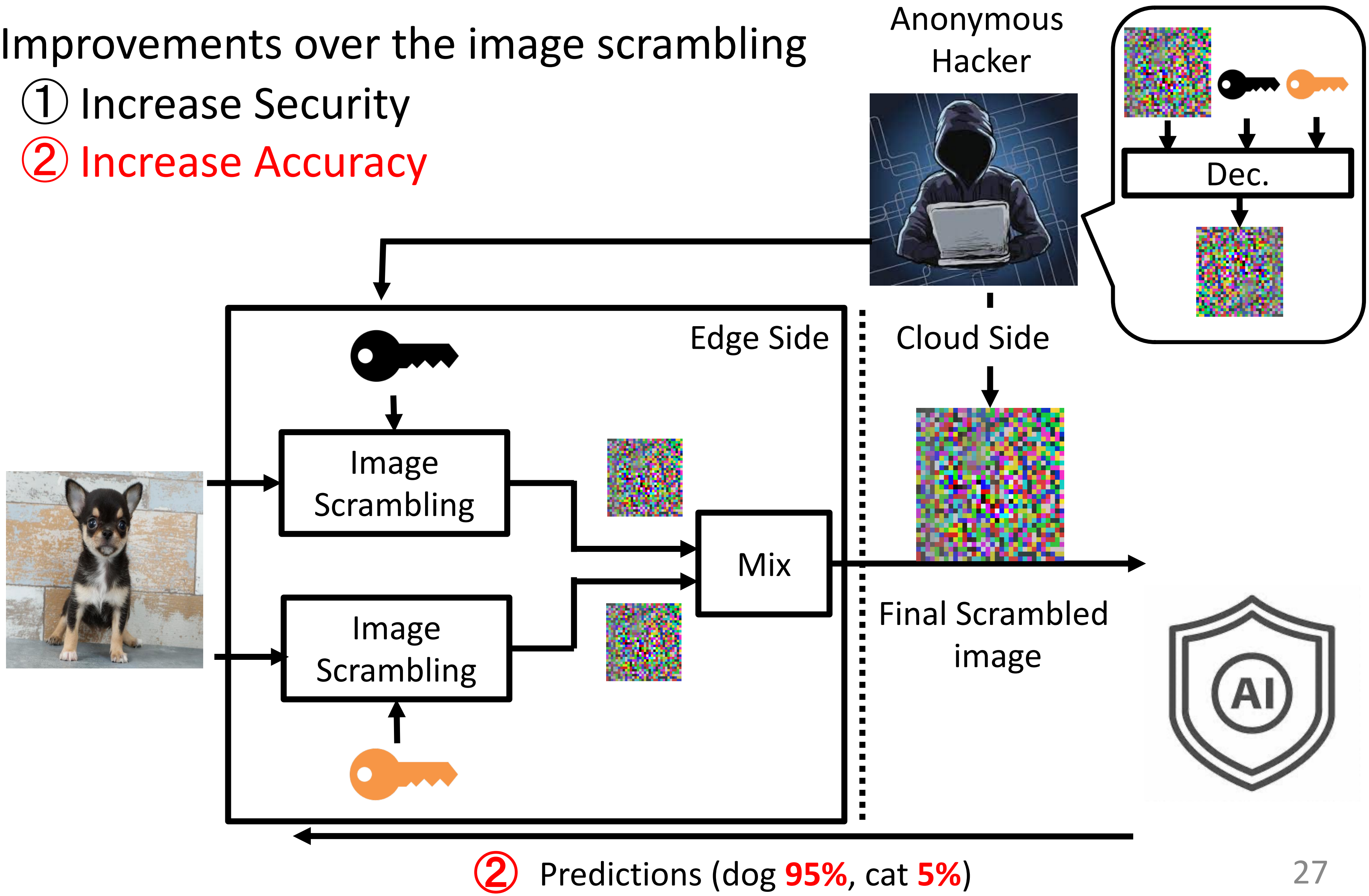


ScrambleMix (Proposed)

