CamPro: Camera-based Anti-Facial Recognition

Wenjun Zhu, Yuan Sun, Jiani Liu, Yushi Cheng, Xiaoyu Ji, Wenyuan Xu

USSLAB, Zhejiang University

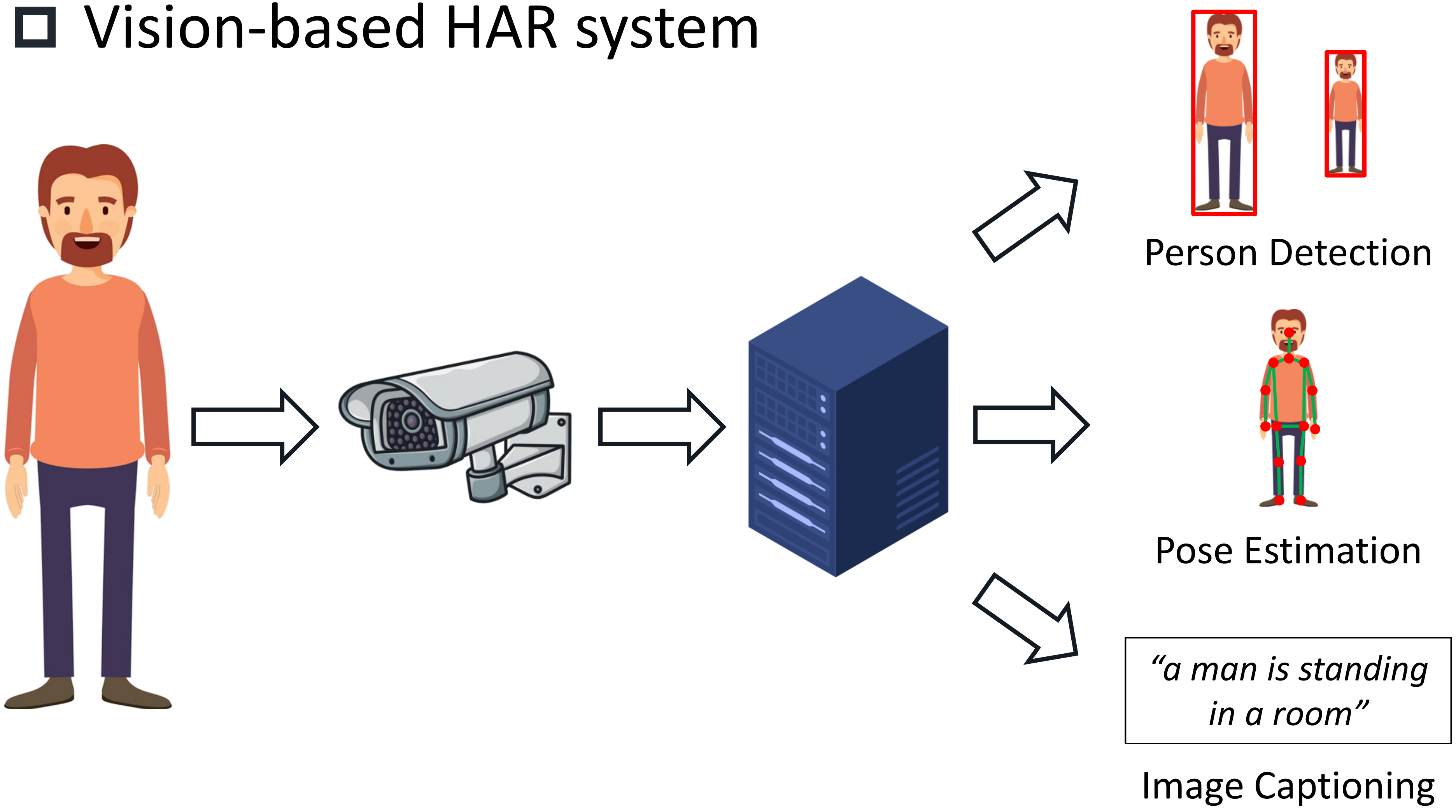
NDSS

浙江大學
ZHEJIANG UNIVERSITY

智能系统安全实验室
UBIQUITOUS SYSTEM SECURITY LAB.

Human Activity Recognition (HAR)

- Vision-based HAR system



Person Detection

Pose Estimation

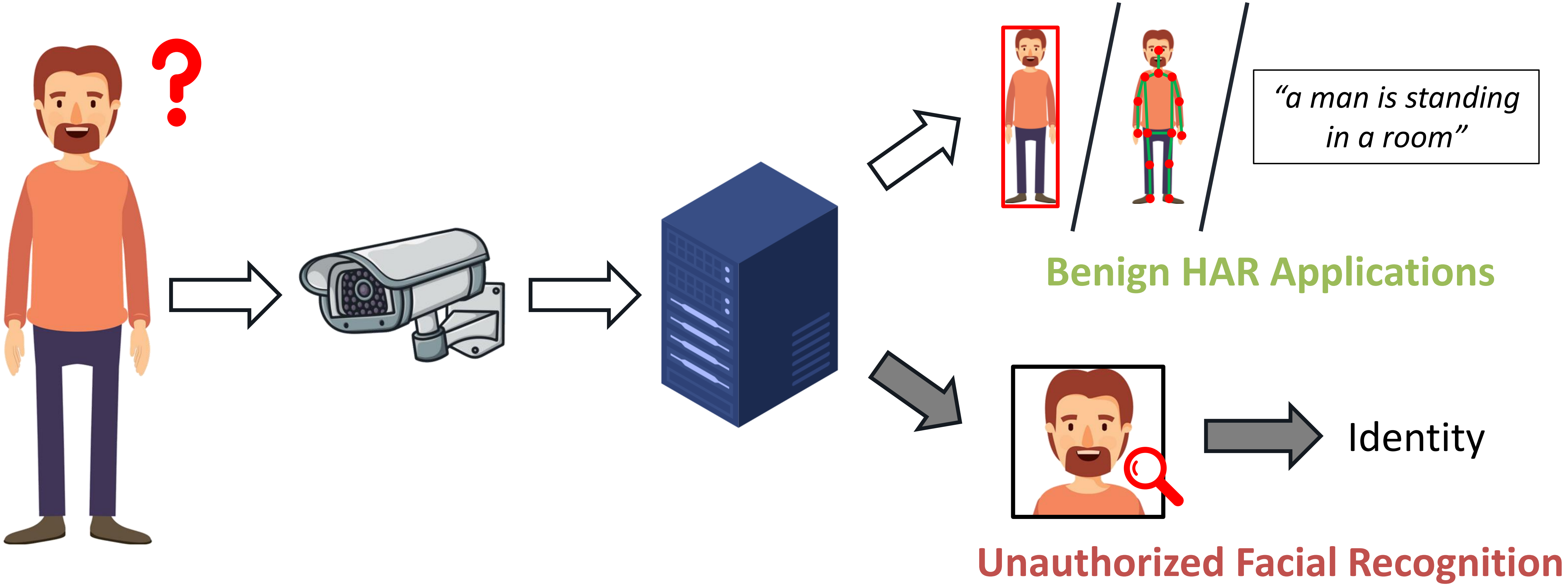
“a man is standing in a room”

Image Captioning

...

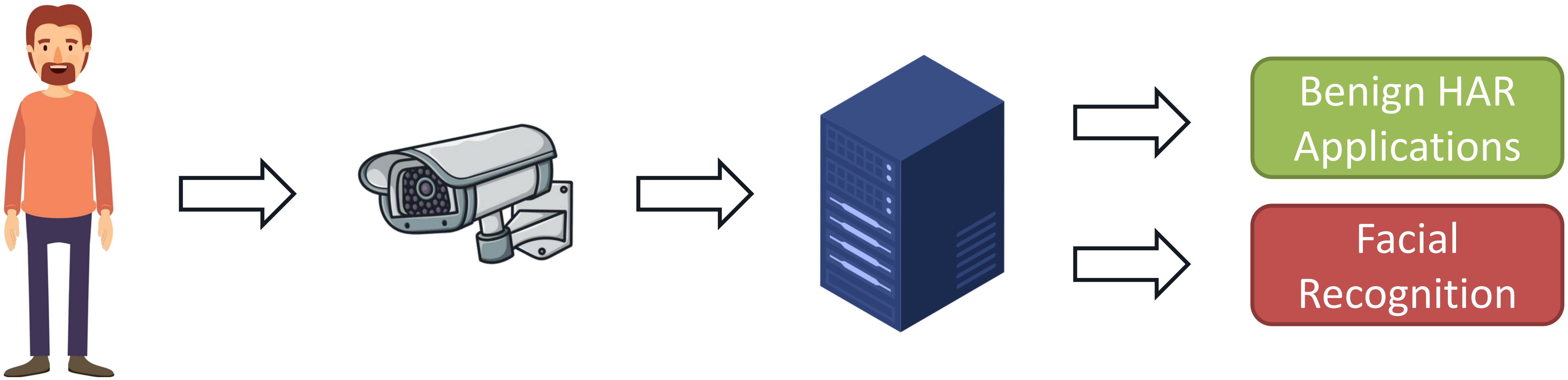
Human Activity Recognition (HAR)

- Vision-based HAR is often linked to privacy concerns.

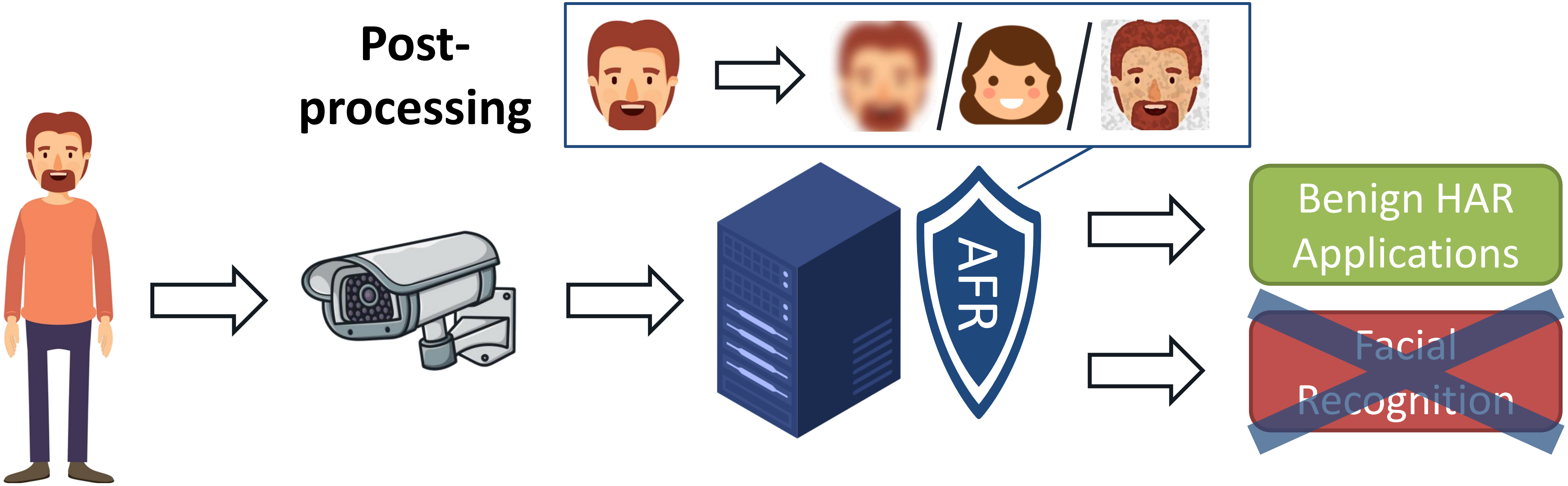


?
"a man is standing in a room"
Benign HAR Applications
Identity
Unauthorized Facial Recognition

Anti-Facial Recognition (AFR)



Anti-Facial Recognition (AFR)

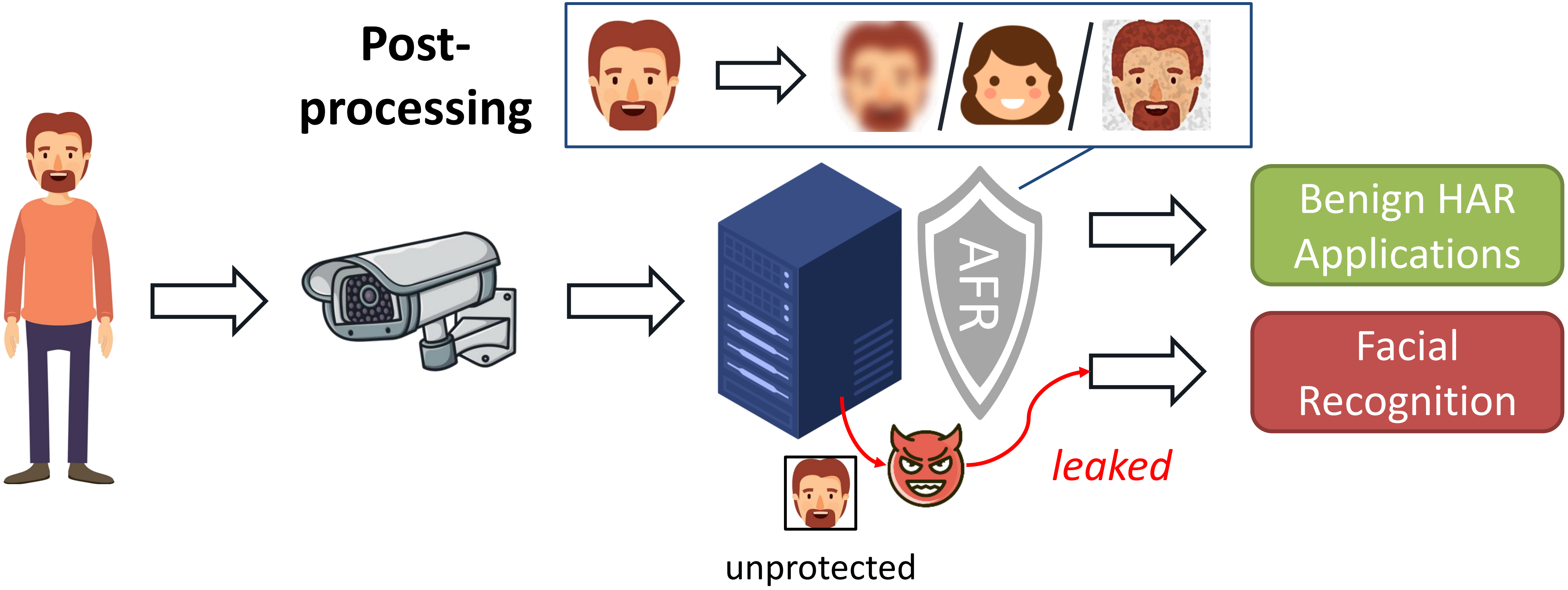


Post-processing

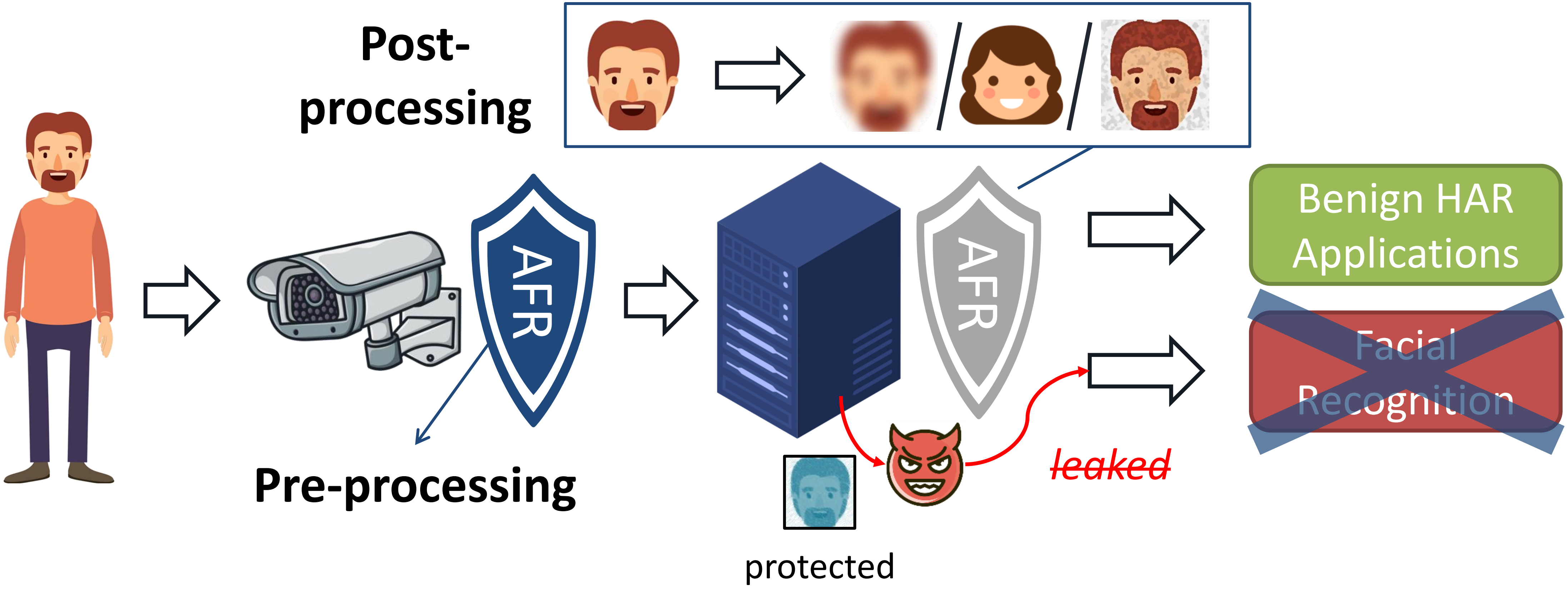
Benign HAR Applications

~~Facial Recognition~~

Anti-Facial Recognition (AFR)



Anti-Facial Recognition (AFR)



Post-processing

Benign HAR Applications

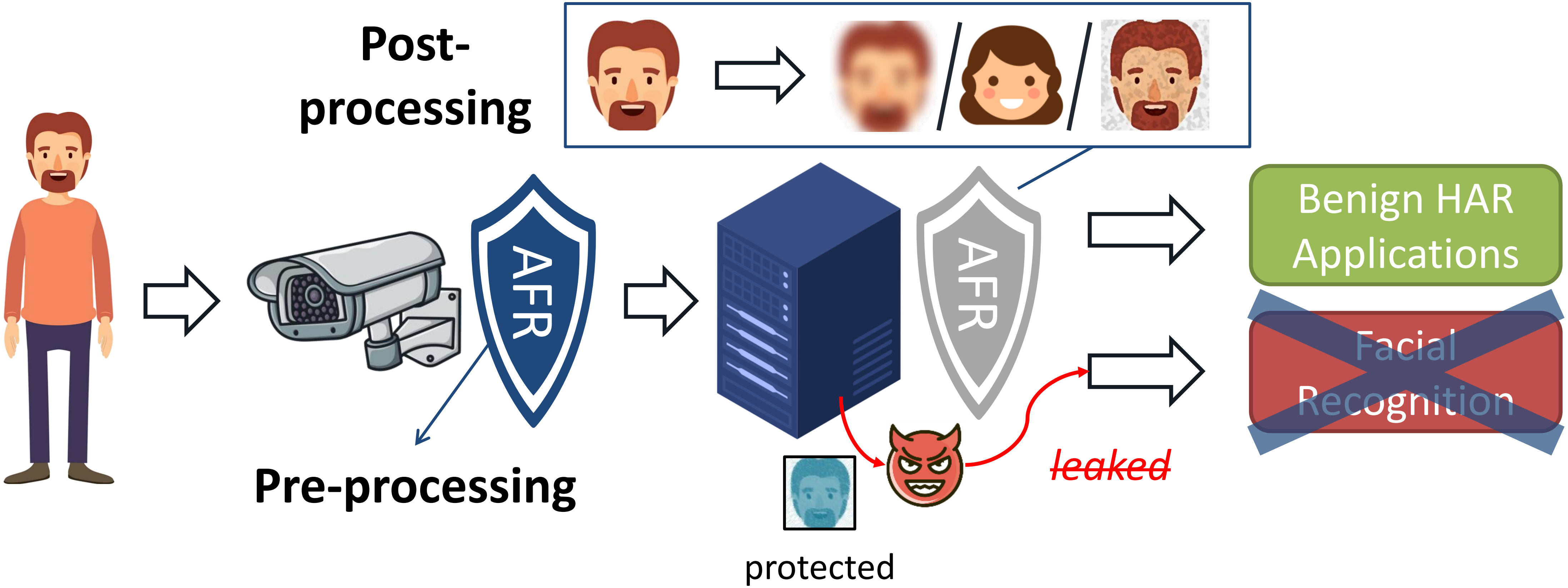
Facial Recognition

Pre-processing

~~leaked~~

protected

Anti-Facial Recognition (AFR)



