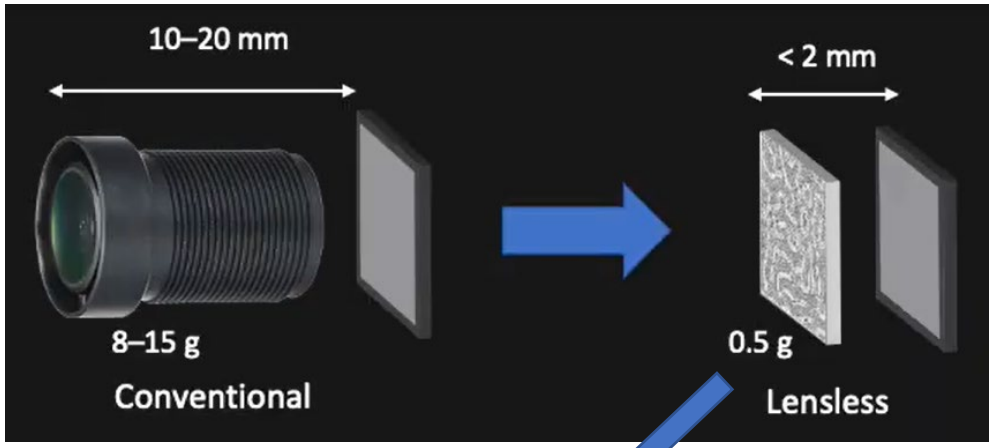
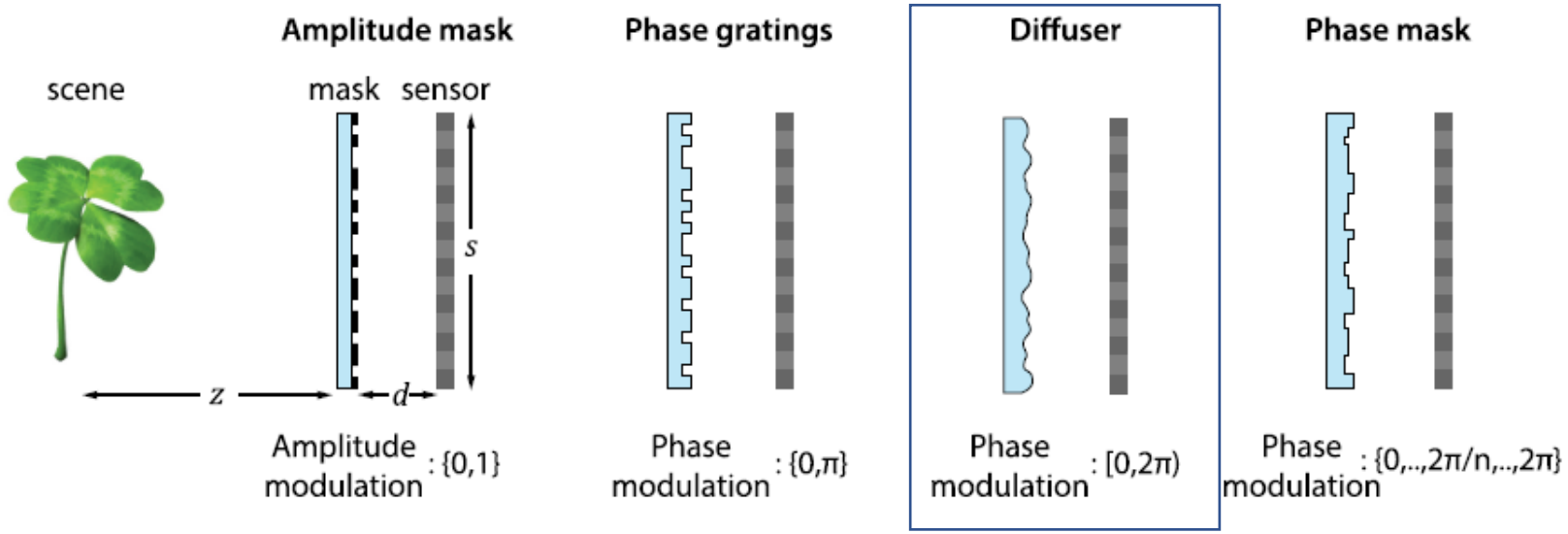
Mask-based Lensless Camera