Improvements over the image scrambling

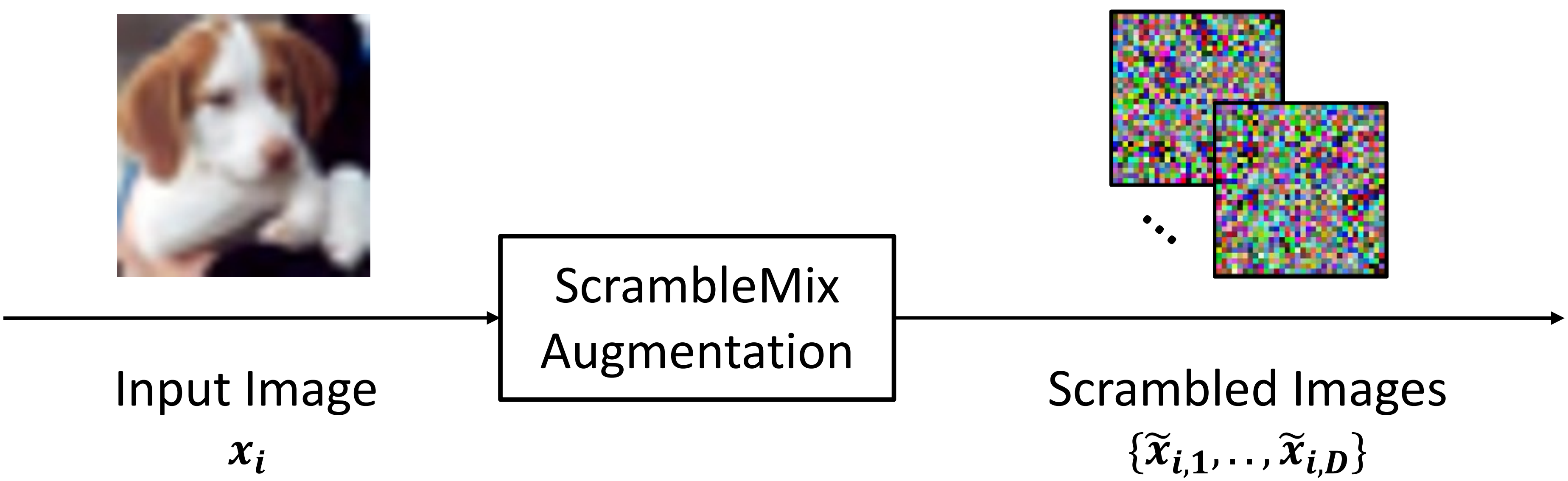
① Increase Security

② Increase Accuracy



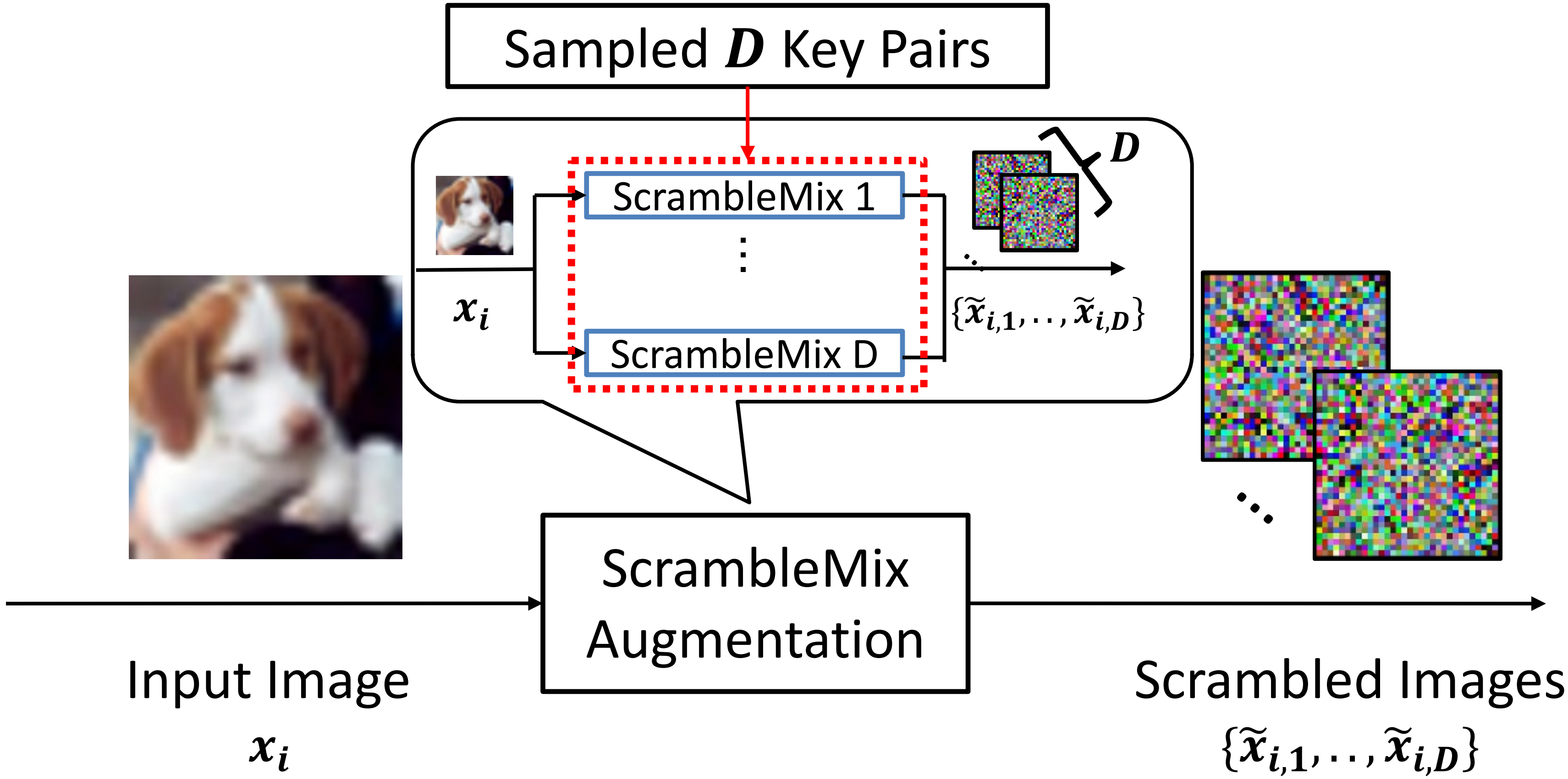
Training with ScrambleMix

1. Do ScrambleMix Augmentation



Training with ScrambleMix