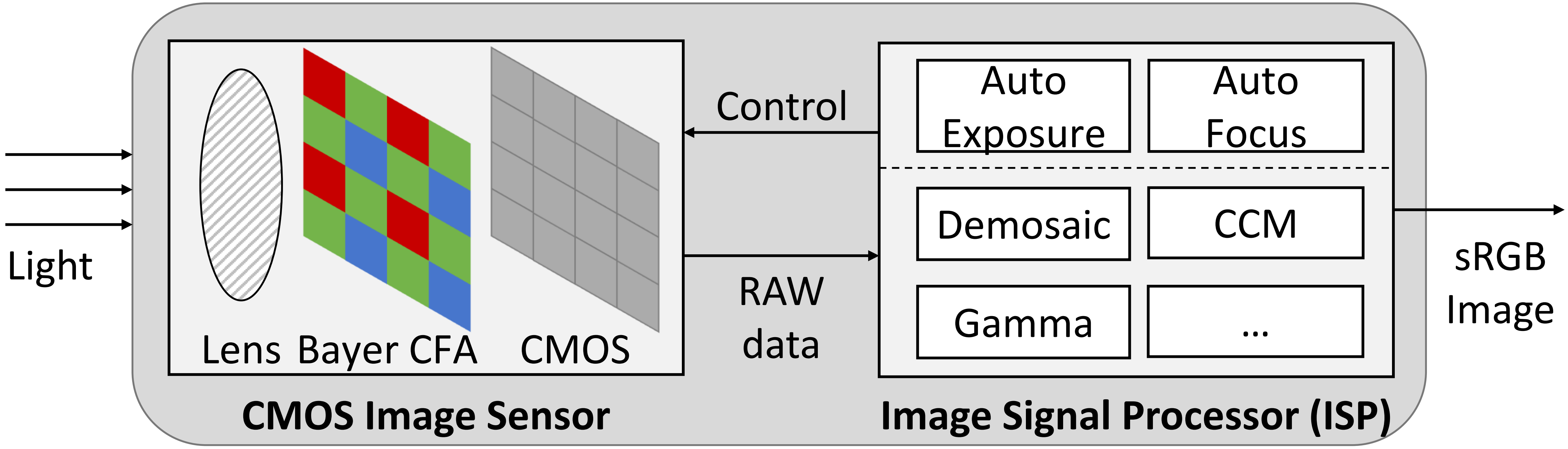
A new paradigm: ***privacy-preserving by birth***

Motivation

How to achieve AFR inside a basic camera module?

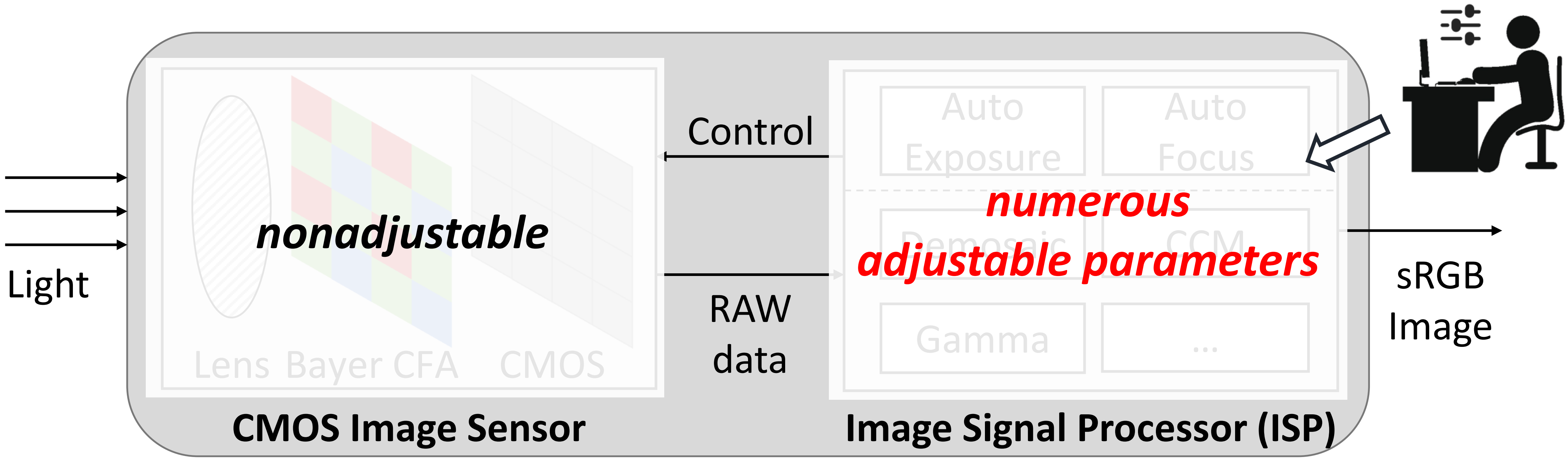
Motivation

How to achieve AFR inside a basic camera module?



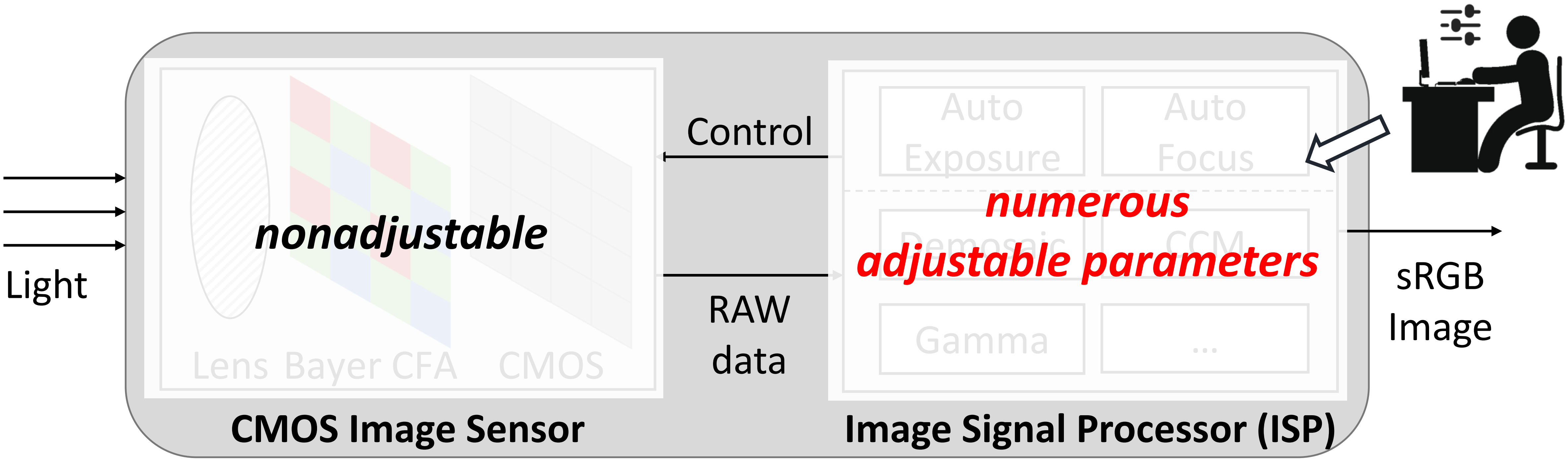
Motivation

How to achieve AFR inside a basic camera module?



Motivation

How to achieve AFR inside a basic camera module?



Basic Idea: achieve AFR by *adjusting ISP parameters*

Image Signal Processing

- Selected two ISP functions
  - Color correction
  - Gamma correction

Image Signal Processing

- Selected two ISP functions
  - Color correction → *a 3x3 matrix*
  - Gamma correction

} adjustable parameters

$$\begin{bmatrix} R_{out} \\ G_{out} \\ B_{out} \end{bmatrix} = \text{clip}_{[0,1]}\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} R_{in} \\ G_{in} \\ B_{in} \end{bmatrix} \right)$$

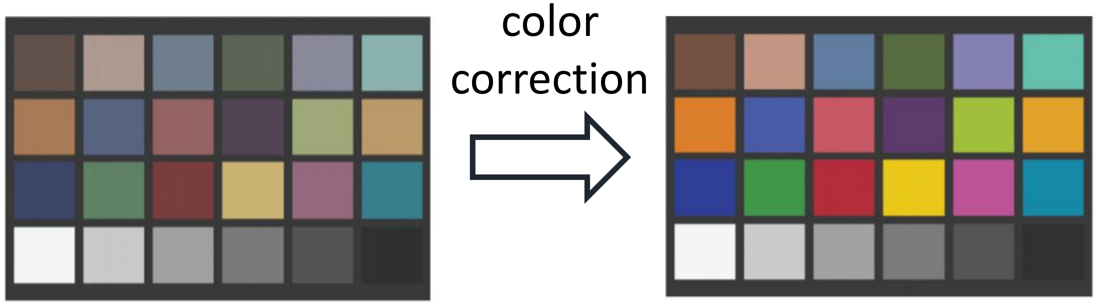


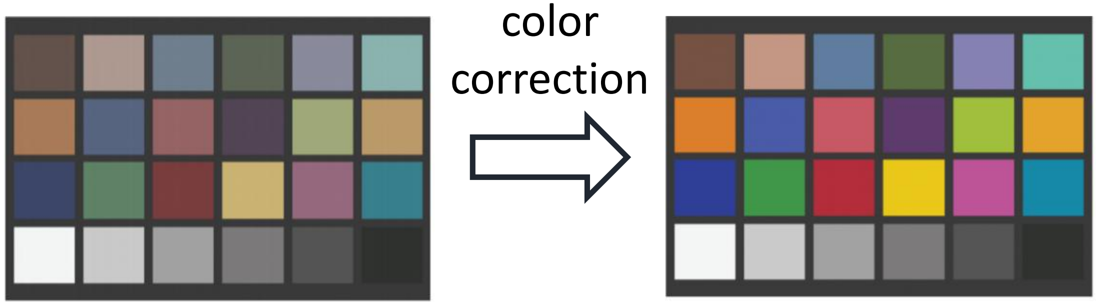
Image Signal Processing

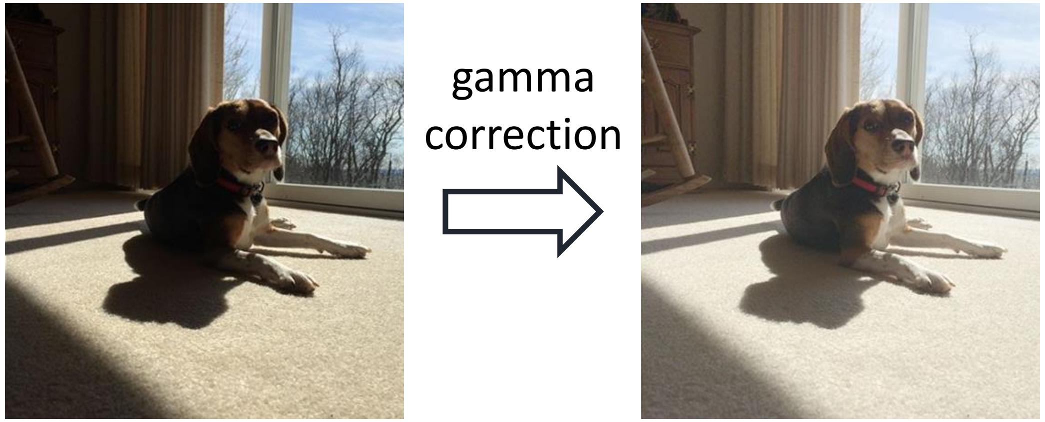
- Selected two ISP functions
  - Color correction → *a 3x3 matrix*
  - Gamma correction → *y-values*

} adjustable parameters

$$\begin{bmatrix} R_{out} \\ G_{out} \\ B_{out} \end{bmatrix} = \text{clip}_{[0,1]}\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} R_{in} \\ G_{in} \\ B_{in} \end{bmatrix} \right)$$

$$y = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i), \quad i = 1, 2, \cdots, k-1$$



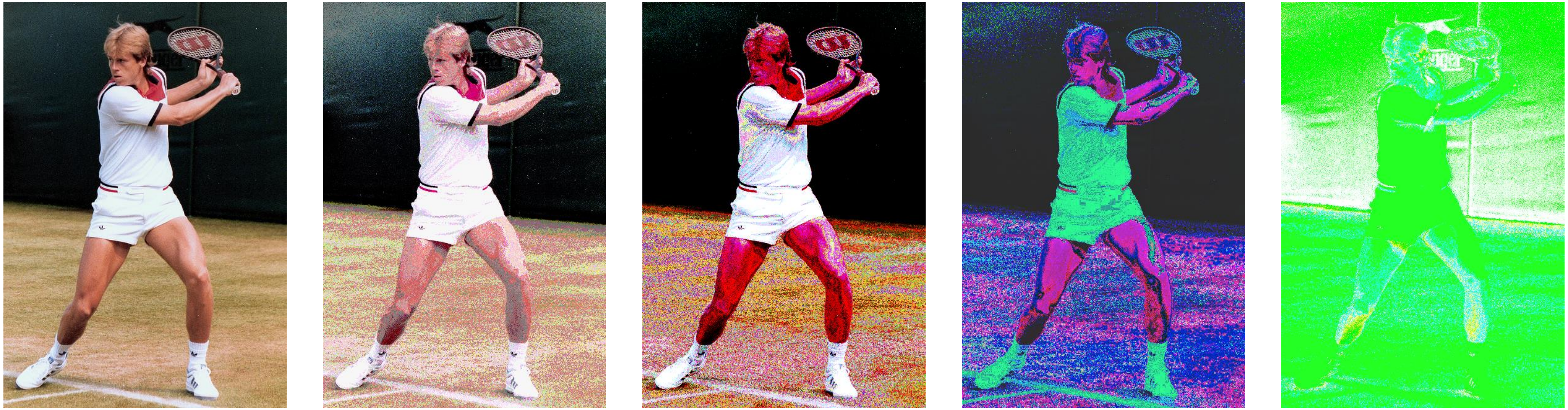


ISP Parameter Adjustment

Magnitude of parameter adjustment

Low → *High*

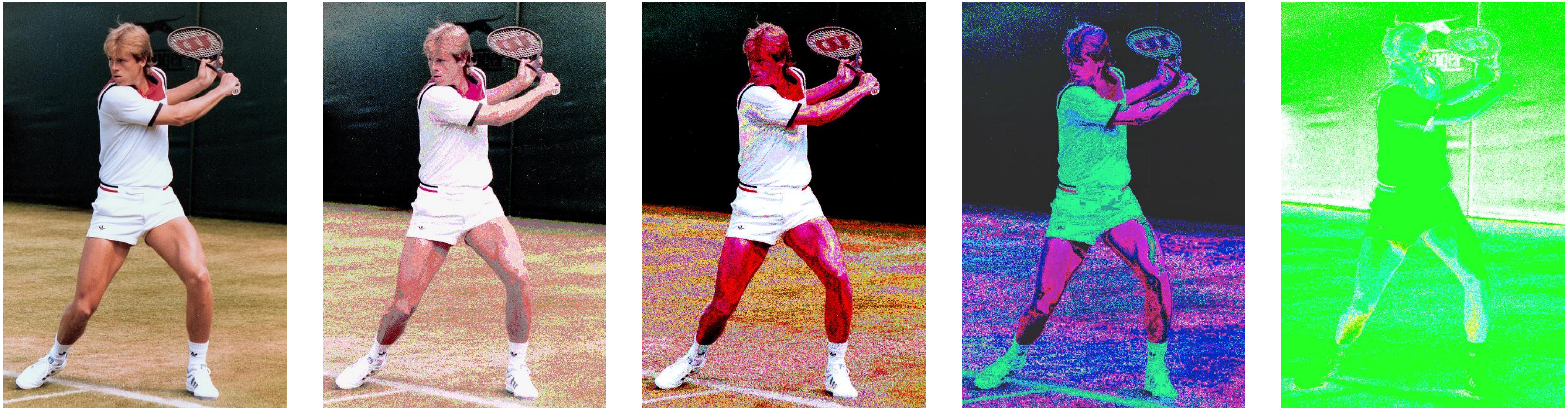
Original image

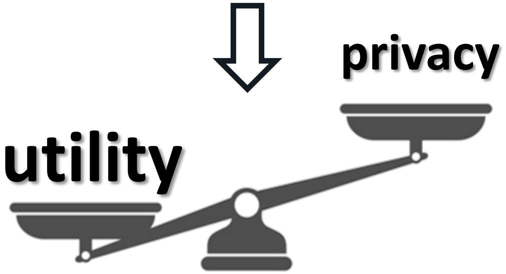


ISP Parameter Adjustment

Original image

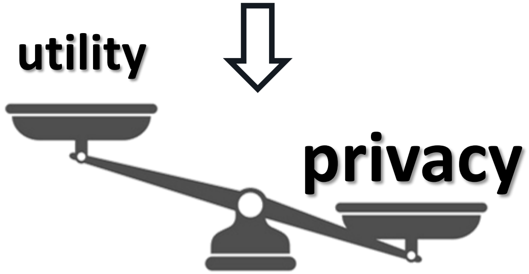
Low **Magnitude of parameter adjustment** *High*





privacy

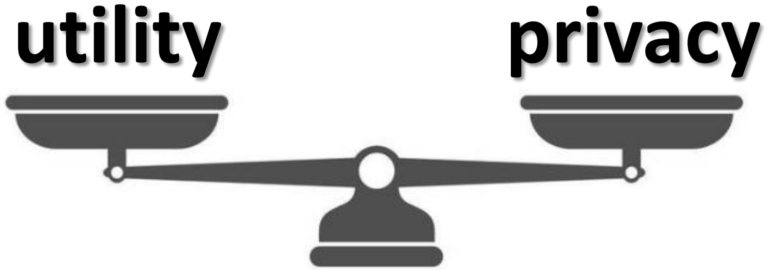
utility



utility

privacy

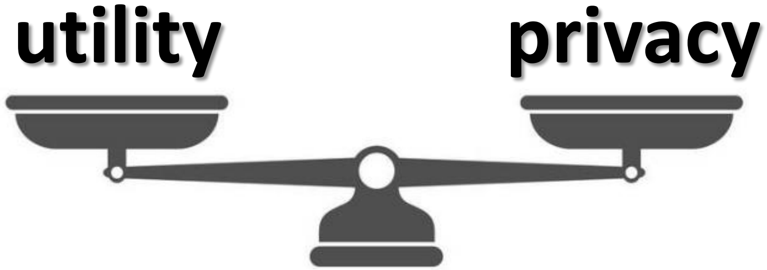
Challenges



Challenges

C1 **Utility of machine perception**

How to make HAR algorithms function properly?



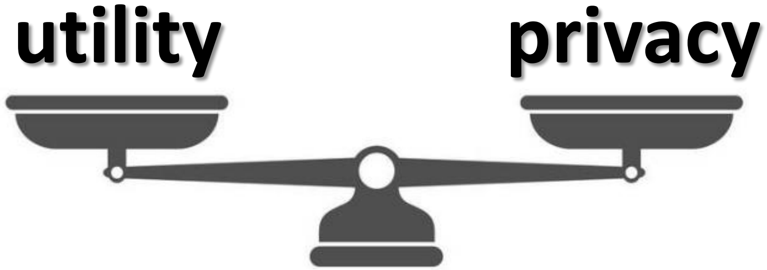
Challenges

C1 **Utility of machine perception**

How to make HAR algorithms function properly?

C2 **Utility of human perception**

How to allow human viewers to see images normally?

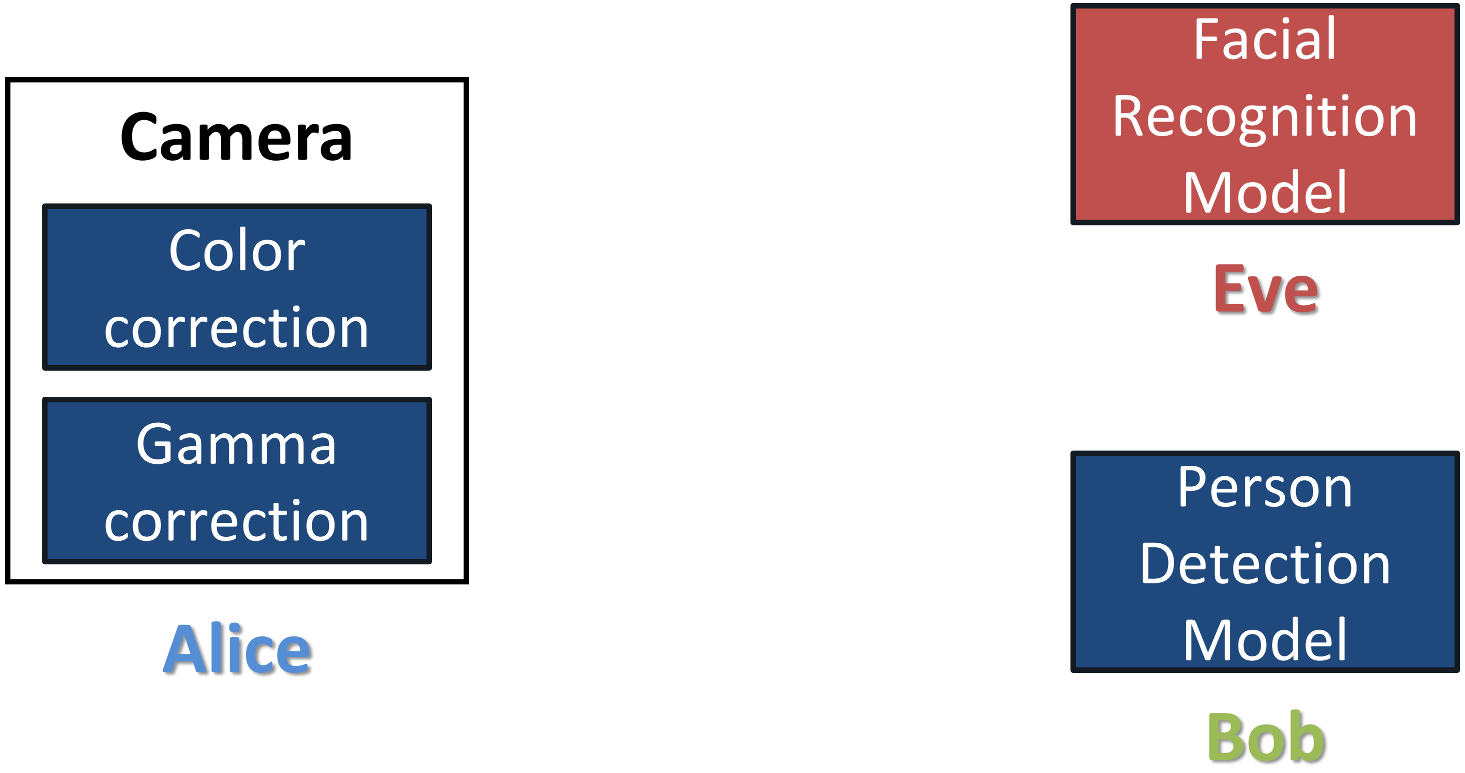


utility

privacy

C1: Optimization Problem

- Three-player game on privacy and utility



Camera

Color correction

Gamma correction

Alice

Facial Recognition Model

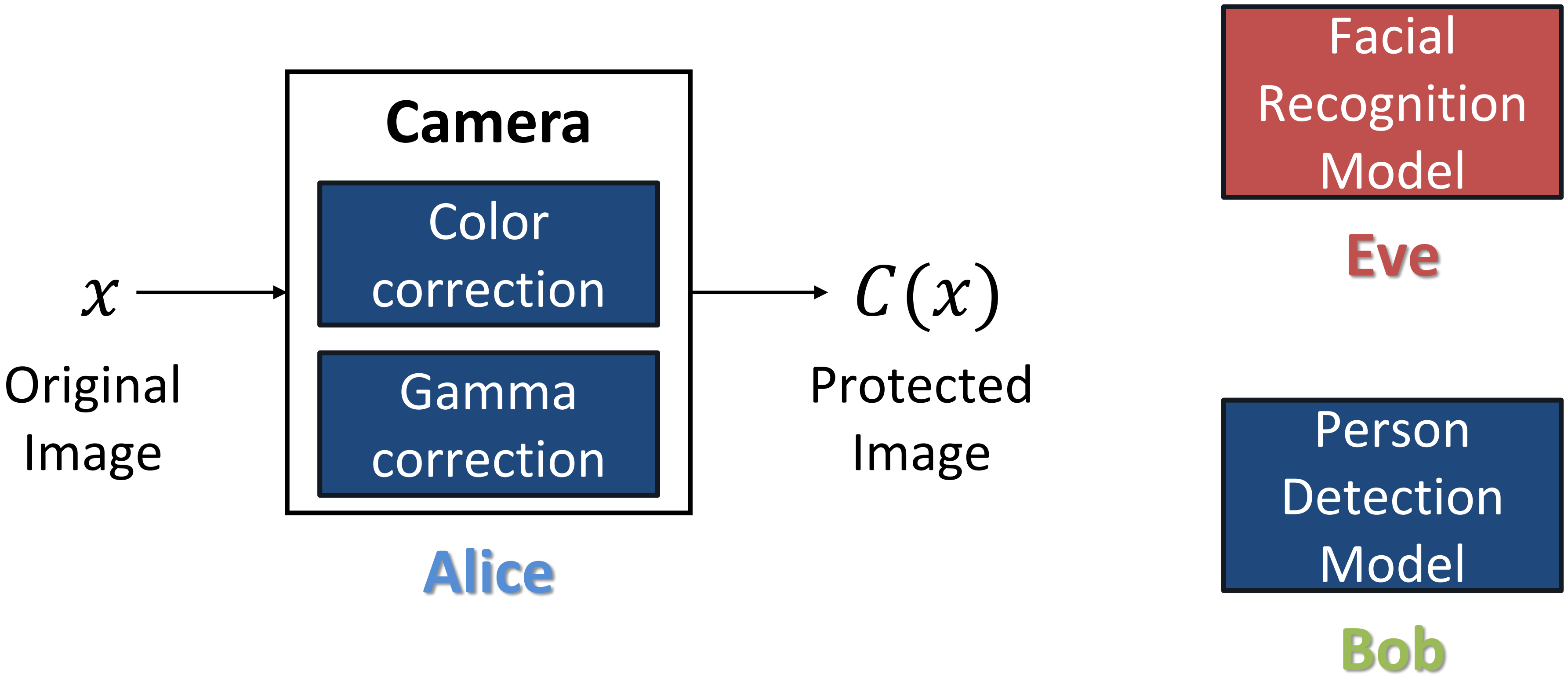
Eve

Person Detection Model

Bob

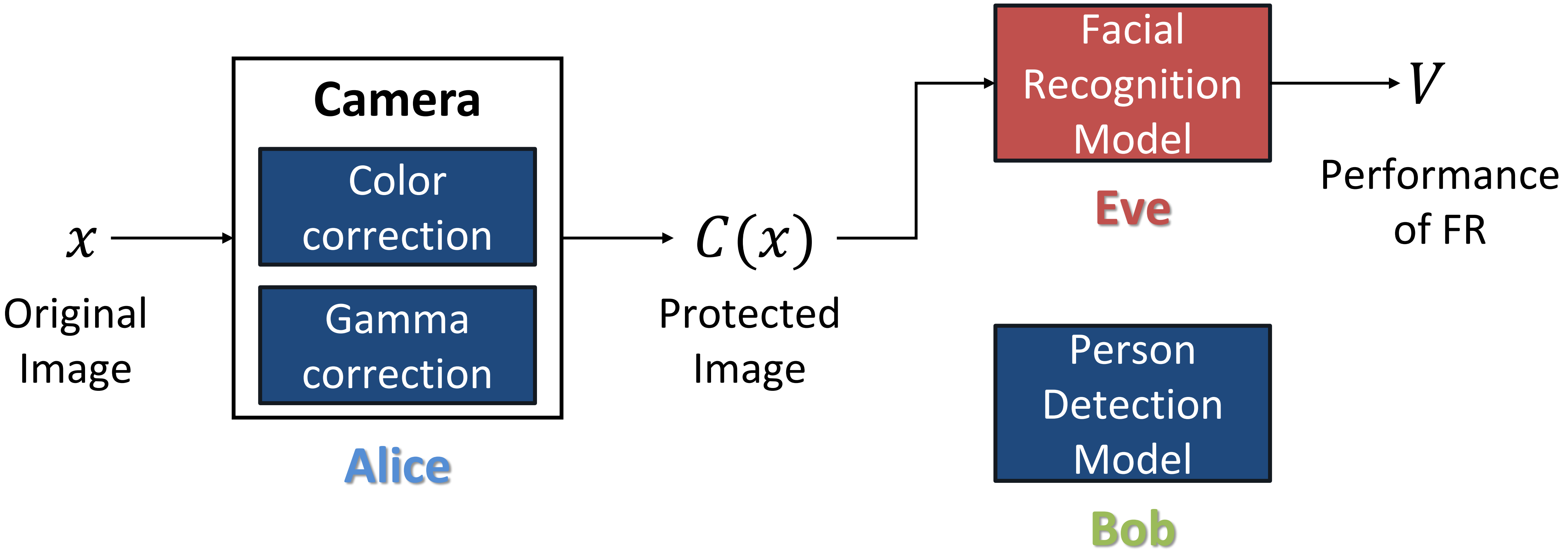
C1: Optimization Problem

- Three-player game on privacy and utility



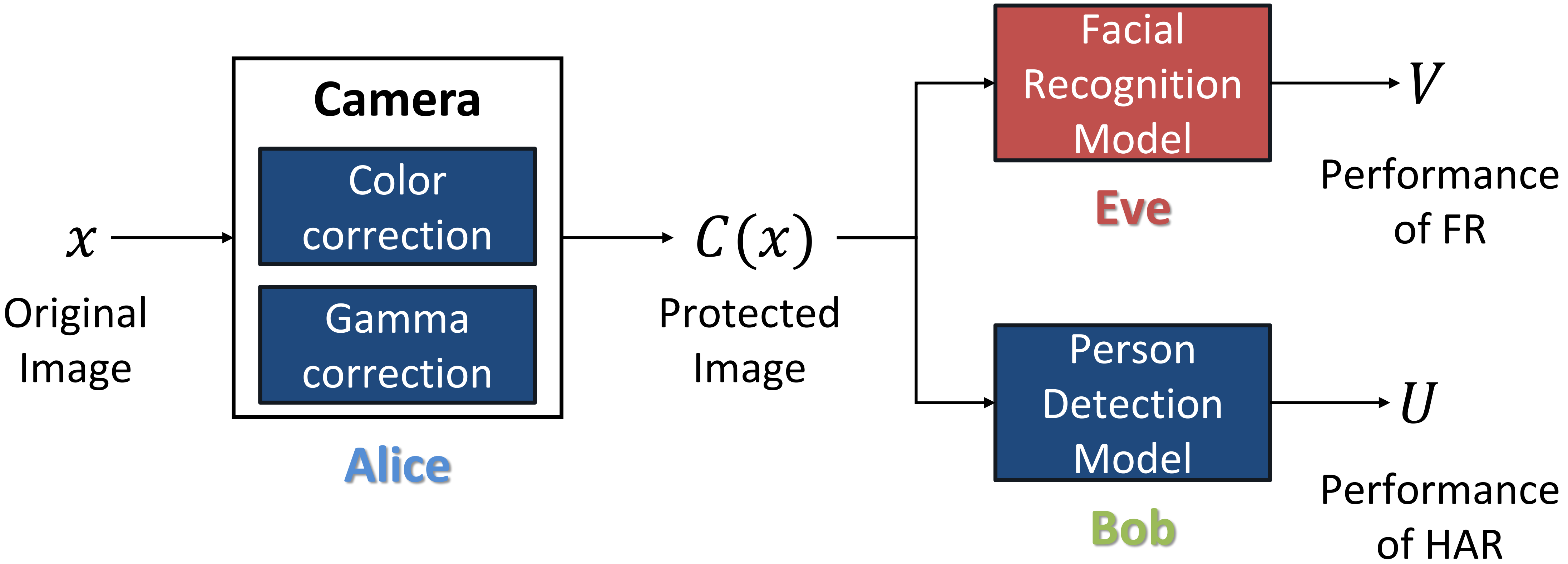
C1: Optimization Problem

- Three-player game on privacy and utility



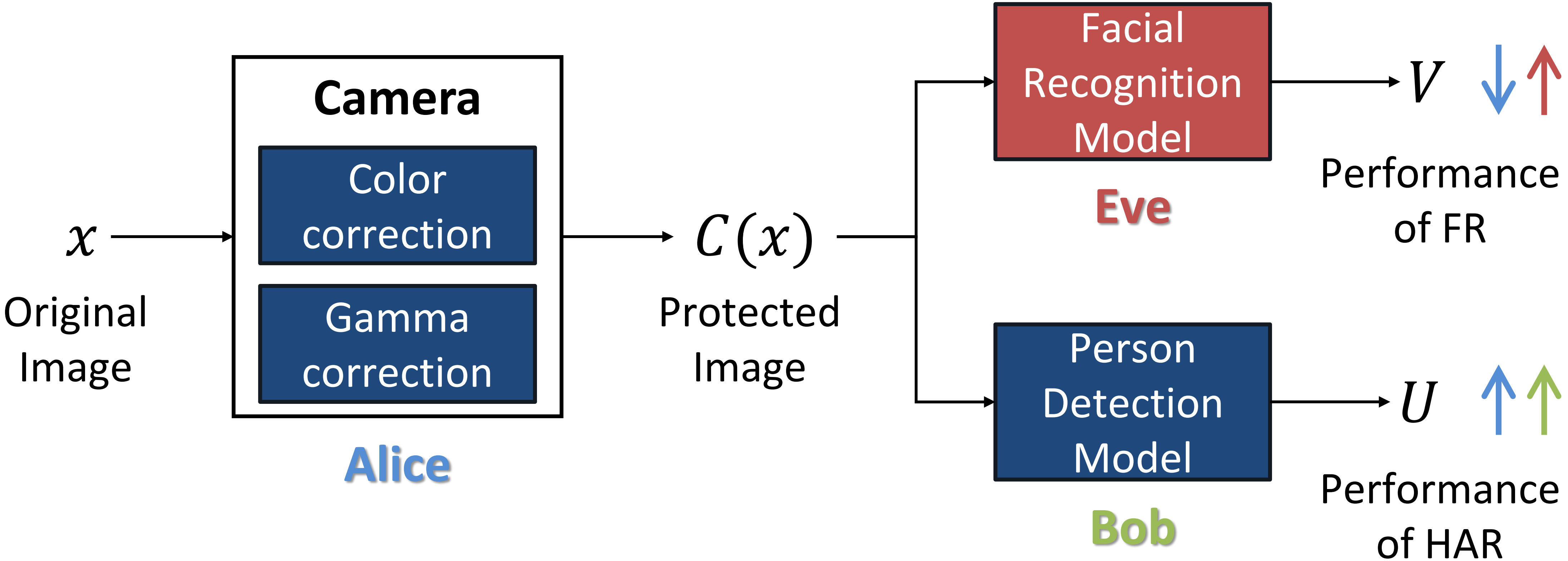
C1: Optimization Problem

- Three-player game on privacy and utility



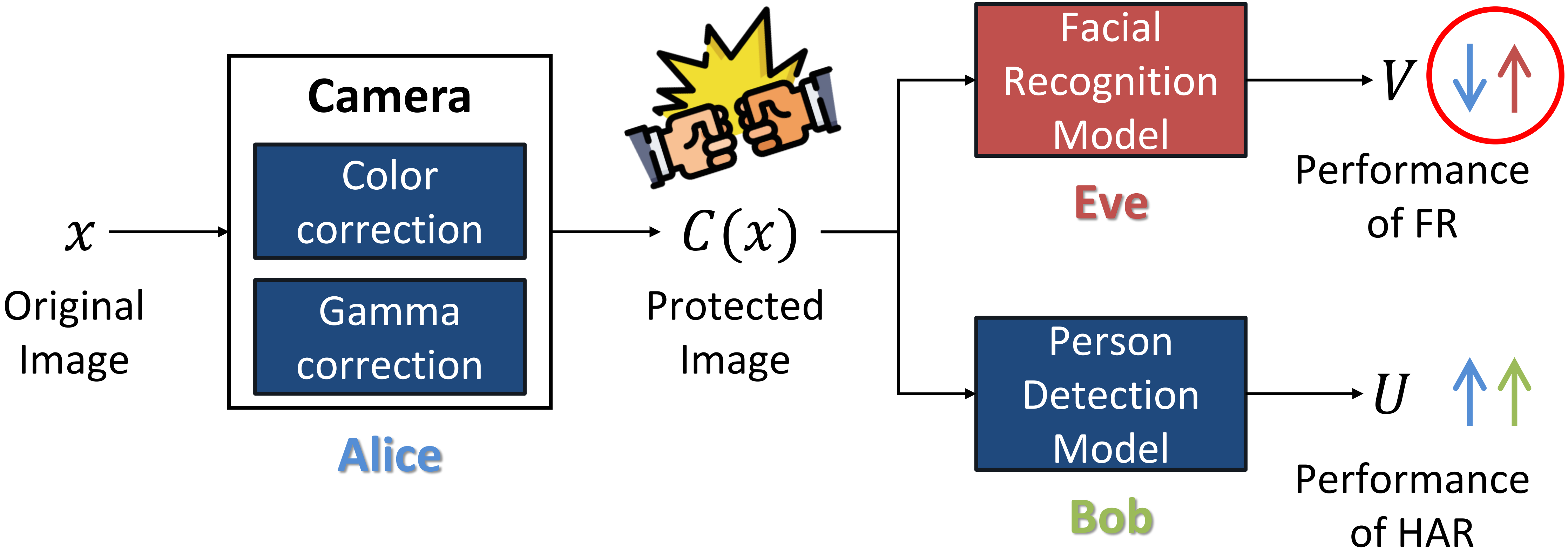
C1: Optimization Problem

- Three-player game on privacy and utility



C1: Optimization Problem

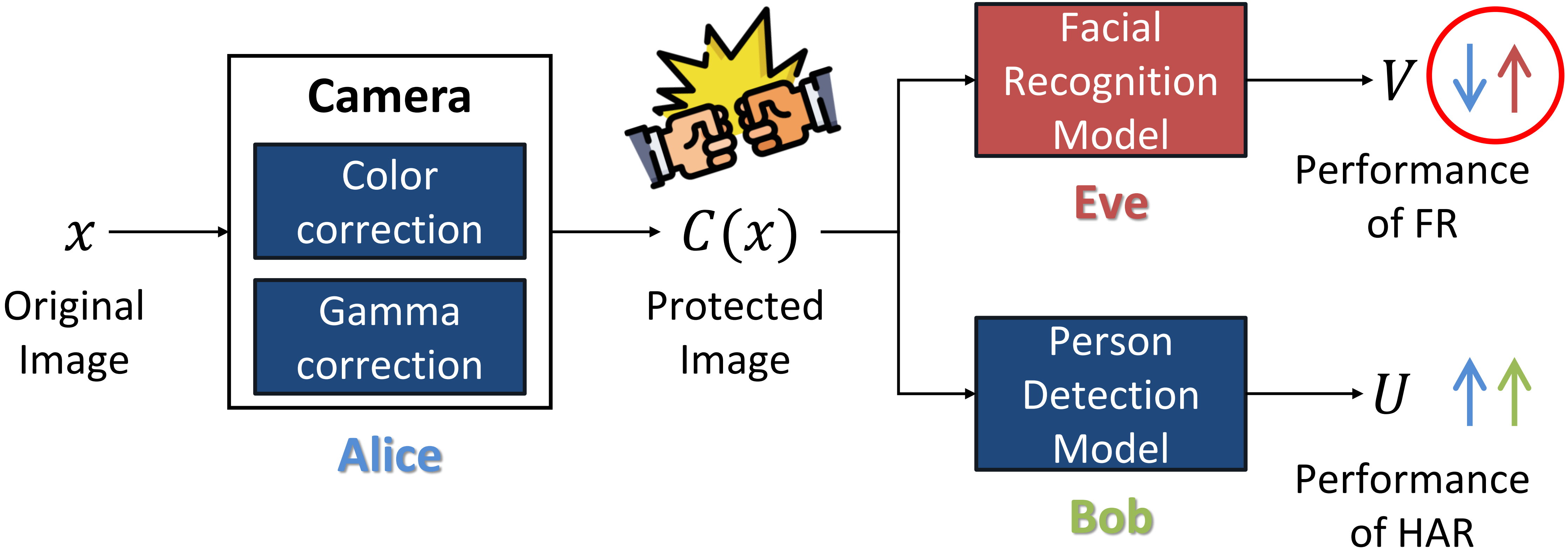
- Three-player game on privacy and utility