1. Do ScrambleMix Augmentation
 - image is augmented $\boldsymbol{D}$ scrambled images.



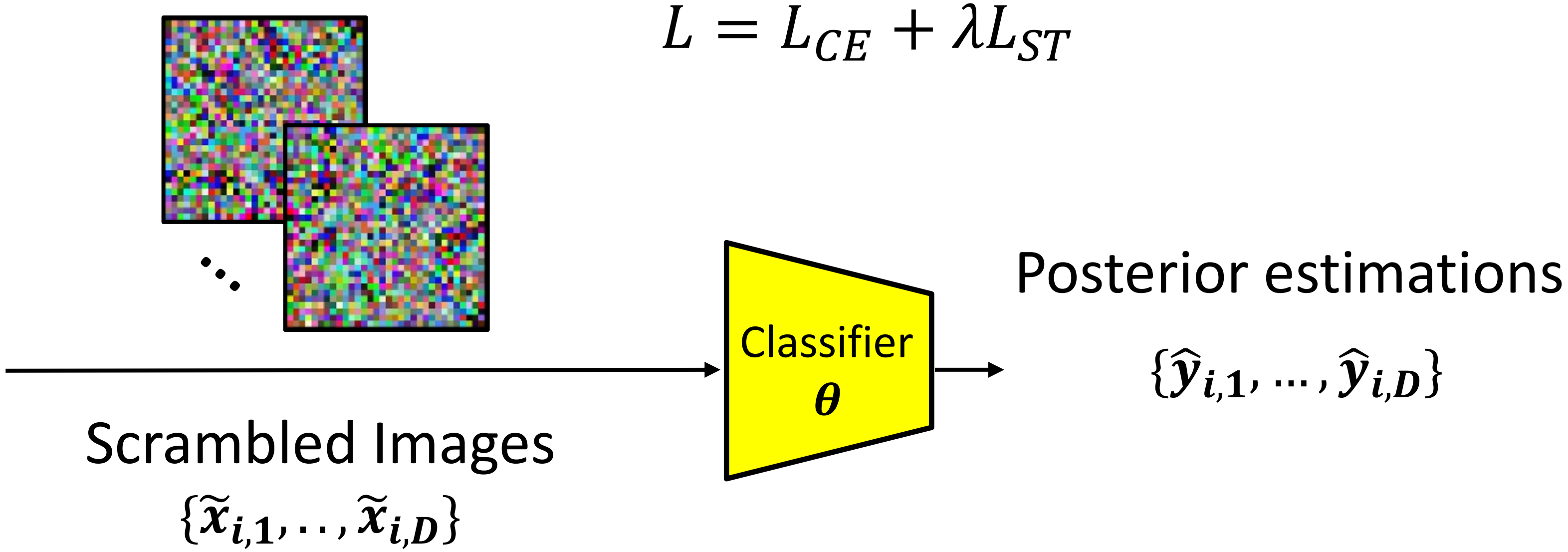
Optimization

2. Compute the loss for optimization

+ L_{CE} : Cross-entropy Loss

+ L_{ST} : Self-teaching Loss (proposed)