C1: Optimization Problem

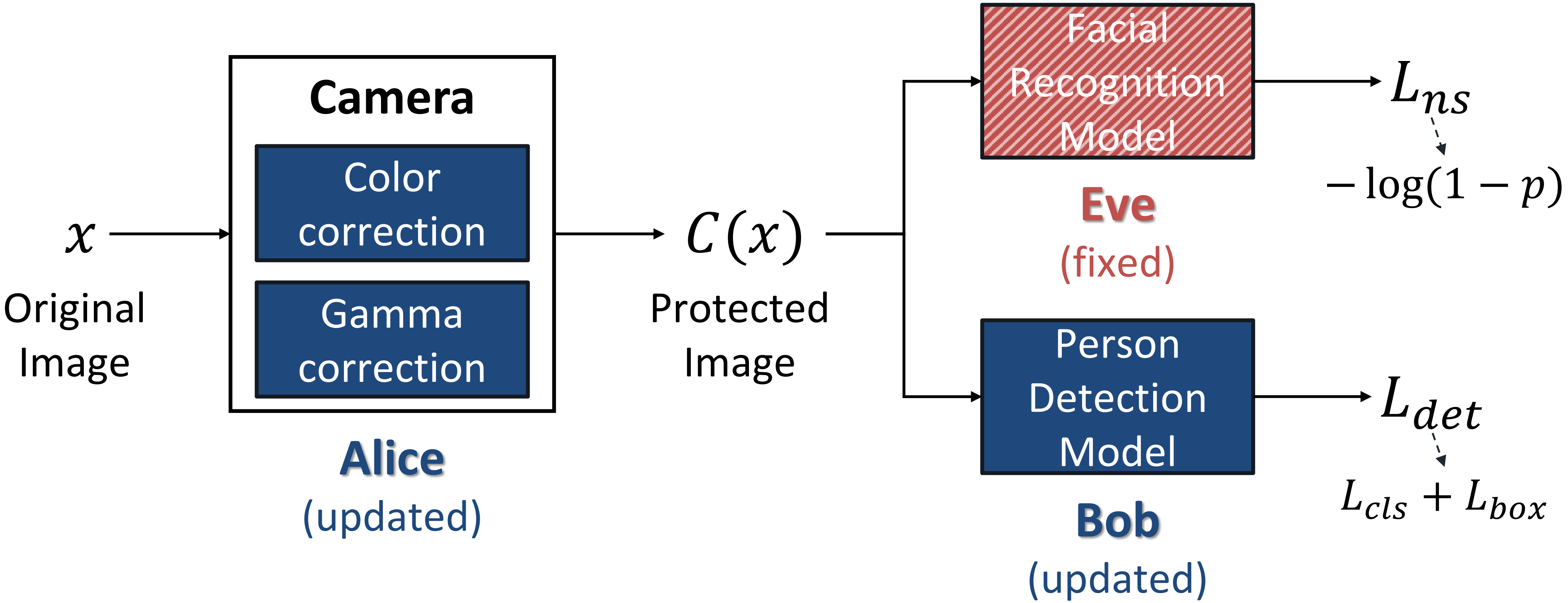
□ Three-player game on privacy and utility

□ Alternating optimization between **Protector** and **Attacker**



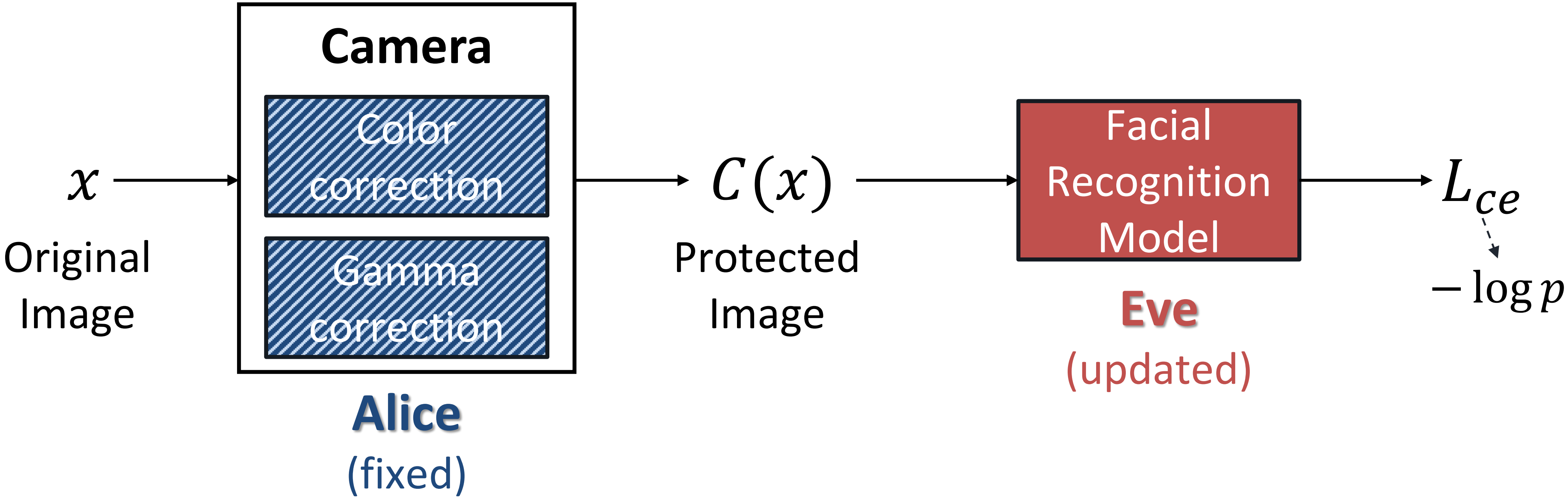
C1: Adversarial Learning Framework

Step 1: Update Protector



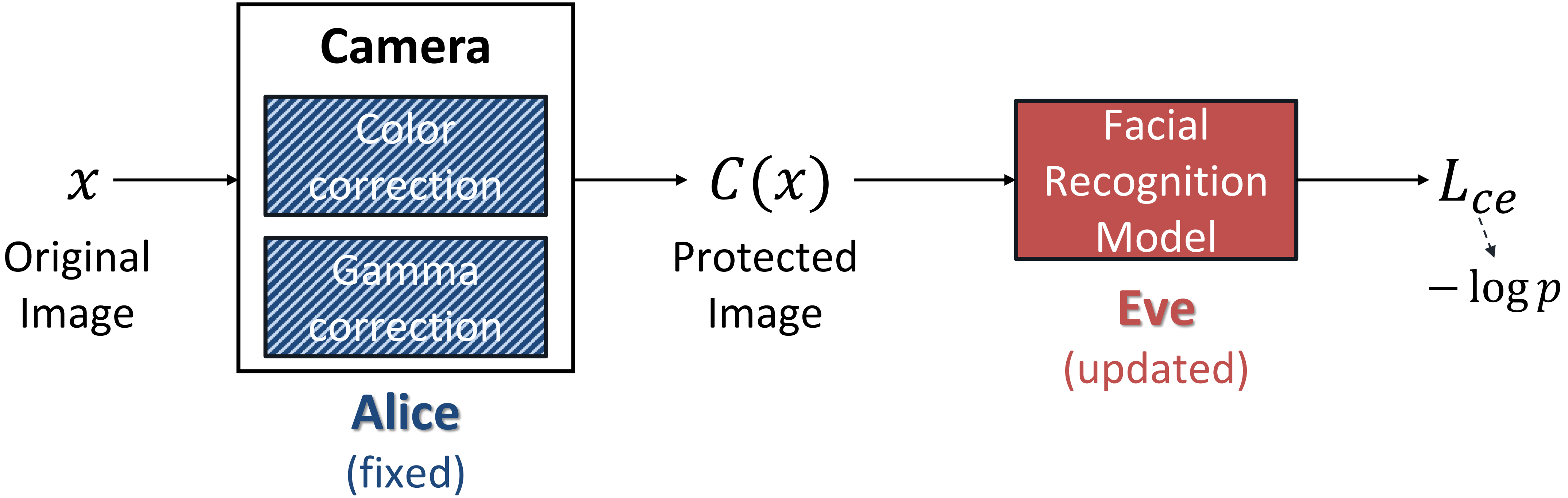
C1: Adversarial Learning Framework

Step 2: Update Attacker



C1: Adversarial Learning Framework

Step 2: Update Attacker



Protector learns **robustness** and **transferability** from an adaptive Attacker.

C2: Capacity Limitation

- □ Adjustments for AFR decrease the image quality unavoidably.

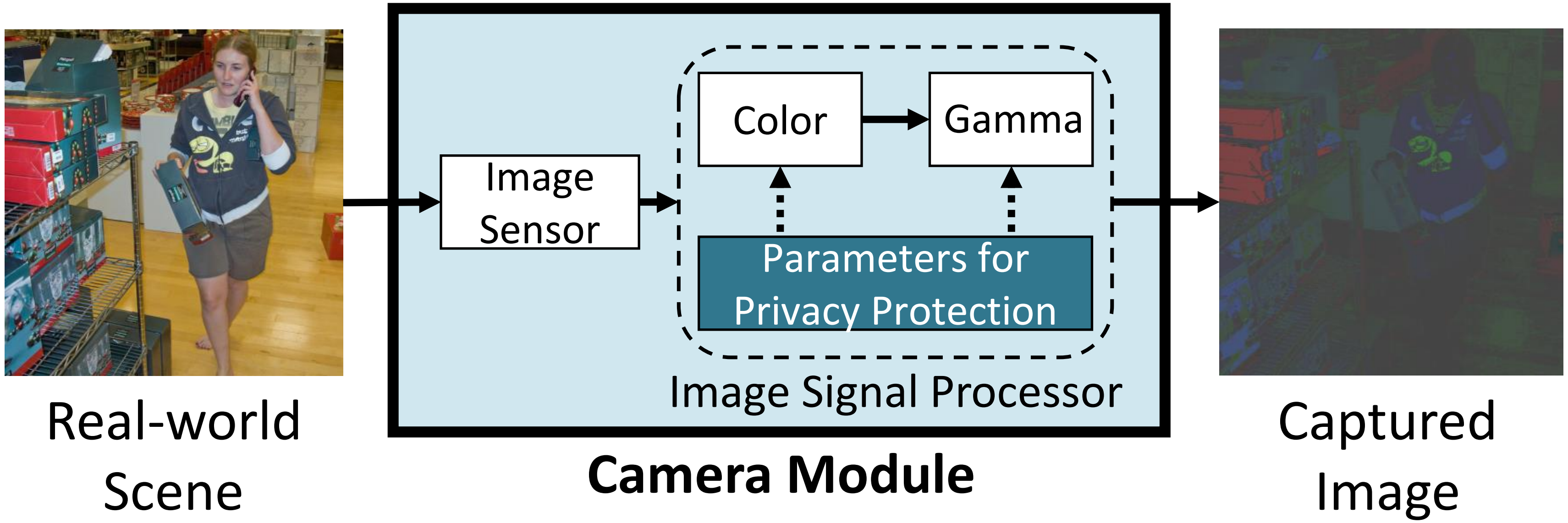


Image Sensor

Color

Gamma

Parameters for Privacy Protection

Image Signal Processor

Real-world Scene

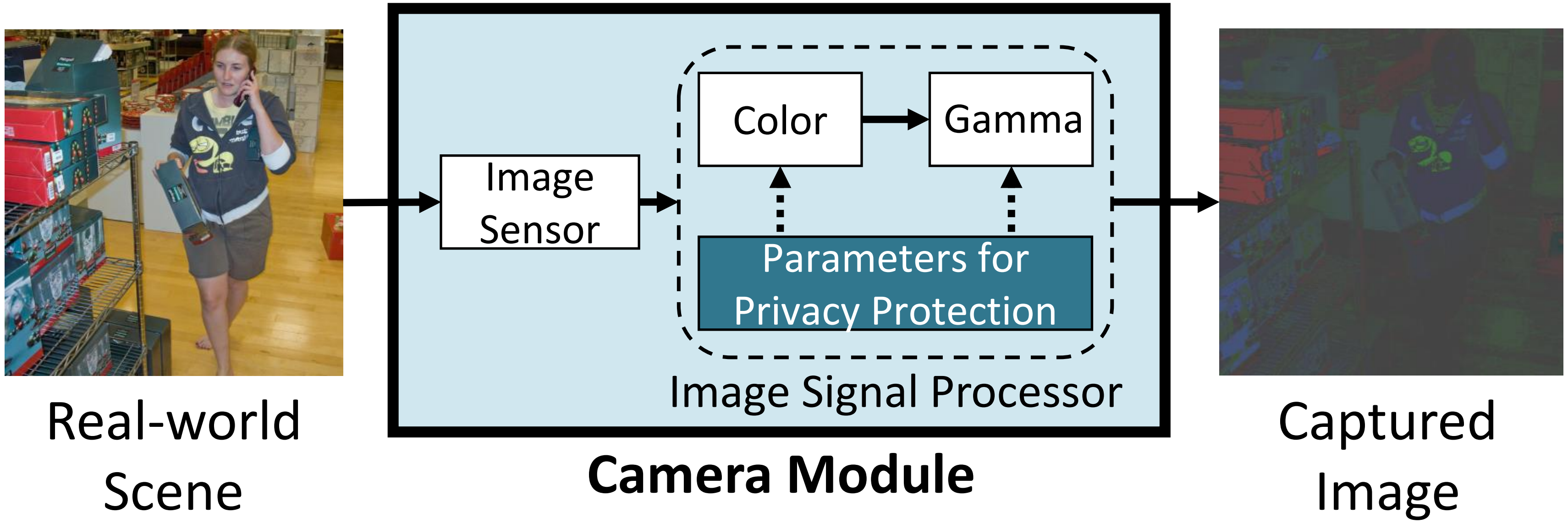
Camera Module

Captured Image

C2: Capacity Limitation

- □ Adjustments for AFR decrease the image quality unavoidably.
- □ **Limited capacity:** only 41 adjustable parameters
  - ➢ 9 in color correction
  - ➢ 32 in gamma correction