$$L = L_{CE} + \lambda L_{ST}$$

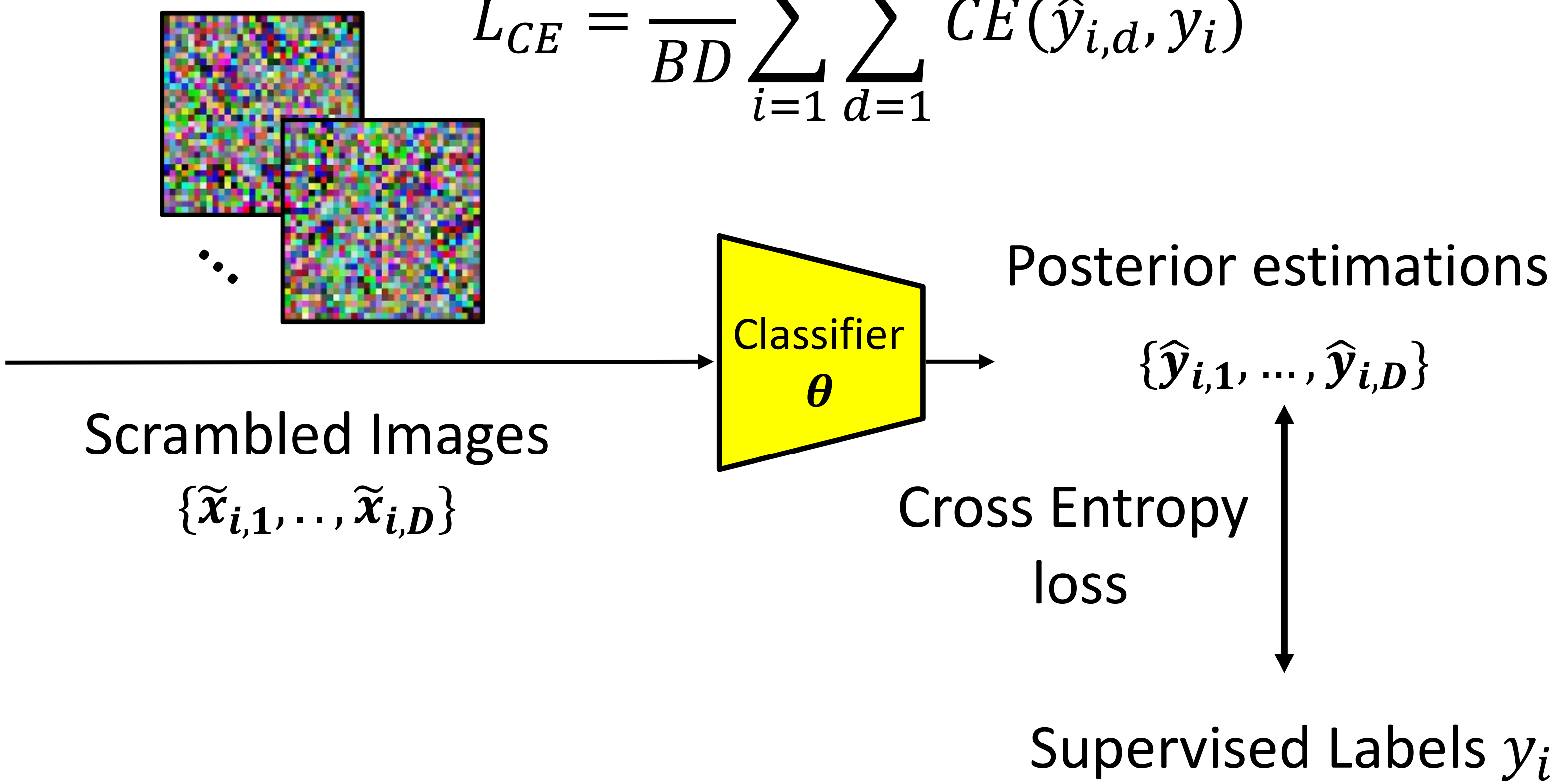


Optimization

2.1. Cross Entropy Loss (CE)

+ Minimize the posterior with supervised labels

$$L_{CE} = \frac{1}{BD}\sum_{i=1}^{B}\sum_{d=1}^{D} CE(\hat{y}_{i,d}, y_i)$$

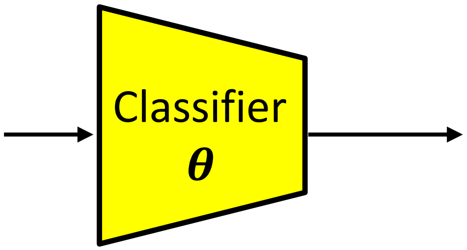


Optimization

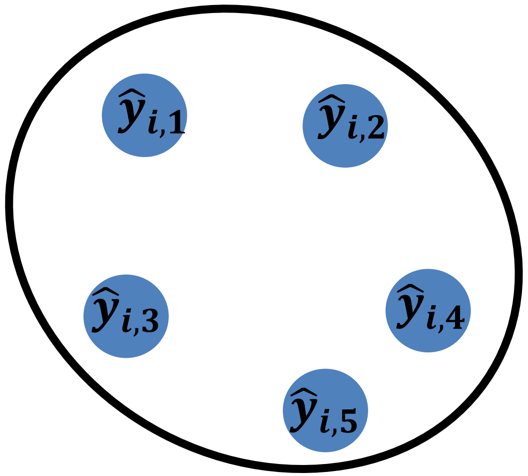
2.2. Self-Teaching Loss (ST)