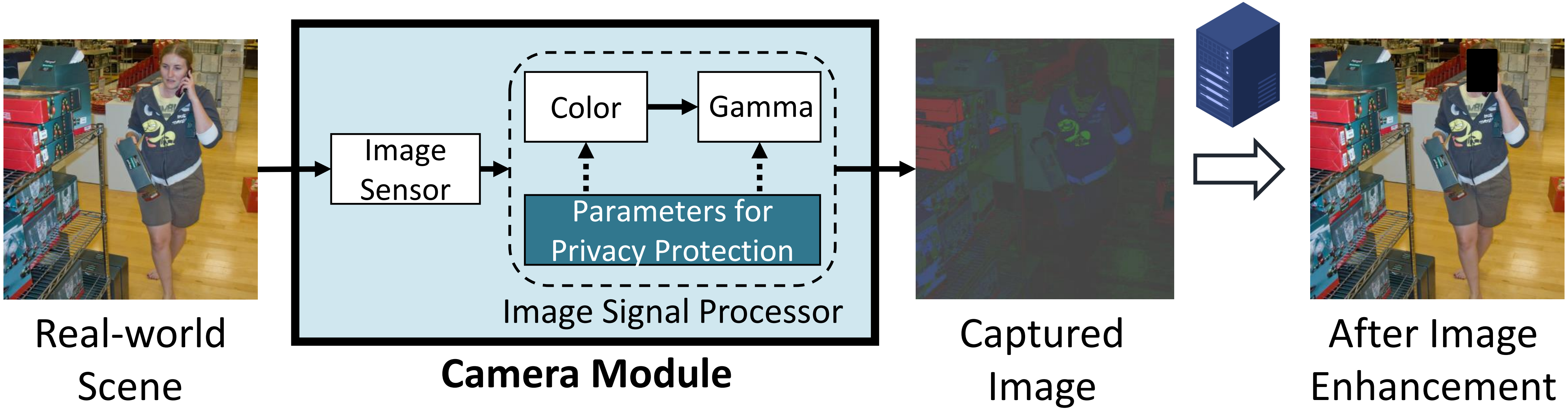
41 ≈ 1/1,000,000 of a DNN



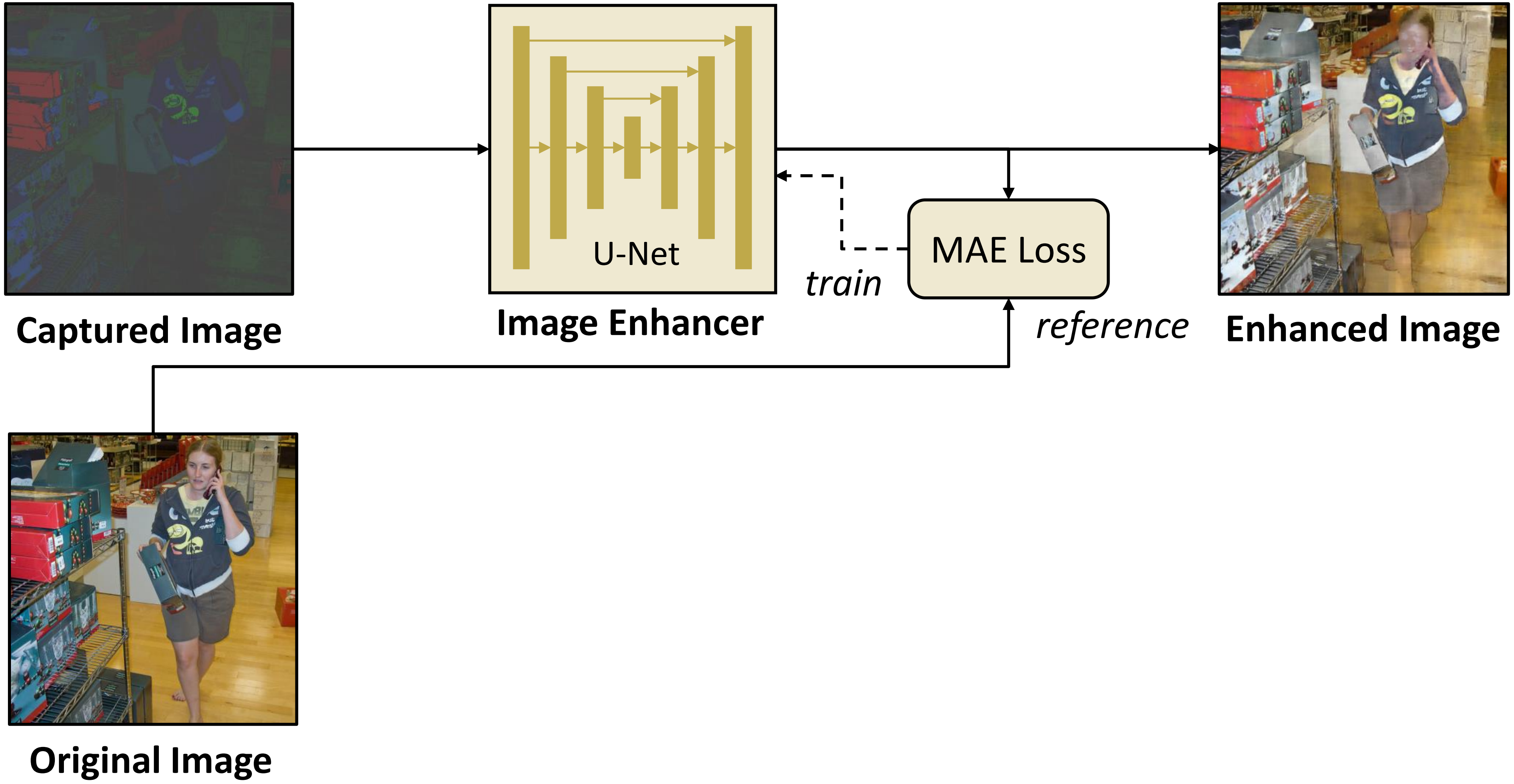
C2: Capacity Limitation

- □ Adjustments for AFR decrease the image quality unavoidably.
- □ **Limited capacity:** only 41 adjustable parameters
  - ➢ 9 in color correction
  - ➢ 32 in gamma correction

41 ≈ 1/1,000,000 of a DNN



C2: Image Enhancer



Captured Image

U-Net

Image Enhancer

train

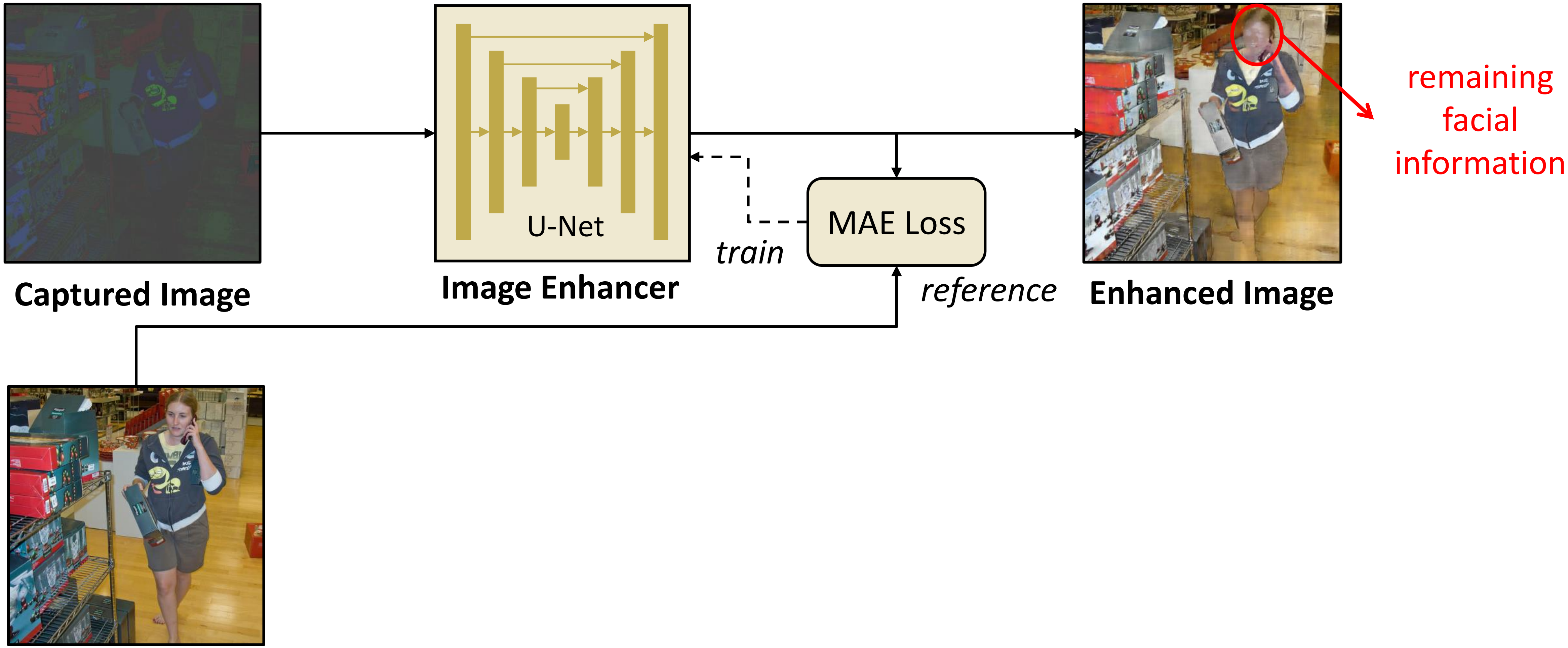
MAE Loss

reference

Enhanced Image

Original Image

C2: Image Enhancer



Captured Image

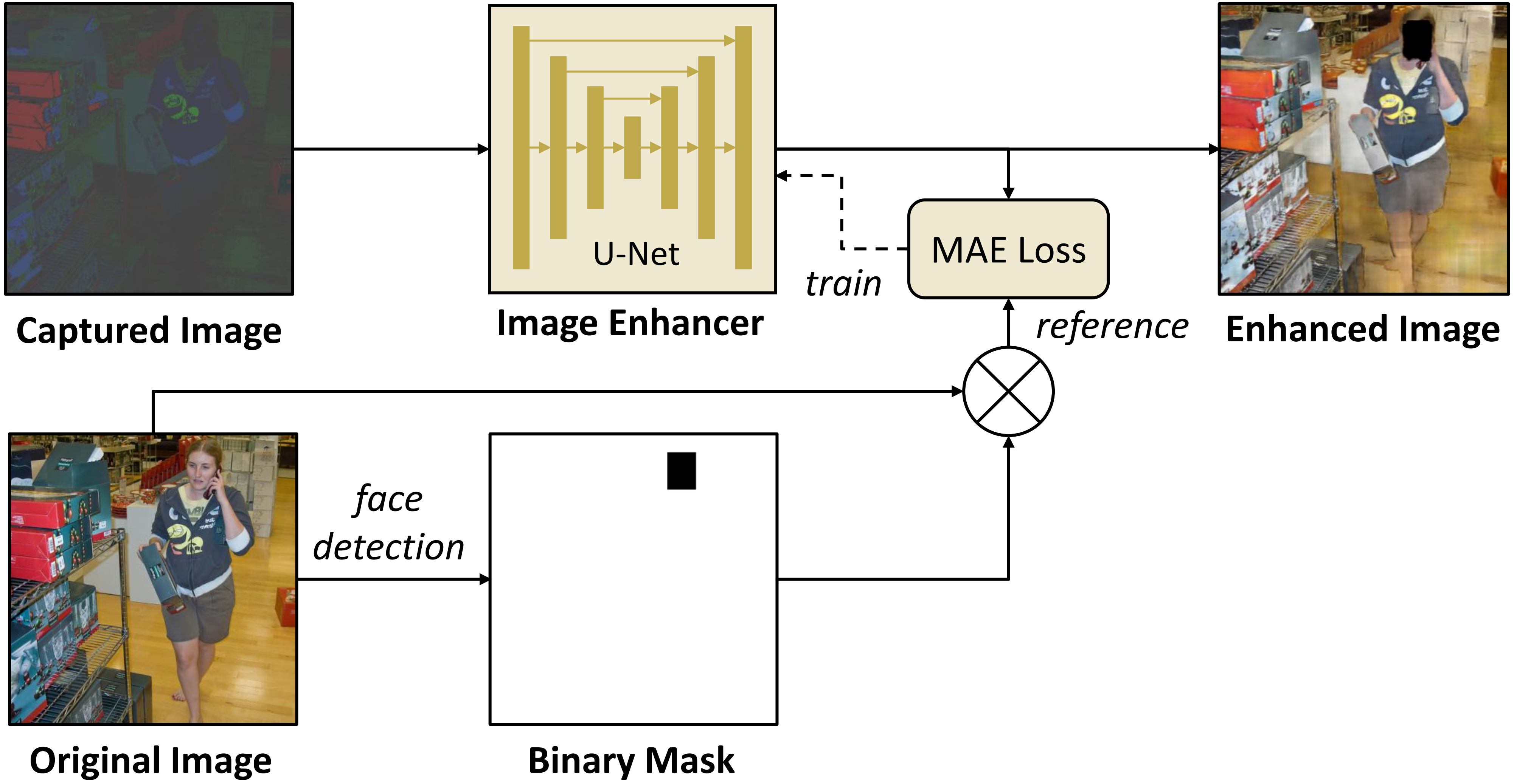
Image Enhancer

Enhanced Image

remaining facial information

Original Image

C2: Image Enhancer



Captured Image

U-Net

Image Enhancer

train

MAE Loss

reference

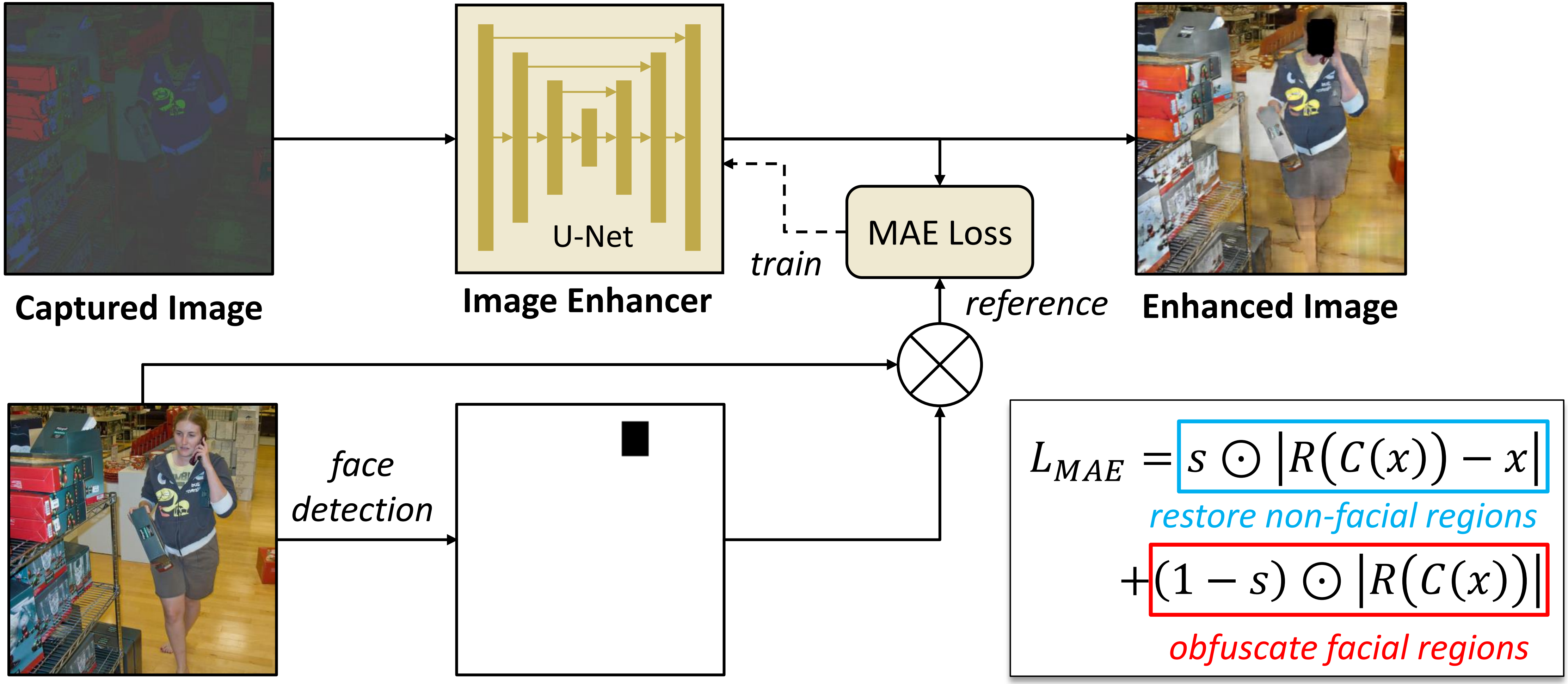
Enhanced Image

face detection

Original Image

Binary Mask

C2: Image Enhancer



Captured Image

U-Net

Image Enhancer

Enhanced Image

MAE Loss

train

reference

Original Image

face detection

Binary Mask

$$L_{MAE} = \boxed{s \odot \left|R\big(C(x)\big) - x\right|}$$

restore non-facial regions

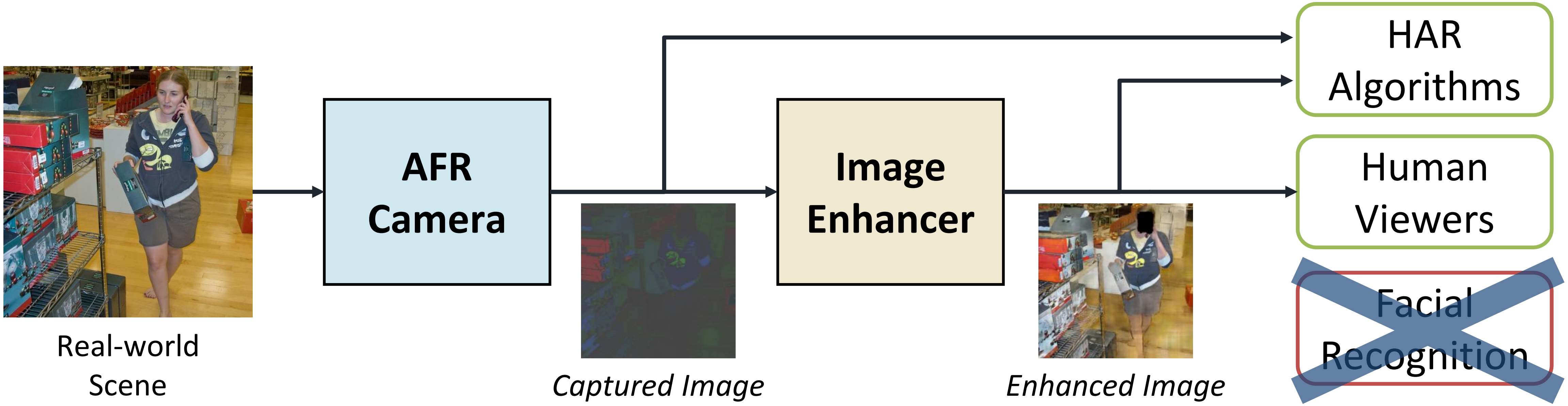
$$+\boxed{(1 - s) \odot \left|R\big(C(x)\big)\right|}$$

obfuscate facial regions

Multiple-task Training

System Design

- CamPro system
  - camera module with ISP parameter adjustments
  - image enhancer to improve the image quality