+ posterior changes due to different keys

2D visualization



$\{\hat{\boldsymbol{y}}_{\boldsymbol{i},\mathbf{1}}, \dots, \hat{\boldsymbol{y}}_{\boldsymbol{i},\boldsymbol{D}}\}$

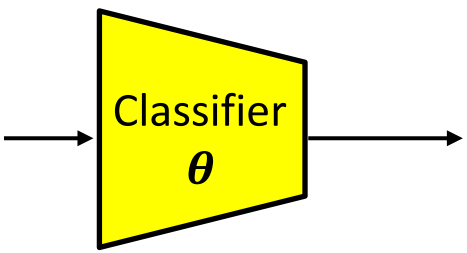


Posterior estimations

Optimization

2.2. Self-Teaching Loss (ST)

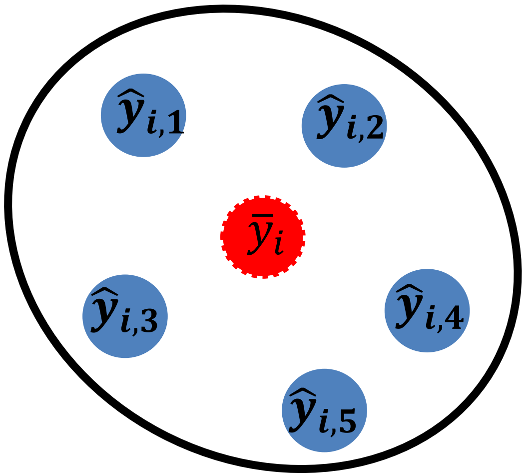
+ Same original image should have same posterior



$\{\hat{y}_{i,1}, \dots, \hat{y}_{i,D}\}$

Posterior estimations

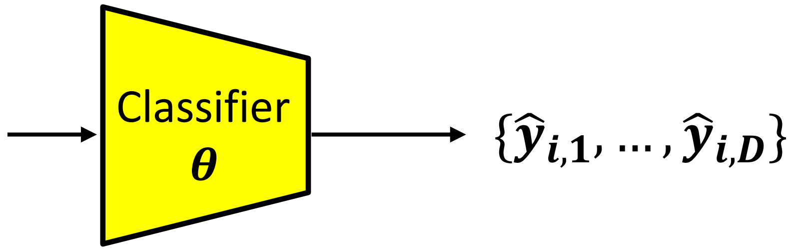
2D visualization



Optimization

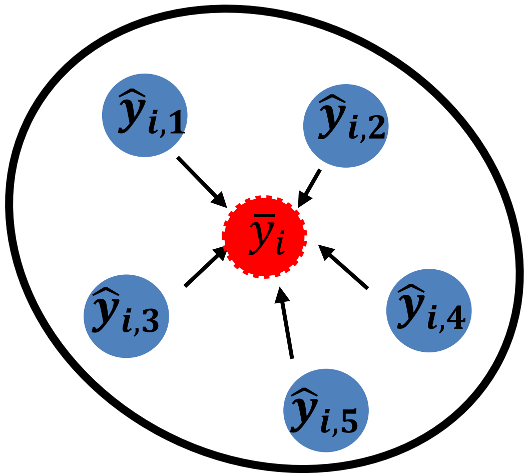
2.2. Self-Teaching Loss (ST)

+ Approach : Minimize each posterior and average posterior



Posterior estimations

2D visualization



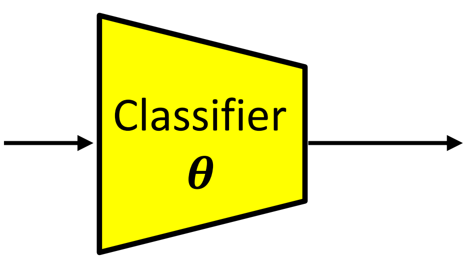
Optimization

2.2. Self-Teaching Loss (ST)

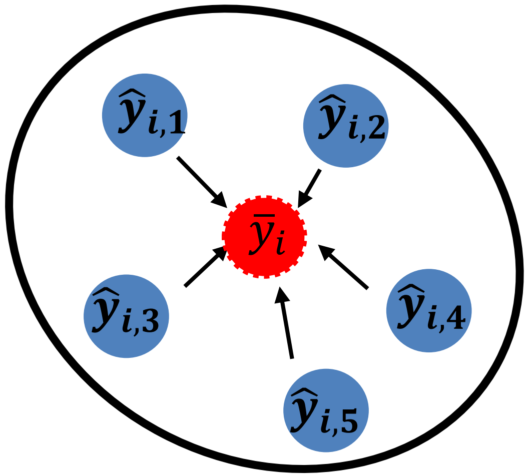
+ Average posterior: $\bar{y}_i$

$$\bar{y}_i = \mathrm{StopGrad}\left[\frac{1}{D}\sum_{d=1}^{D}\hat{\boldsymbol{y}}_{\boldsymbol{i,d}}\right]$$