Real-world Scene

AFR Camera

Captured Image

Image Enhancer

Enhanced Image

HAR Algorithms

Human Viewers

~~Facial Recognition~~

Evaluation

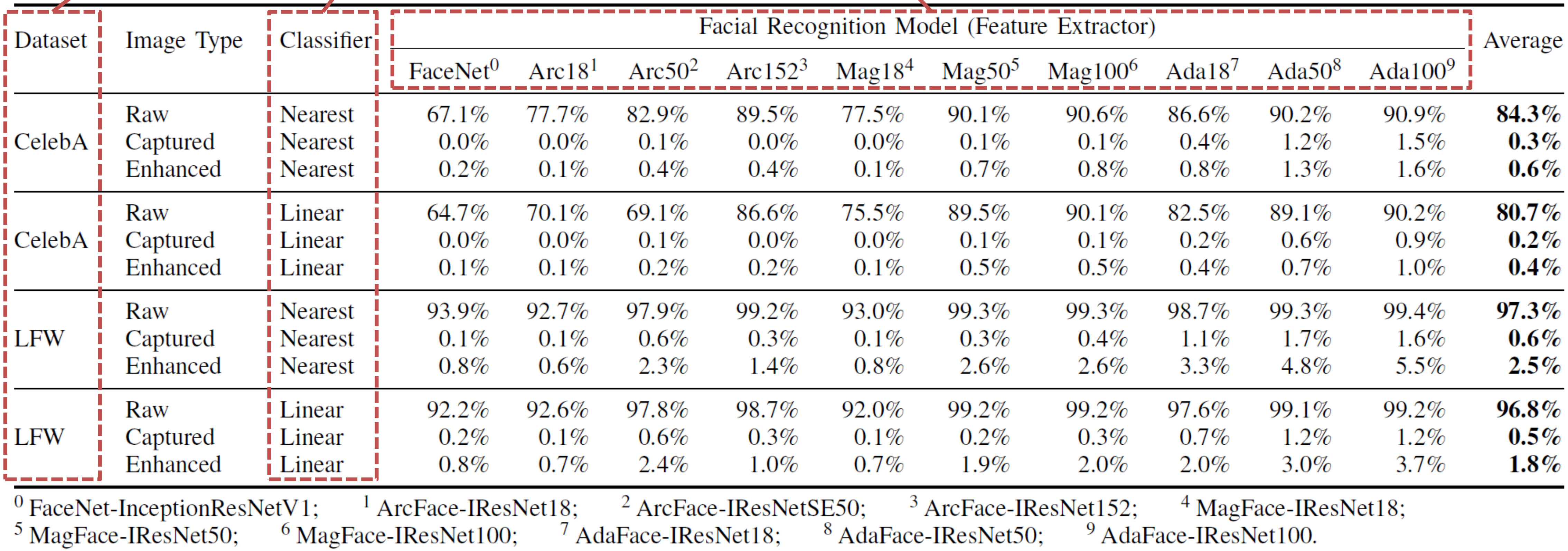
□ **E1:** Privacy protection evaluation

□ **E2:** Utility maintenance evaluation

□ **E3:** Real-world evaluation

E1: Black-box AFR Performance

2 datasets **2 classifiers** **10 models**



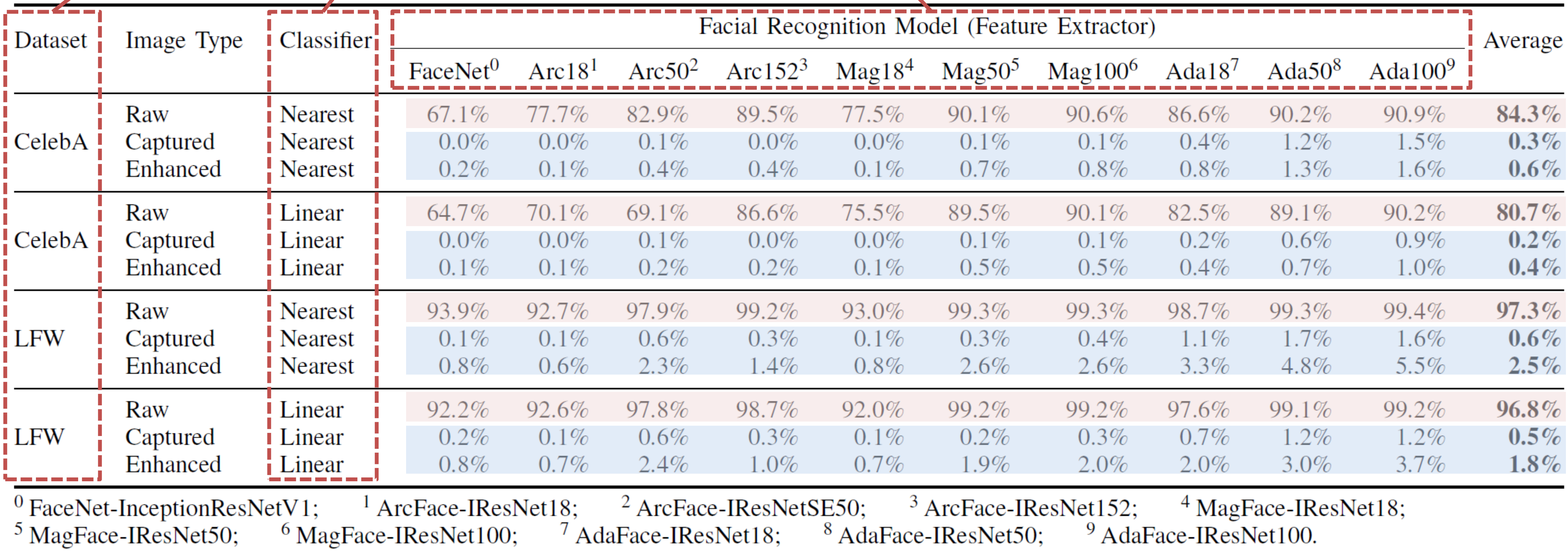
Dataset	Image Type	Classifier	Facial Recognition Model (Feature Extractor)										Average
			FaceNet[0]	Arc18[1]	Arc50[2]	Arc152[3]	Mag18[4]	Mag50[5]	Mag100[6]	Ada18[7]	Ada50[8]	Ada100[9]	
CelebA	Raw	Nearest	67.1%	77.7%	82.9%	89.5%	77.5%	90.1%	90.6%	86.6%	90.2%	90.9%	**84.3%**
	Captured	Nearest	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.4%	1.2%	1.5%	**0.3%**
	Enhanced	Nearest	0.2%	0.1%	0.4%	0.4%	0.1%	0.7%	0.8%	0.8%	1.3%	1.6%	**0.6%**
CelebA	Raw	Linear	64.7%	70.1%	69.1%	86.6%	75.5%	89.5%	90.1%	82.5%	89.1%	90.2%	**80.7%**
	Captured	Linear	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.2%	0.6%	0.9%	**0.2%**
	Enhanced	Linear	0.1%	0.1%	0.2%	0.2%	0.1%	0.5%	0.5%	0.4%	0.7%	1.0%	**0.4%**
LFW	Raw	Nearest	93.9%	92.7%	97.9%	99.2%	93.0%	99.3%	99.3%	98.7%	99.3%	99.4%	**97.3%**
	Captured	Nearest	0.1%	0.1%	0.6%	0.3%	0.1%	0.3%	0.4%	1.1%	1.7%	1.6%	**0.6%**
	Enhanced	Nearest	0.8%	0.6%	2.3%	1.4%	0.8%	2.6%	2.6%	3.3%	4.8%	5.5%	**2.5%**
LFW	Raw	Linear	92.2%	92.6%	97.8%	98.7%	92.0%	99.2%	99.2%	97.6%	99.1%	99.2%	**96.8%**
	Captured	Linear	0.2%	0.1%	0.6%	0.3%	0.1%	0.2%	0.3%	0.7%	1.2%	1.2%	**0.5%**
	Enhanced	Linear	0.8%	0.7%	2.4%	1.0%	0.7%	1.9%	2.0%	2.0%	3.0%	3.7%	**1.8%**

[0] FaceNet-InceptionResNetV1; [1] ArcFace-IResNet18; [2] ArcFace-IResNetSE50; [3] ArcFace-IResNet152; [4] MagFace-IResNet18; [5] MagFace-IResNet50; [6] MagFace-IResNet100; [7] AdaFace-IResNet18; [8] AdaFace-IResNet50; [9] AdaFace-IResNet100.

NDSS SYMPOSIUM/2024

E1: Black-box AFR Performance

2 datasets **2 classifiers** **10 models**



Dataset	Image Type	Classifier	Facial Recognition Model (Feature Extractor)										Average
			FaceNet[0]	Arc18[1]	Arc50[2]	Arc152[3]	Mag18[4]	Mag50[5]	Mag100[6]	Ada18[7]	Ada50[8]	Ada100[9]	
CelebA	Raw	Nearest	67.1%	77.7%	82.9%	89.5%	77.5%	90.1%	90.6%	86.6%	90.2%	90.9%	**84.3%**
	Captured	Nearest	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.4%	1.2%	1.5%	**0.3%**
	Enhanced	Nearest	0.2%	0.1%	0.4%	0.4%	0.1%	0.7%	0.8%	0.8%	1.3%	1.6%	**0.6%**
CelebA	Raw	Linear	64.7%	70.1%	69.1%	86.6%	75.5%	89.5%	90.1%	82.5%	89.1%	90.2%	**80.7%**
	Captured	Linear	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.2%	0.6%	0.9%	**0.2%**
	Enhanced	Linear	0.1%	0.1%	0.2%	0.2%	0.1%	0.5%	0.5%	0.4%	0.7%	1.0%	**0.4%**
LFW	Raw	Nearest	93.9%	92.7%	97.9%	99.2%	93.0%	99.3%	99.3%	98.7%	99.3%	99.4%	**97.3%**
	Captured	Nearest	0.1%	0.1%	0.6%	0.3%	0.1%	0.3%	0.4%	1.1%	1.7%	1.6%	**0.6%**
	Enhanced	Nearest	0.8%	0.6%	2.3%	1.4%	0.8%	2.6%	2.6%	3.3%	4.8%	5.5%	**2.5%**
LFW	Raw	Linear	92.2%	92.6%	97.8%	98.7%	92.0%	99.2%	99.2%	97.6%	99.1%	99.2%	**96.8%**
	Captured	Linear	0.2%	0.1%	0.6%	0.3%	0.1%	0.2%	0.3%	0.7%	1.2%	1.2%	**0.5%**
	Enhanced	Linear	0.8%	0.7%	2.4%	1.0%	0.7%	1.9%	2.0%	2.0%	3.0%	3.7%	**1.8%**

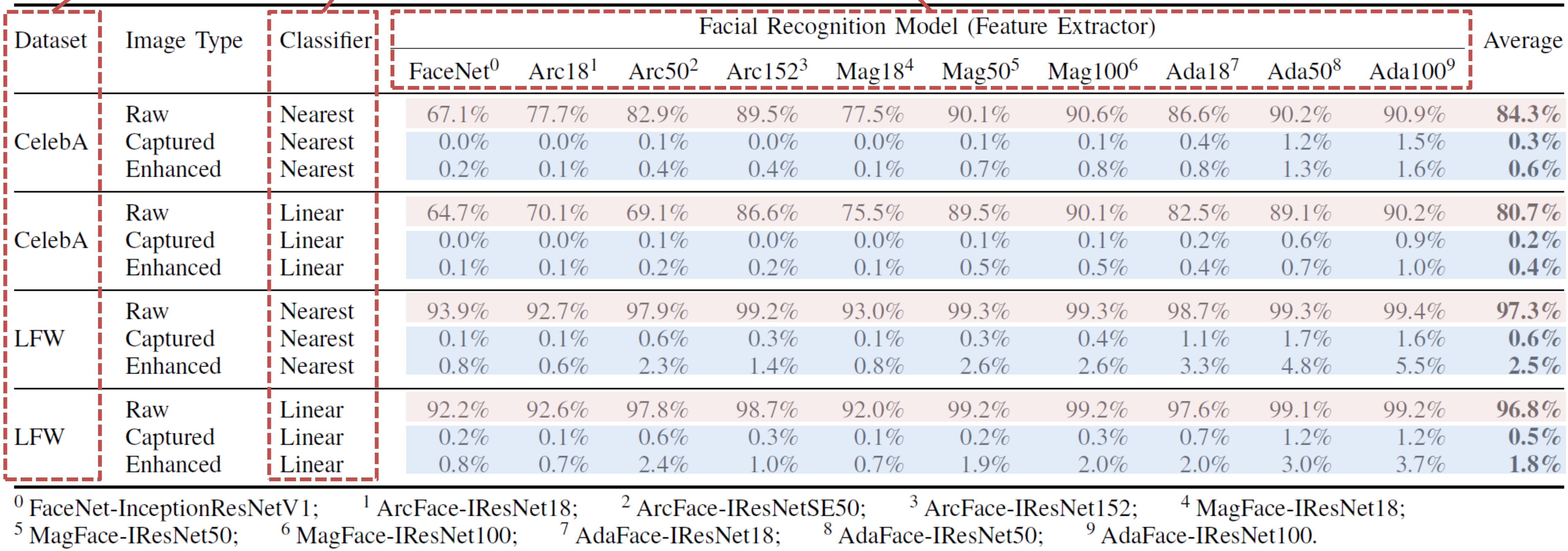
[0] FaceNet-InceptionResNetV1; [1] ArcFace-IResNet18; [2] ArcFace-IResNetSE50; [3] ArcFace-IResNet152; [4] MagFace-IResNet18; [5] MagFace-IResNet50; [6] MagFace-IResNet100; [7] AdaFace-IResNet18; [8] AdaFace-IResNet50; [9] AdaFace-IResNet100.

E1: Black-box AFR Performance

2 datasets

2 classifiers

10 models



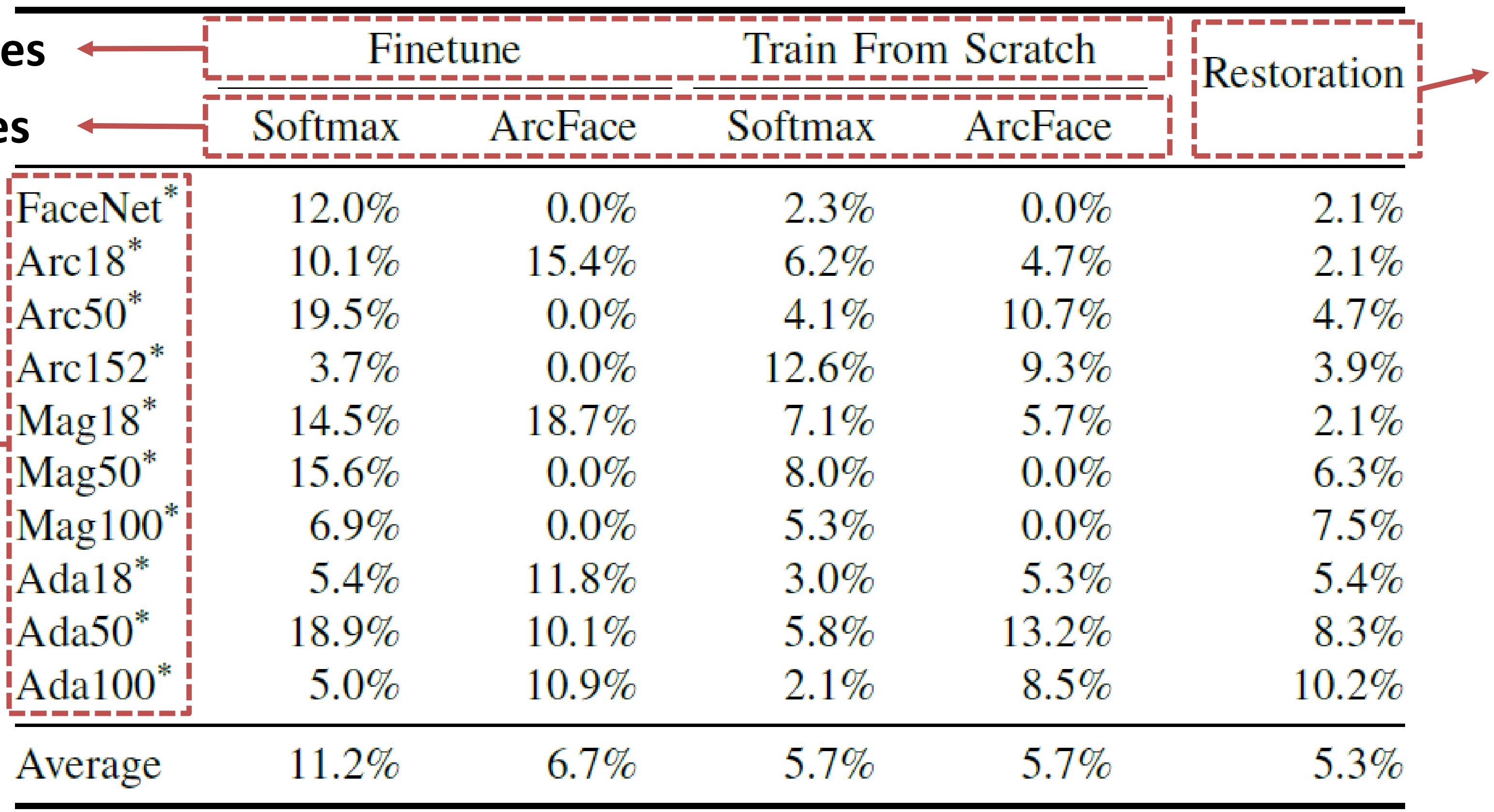
Dataset	Image Type	Classifier	Facial Recognition Model (Feature Extractor)										Average
			FaceNet[0]	Arc18[1]	Arc50[2]	Arc152[3]	Mag18[4]	Mag50[5]	Mag100[6]	Ada18[7]	Ada50[8]	Ada100[9]	
CelebA	Raw	Nearest	67.1%	77.7%	82.9%	89.5%	77.5%	90.1%	90.6%	86.6%	90.2%	90.9%	**84.3%**
	Captured	Nearest	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.4%	1.2%	1.5%	**0.3%**
	Enhanced	Nearest	0.2%	0.1%	0.4%	0.4%	0.1%	0.7%	0.8%	0.8%	1.3%	1.6%	**0.6%**
CelebA	Raw	Linear	64.7%	70.1%	69.1%	86.6%	75.5%	89.5%	90.1%	82.5%	89.1%	90.2%	**80.7%**
	Captured	Linear	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.1%	0.2%	0.6%	0.9%	**0.2%**
	Enhanced	Linear	0.1%	0.1%	0.2%	0.2%	0.1%	0.5%	0.5%	0.4%	0.7%	1.0%	**0.4%**
LFW	Raw	Nearest	93.9%	92.7%	97.9%	99.2%	93.0%	99.3%	99.3%	98.7%	99.3%	99.4%	**97.3%**
	Captured	Nearest	0.1%	0.1%	0.6%	0.3%	0.1%	0.3%	0.4%	1.1%	1.7%	1.6%	**0.6%**
	Enhanced	Nearest	0.8%	0.6%	2.3%	1.4%	0.8%	2.6%	2.6%	3.3%	4.8%	5.5%	**2.5%**
LFW	Raw	Linear	92.2%	92.6%	97.8%	98.7%	92.0%	99.2%	99.2%	97.6%	99.1%	99.2%	**96.8%**
	Captured	Linear	0.2%	0.1%	0.6%	0.3%	0.1%	0.2%	0.3%	0.7%	1.2%	1.2%	**0.5%**
	Enhanced	Linear	0.8%	0.7%	2.4%	1.0%	0.7%	1.9%	2.0%	2.0%	3.0%	3.7%	**1.8%**

[0] FaceNet-InceptionResNetV1; [1] ArcFace-IResNet18; [2] ArcFace-IResNetSE50; [3] ArcFace-IResNet152; [4] MagFace-IResNet18; [5] MagFace-IResNet50; [6] MagFace-IResNet100; [7] AdaFace-IResNet18; [8] AdaFace-IResNet50; [9] AdaFace-IResNet100.

The AFR effects of CamPro can transfer to various models, classifiers, and datasets.

NDSS SYMPOSIUM/2024

E1: White-box Adaptive Attack



	Finetune		Train From Scratch		Restoration
	Softmax	ArcFace	Softmax	ArcFace	
FaceNet*	12.0%	0.0%	2.3%	0.0%	2.1%
Arc18*	10.1%	15.4%	6.2%	4.7%	2.1%
Arc50*	19.5%	0.0%	4.1%	10.7%	4.7%
Arc152*	3.7%	0.0%	12.6%	9.3%	3.9%
Mag18*	14.5%	18.7%	7.1%	5.7%	2.1%
Mag50*	15.6%	0.0%	8.0%	0.0%	6.3%
Mag100*	6.9%	0.0%	5.3%	0.0%	7.5%
Ada18*	5.4%	11.8%	3.0%	5.3%	5.4%
Ada50*	18.9%	10.1%	5.8%	13.2%	8.3%
Ada100*	5.0%	10.9%	2.1%	8.5%	10.2%
Average	11.2%	6.7%	5.7%	5.7%	5.3%

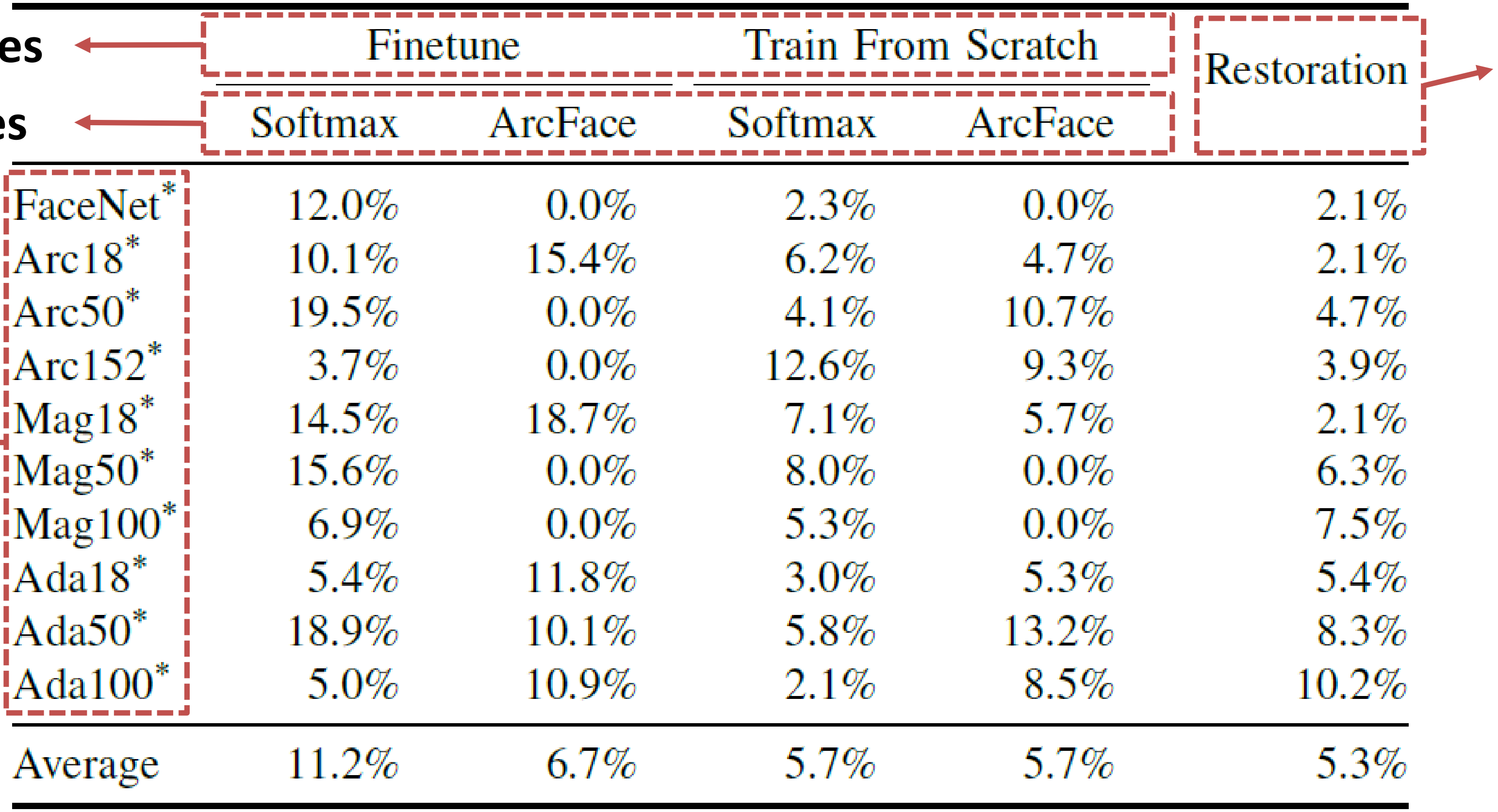
2 training modes

2 training losses

10 models

Image restoration with U-Net

E1: White-box Adaptive Attack



	Finetune		Train From Scratch		Restoration
	Softmax	ArcFace	Softmax	ArcFace	
FaceNet*	12.0%	0.0%	2.3%	0.0%	2.1%
Arc18*	10.1%	15.4%	6.2%	4.7%	2.1%
Arc50*	19.5%	0.0%	4.1%	10.7%	4.7%
Arc152*	3.7%	0.0%	12.6%	9.3%	3.9%
Mag18*	14.5%	18.7%	7.1%	5.7%	2.1%
Mag50*	15.6%	0.0%	8.0%	0.0%	6.3%
Mag100*	6.9%	0.0%	5.3%	0.0%	7.5%
Ada18*	5.4%	11.8%	3.0%	5.3%	5.4%
Ada50*	18.9%	10.1%	5.8%	13.2%	8.3%
Ada100*	5.0%	10.9%	2.1%	8.5%	10.2%
Average	11.2%	6.7%	5.7%	5.7%	5.3%

2 training modes

2 training losses

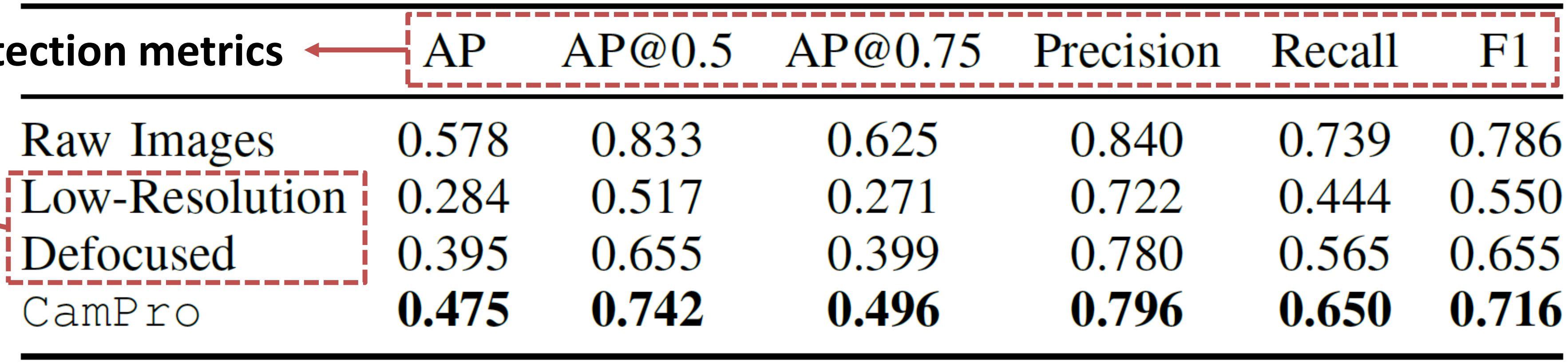
10 models

Image restoration with U-Net

CamPro is, to some extent, resistant to white-box adaptive attacks.

E2: Quantitative Results

- Person detection performance



Detection metrics

2 baseline methods

	AP	AP@0.5	AP@0.75	Precision	Recall	F1
Raw Images	0.578	0.833	0.625	0.840	0.739	0.786
Low-Resolution	0.284	0.517	0.271	0.722	0.444	0.550
Defocused	0.395	0.655	0.399	0.780	0.565	0.655
CamPro	**0.475**	**0.742**	**0.496**	**0.796**	**0.650**	**0.716**

E2: Quantitative Results

□ Person detection performance

Detection metrics

2 baseline methods

	AP	AP@0.5	AP@0.75	Precision	Recall	F1
Raw Images	0.578	0.833	0.625	0.840	0.739	0.786
Low-Resolution	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Defocused	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CamPro	**0.475**	**0.742**	**0.496**	**0.796**	**0.650**	**0.716**

CamPro decreases 18% AP of person detection, outperforming the baselines (51% and 32%).

□ Image quality

Treat raw images as ground truth

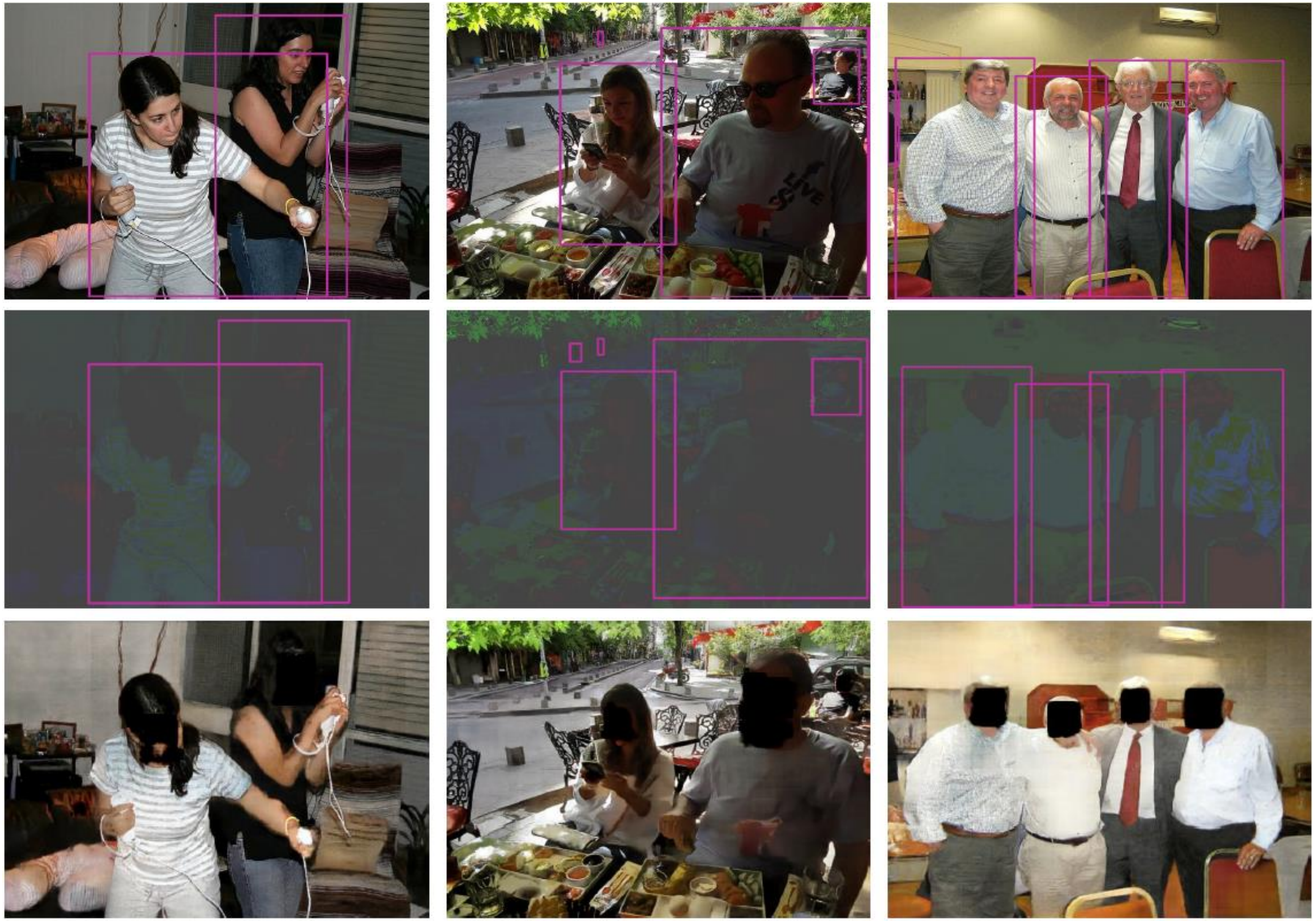
Image Type	RMSE ↓	PSNR ↑	SSIM ↑	MS-SSIM ↑
Captured	0.299	10.8 dB	0.437	0.195
Enhanced	**0.093**	**21.5** dB	**0.749**	**0.761**

E2: Qualitative Results

- Raw images

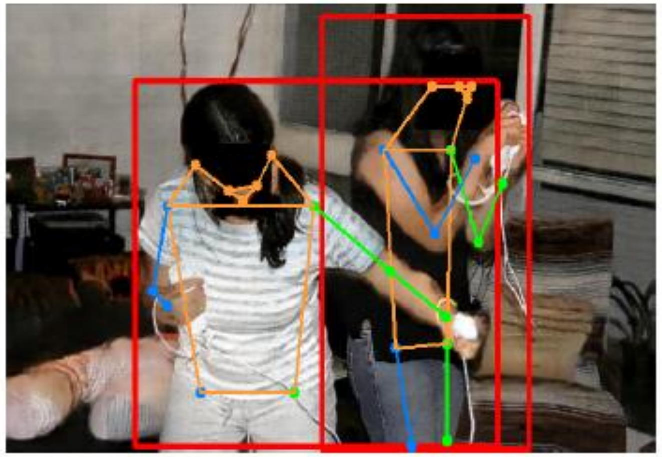
- Captured images

- Enhanced images

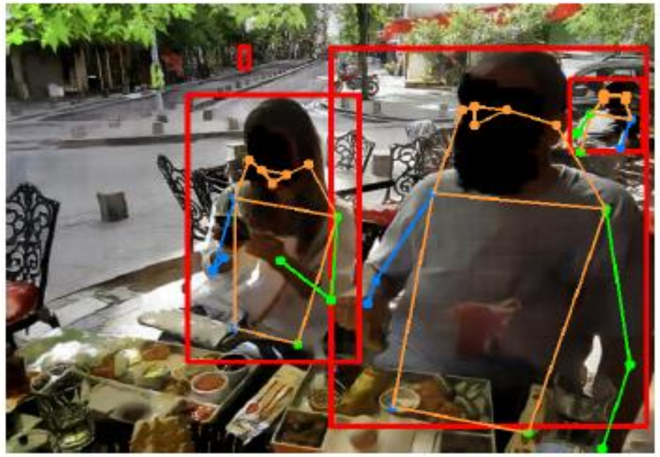


E2: Generalization Ability

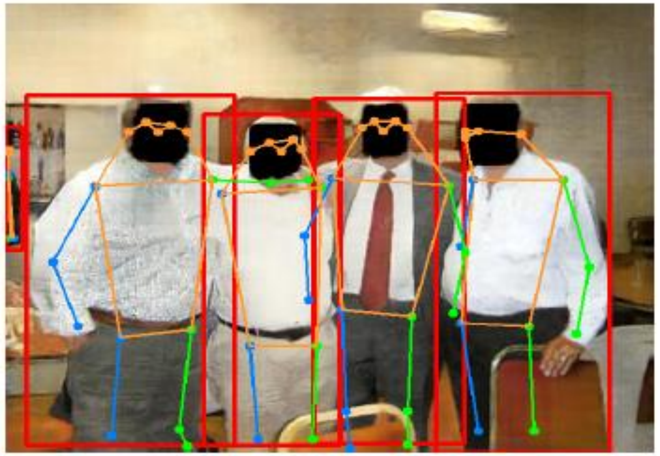
□ Generalized to **pose estimation** and **image captioning**