2D visualization



$\{\hat{\boldsymbol{y}}_{\boldsymbol{i},1},\ldots,\hat{\boldsymbol{y}}_{\boldsymbol{i},\boldsymbol{D}}\}$



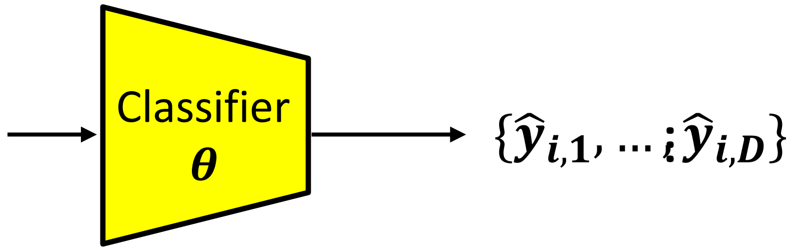
Posterior estimations

Optimization

2.2. Self-Teaching Loss (ST)

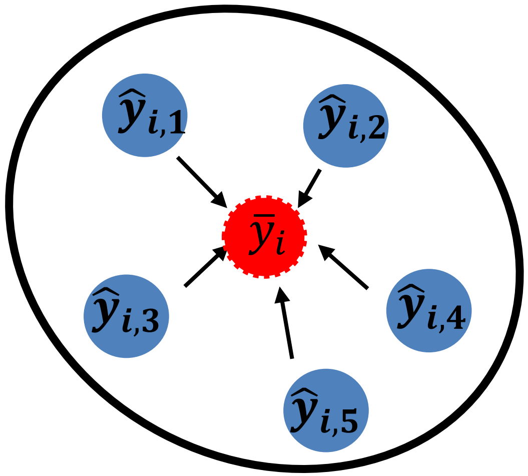
+ Minimize the posterior with supervised labels

$$L_{ST} = \frac{1}{BD}\sum_{i=1}^{B}\sum_{d=1}^{D} KL(\hat{y}_{i,d}||\bar{y}_i)$$