(d) *“Two people playing a video game in a living room.”*

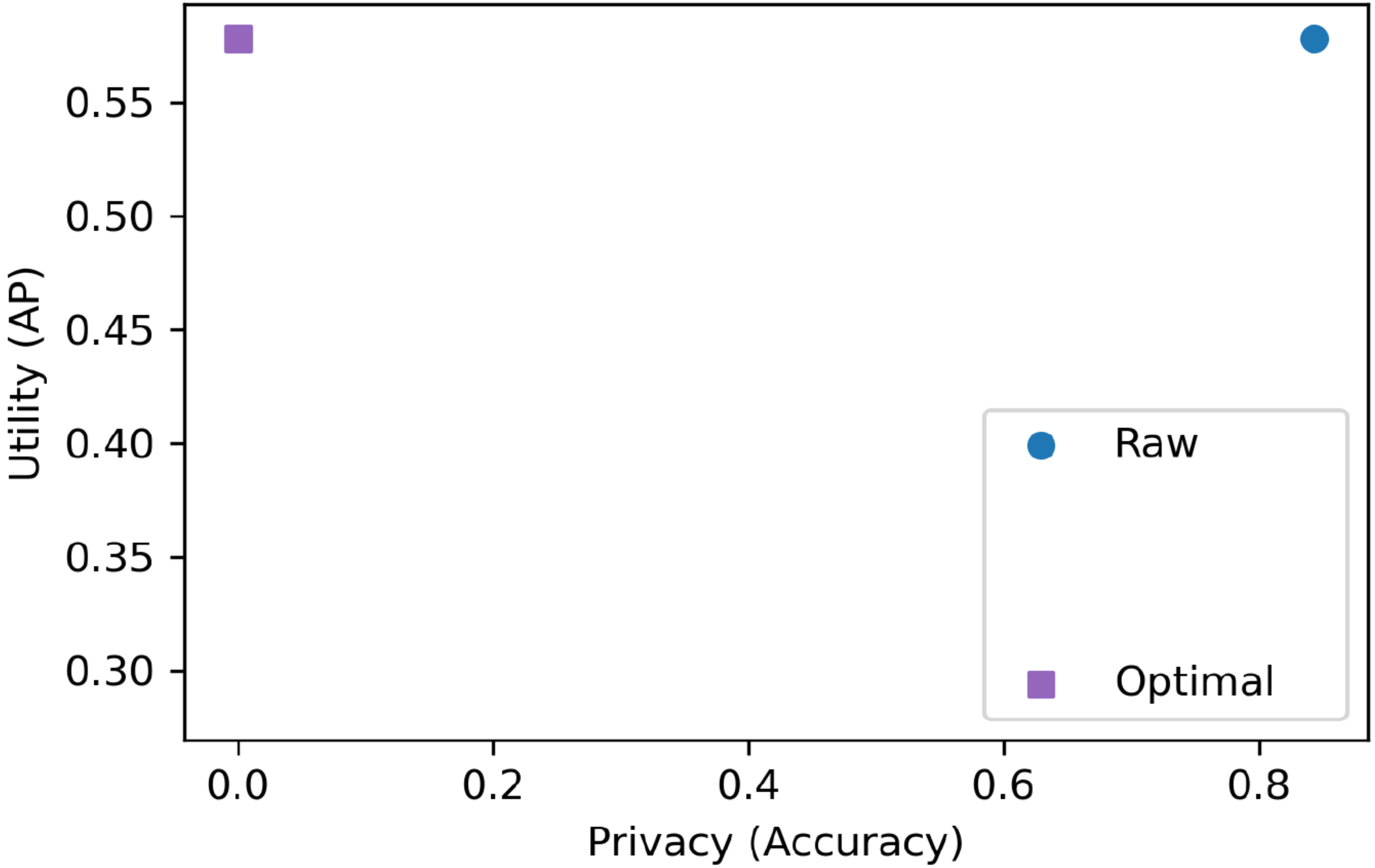


(e) *“A man and a woman sitting at a table with food.”*

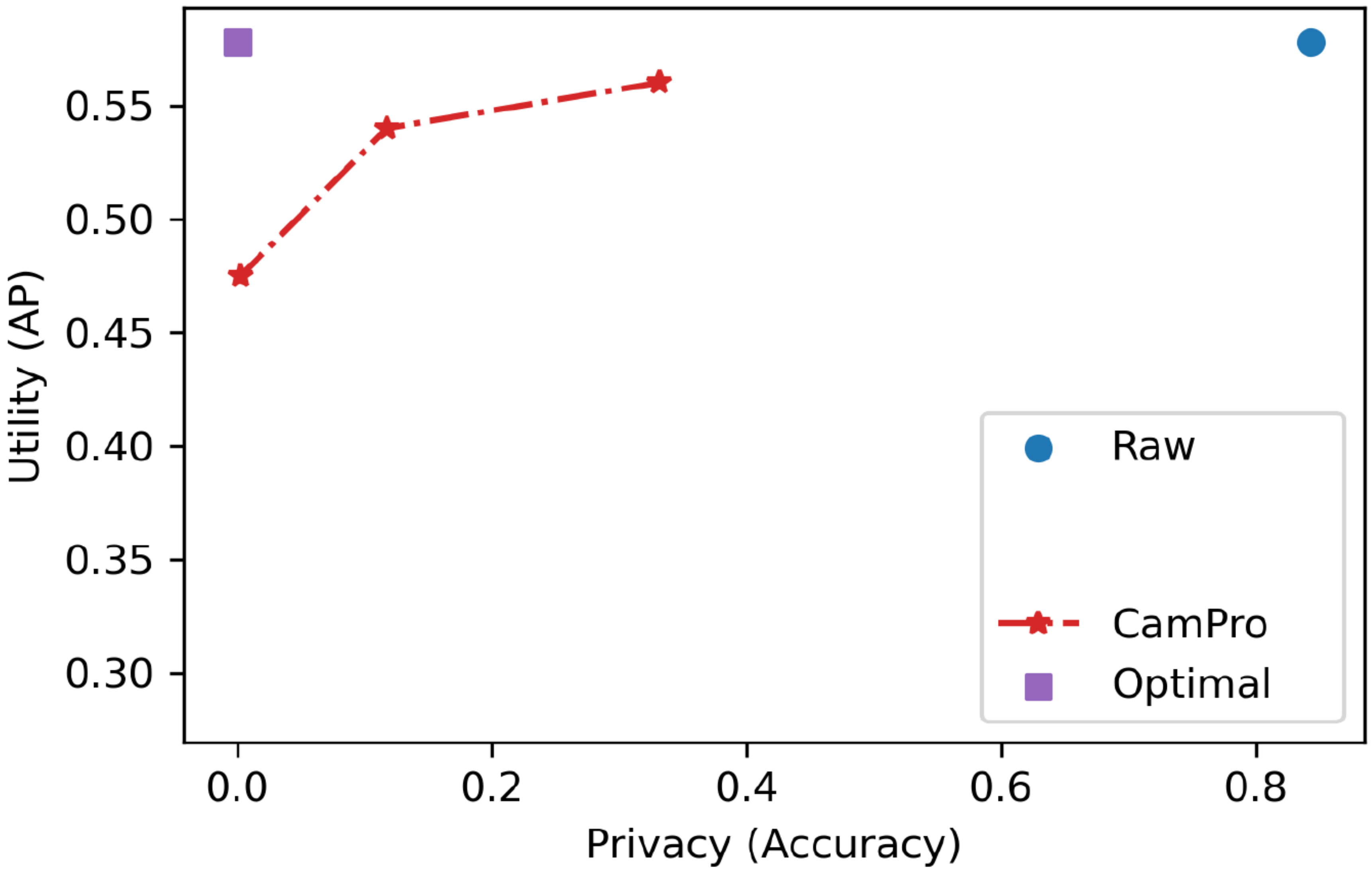


(f) *“A group of men standing next to each other.”*

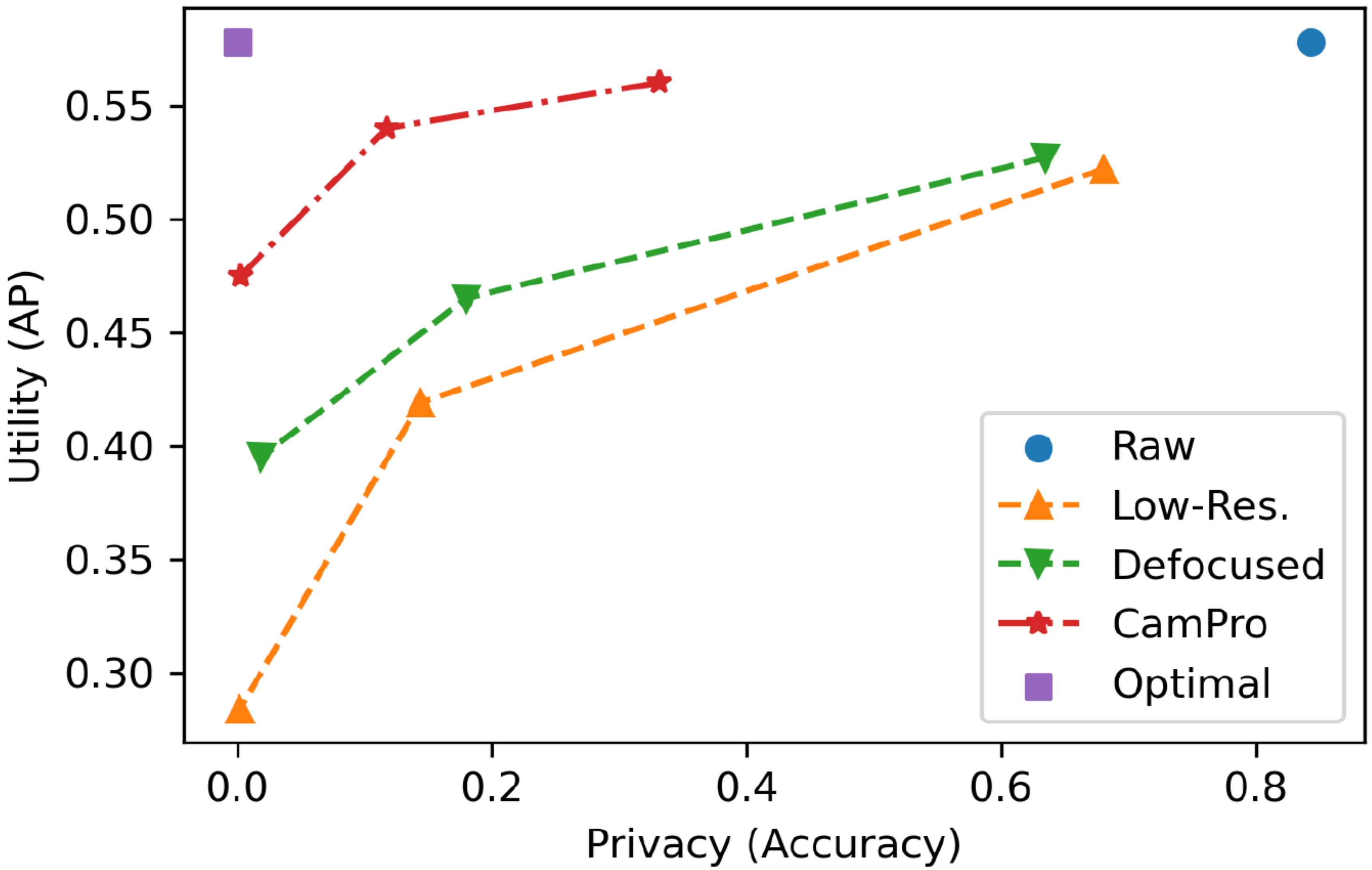
E1 & E2: Privacy-Utility Tradeoff Analysis



E1 & E2: Privacy-Utility Tradeoff Analysis

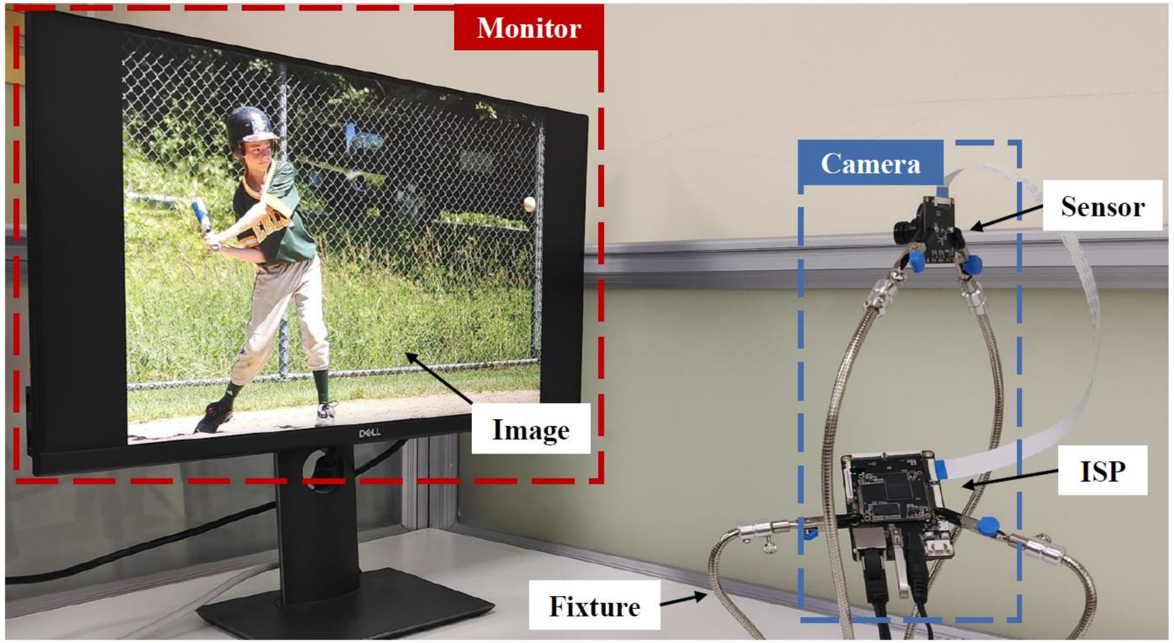


E1 & E2: Privacy-Utility Tradeoff Analysis



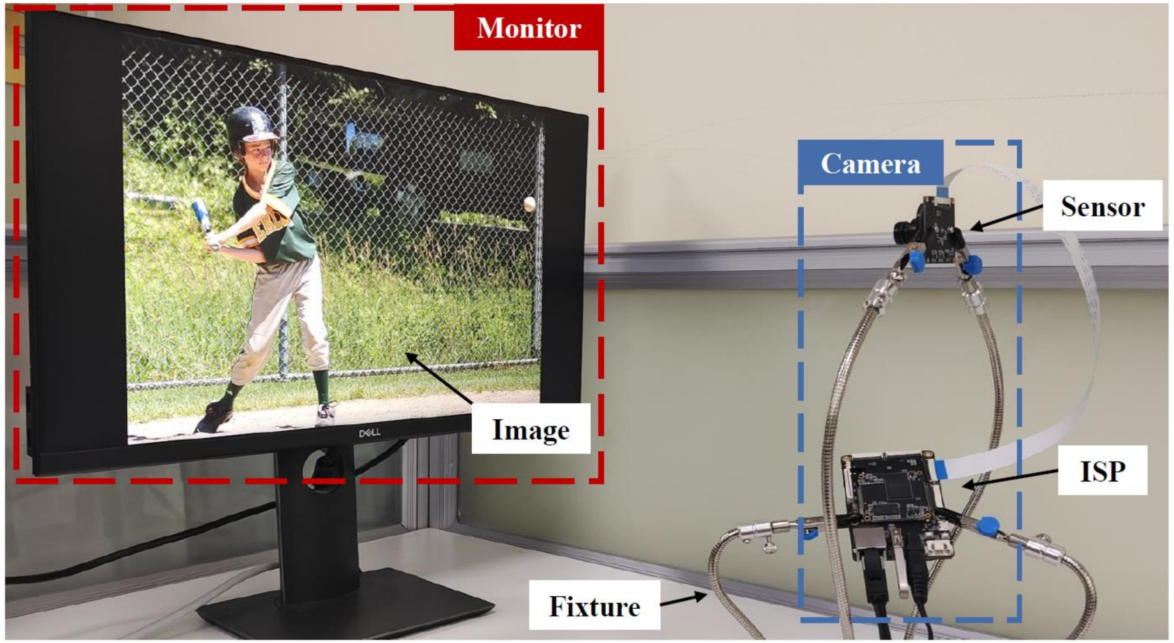
E3: Real-world Evaluation

- A prototype camera module (Sensor: IMX415 + ISP: RV1126)



E3: Real-world Evaluation

- □ A prototype camera module (Sensor: IMX415 + ISP: RV1126)
- □ Real-world captured images are close to simulation results.



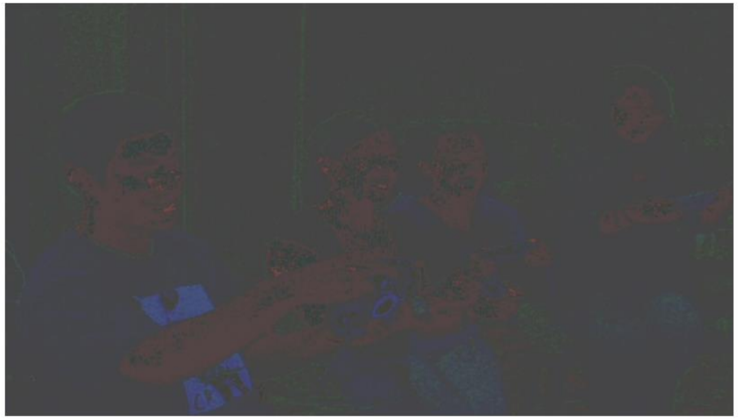
Simulated image

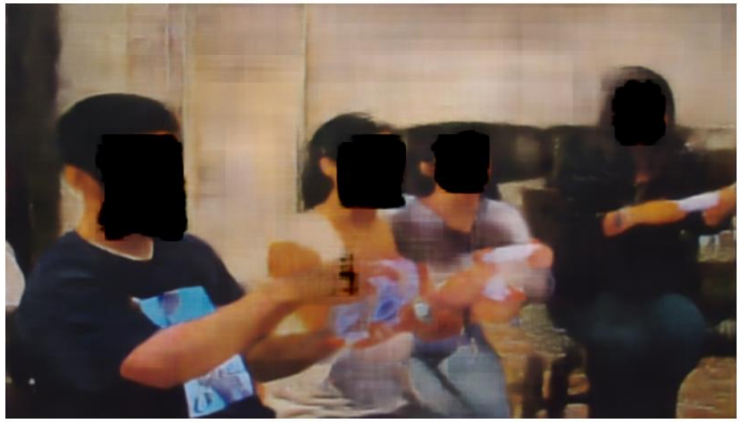
Real captured image

30dB PSNR

E3: Real-world Evaluation

- □ Due to **shooting noises**, real-world results are **better on privacy** and **worse on utility** than simulation ones.
  - ➢ Accuracy on LFW: 95.9% (Raw) → 0.13% (Cap.) / 0.28% (Enh.)
  - ➢ AP of person detection = 0.648
  - ➢ RMSE = 0.129; PSNR = 17.9; SSIM = 0.622



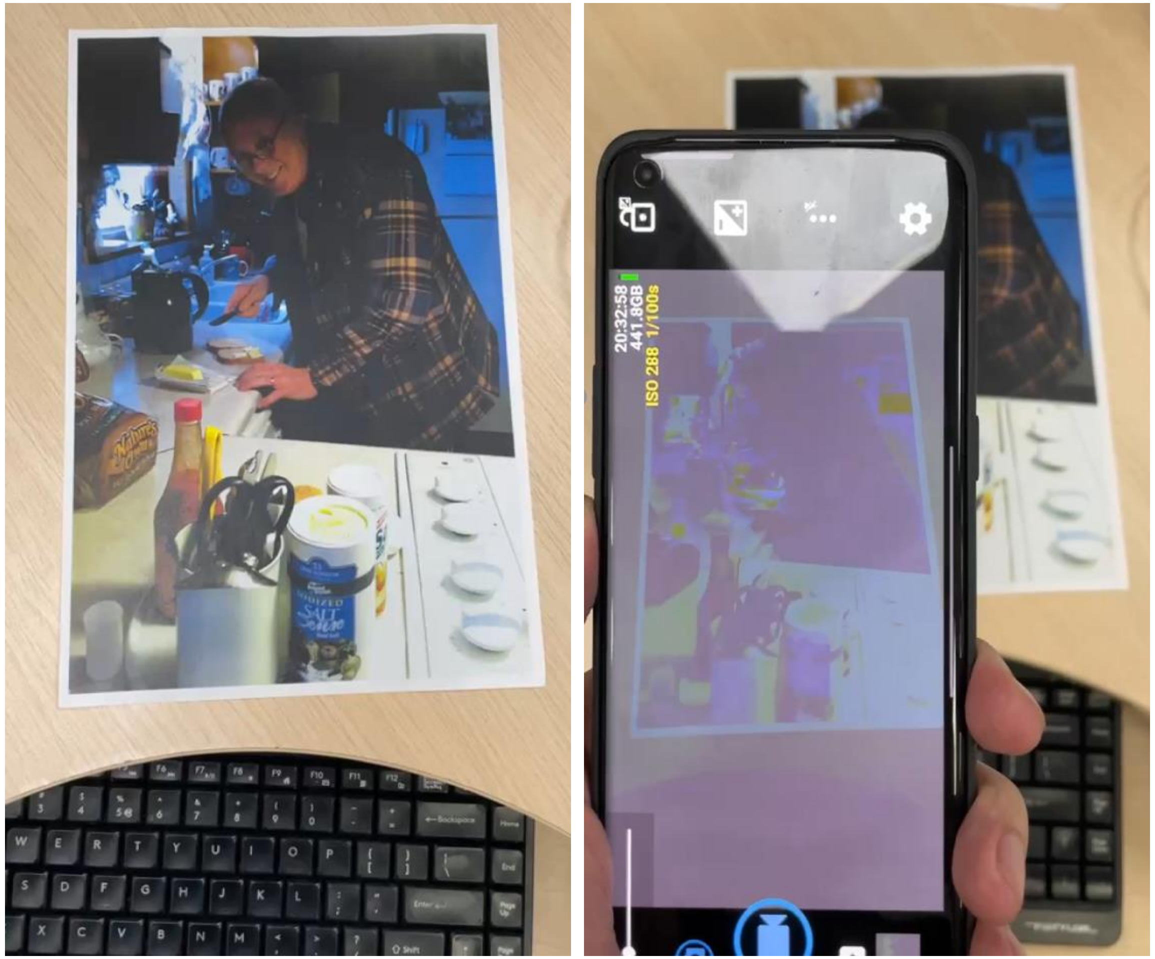


E3: Deployment on Android

- Android camera subsystem parameters
  - ColorSpaceTransform
  - TonemapCurve

Tested Android Smartphones

Device Model	OS	Android version
Google Pixel	stock Android	10
Samsung S20 FE	One UI 3.1	11
Huawei Nova 4	EMUI 10.0.0	10
OPPO Find X5 Pro	ColorOS 13.1	13
iQOO Neo5 SE	OriginOS 3	13
iQOO Neo6 SE	OriginOS 3	13
Redmi K30S Ultra	MIUI 14.0.5	12
MEIZU 16th Plus	Flyme 8.1.8.0A	8



Conclusion

- Propose a new paradigm, privacy-preserving by birth
- Optimize ISP parameters to achieve anti-facial recognition
- Generalized to various facial recognition algorithms and even resistant to white-box adaptive attacks

CamPro: Camera-based Anti-Facial Recognition

Wenjun Zhu, Yuan Sun, Jiani Liu, Yushi Cheng, Xiaoyu Ji, Wenyuan Xu

USSLAB, Zhejiang University

USSLAB Website:
www.usslab.org

Contact Authors:
zwj_@zju.edu.cn
xji@zju.edu.cn
wyxu@zju.edu.cn

Artifact Evaluated
NDSS SYMPOSIUM
Available
Functional
Reproduced

Evaluated Artifact:
zenodo.org/records/10156141

Code Release:
github.com/forget2save/CamPro