Posterior estimations

2D visualization

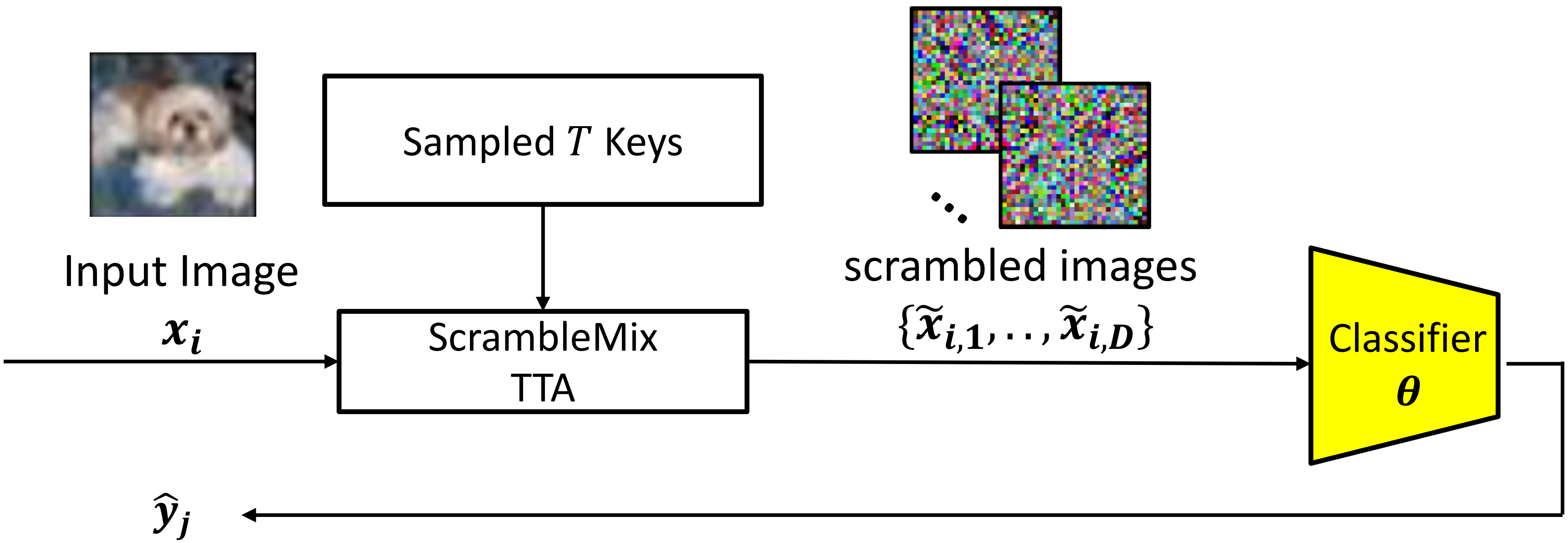


Inference

ScrambleMix TTA (Test Time Augmentation)

- T Keys : same as training phase's keys

$$\hat{y}_j = \frac{1}{T}\sum_{t=1}^{T}\hat{y}_{j,t}$$



Experiment

Baseline

+ InstaHide [Haung 2020]
+ DataMix [Liu 2020]
+ Image Scrambling
  - Learnable Encryption [Tanaka 2018]
  - Random Pixel-wise Encryption [Sirichoptedumrong 2019]

Proposed

+ ScrambleMix

Evaluation

1. Classification task: on Cifar10/100, SVHN
2. Security score : on InstaHide attack[Carlini 2020]

Results (T=1, w/o Test-Time Augmentation)

WideResNet40x10

Accuracy scores	CIFAR10	CIFAR100	SVHN
DataMix	66.89	38.31	19.60
InstaHide	53.58	39.06	52.47
LE	91.34	70.62	96.50
Random PE	92.23	70.82	96.83
ScrambleMix (Proposed)	**93.08**	**71.71**	**96.96**

Shakedrop

Accuracy scores	CIFAR10	CIFAR100	SVHN
DataMix	80.10	50.97	93.42
InstaHide	52.93	39.95	52.87
LE	94.02	77.59	97.26
Random PE	93.51	77.10	97.26
ScrambleMix (Proposed)	**95.02**	**79.39**	**97.47**

Results (T>=1, with Test-Time Augmentation)

Our approach : better on several scores

+ Even if T is small, our approach can get a comparable result

WideResNet40x10

Accuracy scores	**CIFAR10**	**CIFAR100**	**SVHN**
InstaHide, T=10	**94.92**	**78.32**	94.97
ScrambleMix, **T=4**	93.12	71.87	**97.01**

Shakedrop

Accuracy scores	**CIFAR10**	**CIFAR100**	**SVHN**
InstaHide, T=10	92.91	74.06	93.38
ScrambleMix, **T=4**	**95.31**	**79.41**	**97.54**

Results (Security Evaluation)

Attacked Results by InstaHide Attack [Carlini 2020]

+ Evaluate by inception score: high inception score means unsecure state

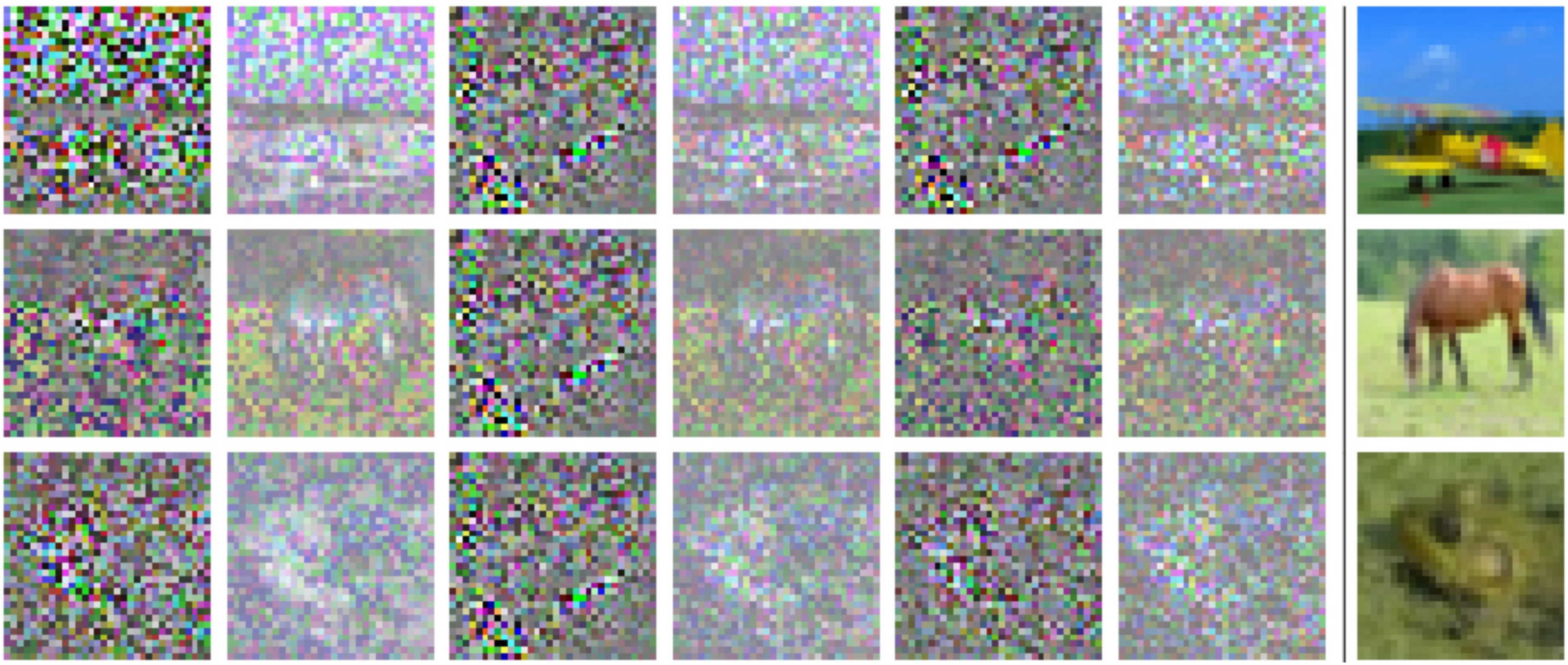
+ Our approach keeps low score (→ keep security)

	InstaHide	**ScrambleMix**
Non-attacked Scrambled Image	1.394 	1.012 
Attacked Scramble Image	**+1.383** 2.777 	**+0.165** **1.177** 

Results (Security Evaluation)

Attacked Results by InstaHide Attack

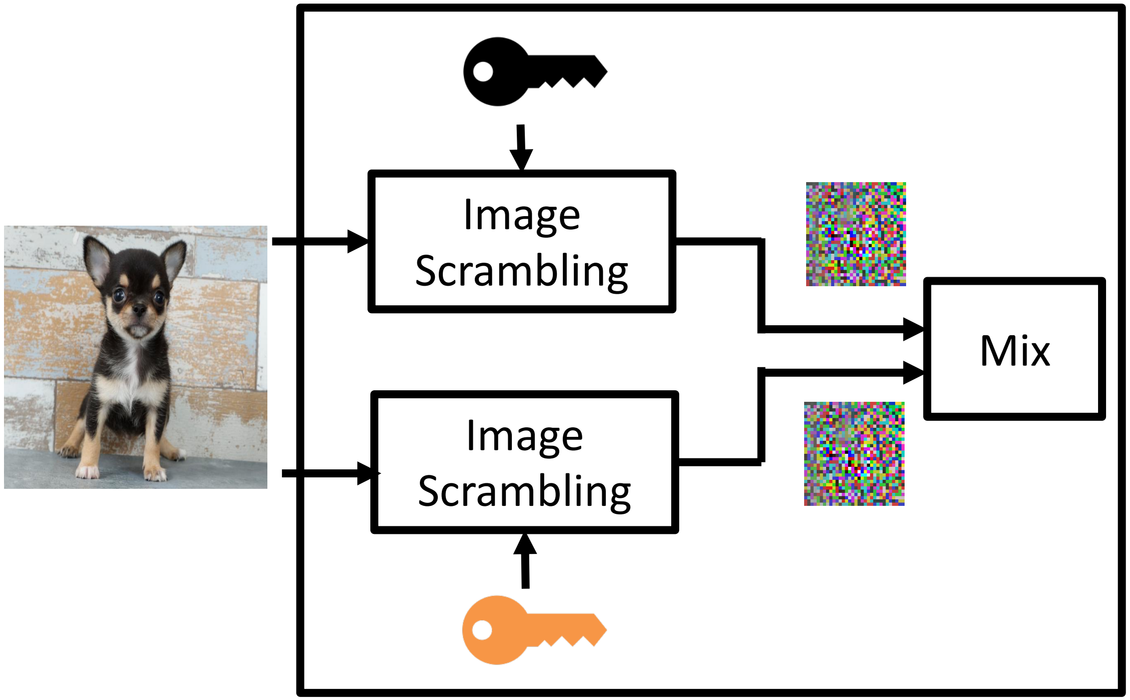
m=0.1		m=0.3		m=0.5		
Scramble Mix	Attacked	Scramble Mix	Attacked	Scramble Mix	Attacked	Original



Summary

ScrambleMix : new scrambling method for **edge-cloud machine learning**

- improve **classification accuracy** over almost settings
- improve **security** over the strong attack method



Overview of ScrambleMix

GitHub / slide