Zhajiang university, Zhang yinger

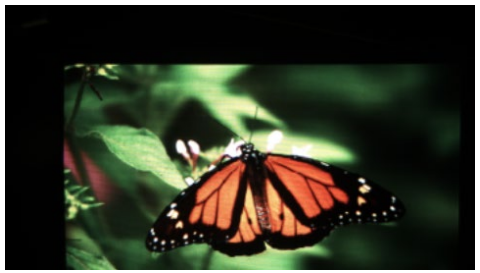
Introduction of Lensless camera

- Size
- Weight
- Cost
- Visual privacy

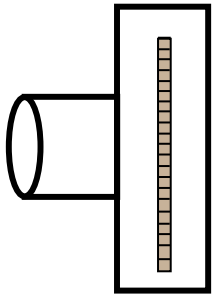




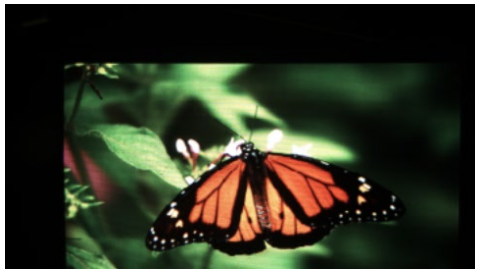
Challenge in Lensless Camera



Scene

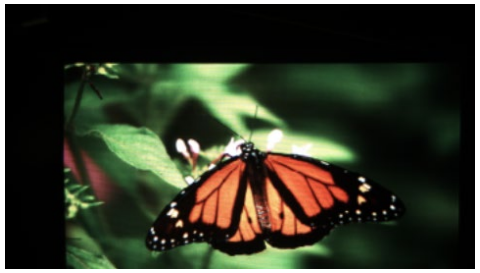


Lensed camera

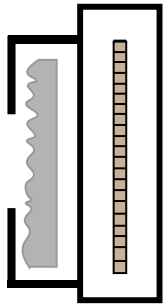


Measurement

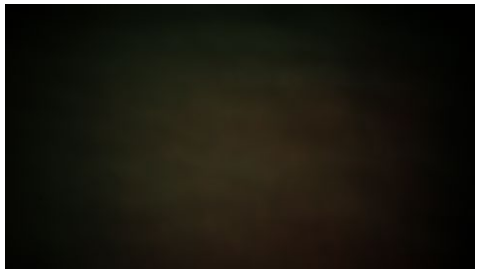
- Lens focuses scene onto sensor
- Mesurement resembles scene



Scene

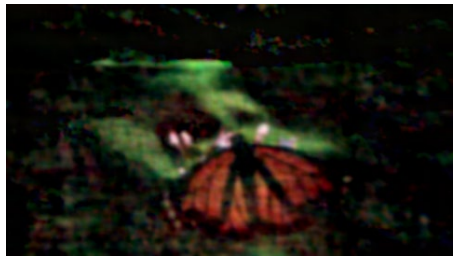


Lensless camera

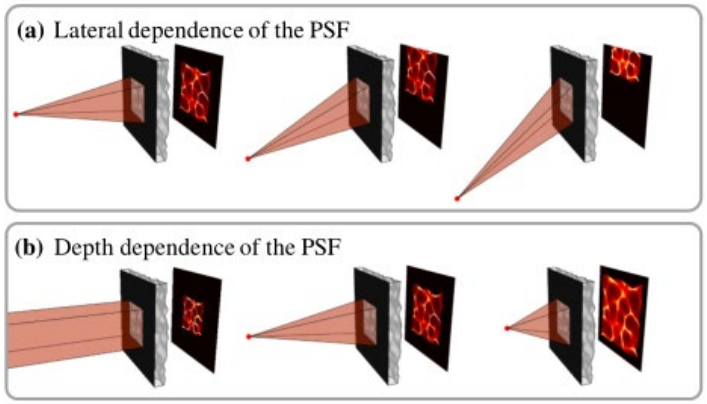


Measurement

- Mesurement is highly multiplexed
- Does not resemble scene
- Needs reconstruction algorithms



Physical Model



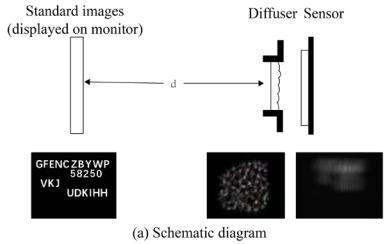
Forward Model $b = Hv \longrightarrow \mathbf{b}(x,y) = crop[\mathbf{h}(x,y) * \mathbf{x}(x,y)]$

Convolution Approximation:
Simpliy the calibration

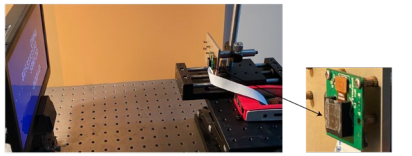
Inverse problem $\hat{v} = arg \min_{v \geq 0} \frac{1}{2} \left|\left|b - Hv\right|\right|_2^2 + \tau ||\Psi v||_1$

Our Study

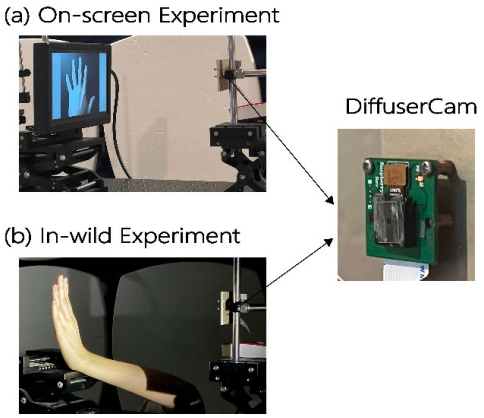
- Work1: Text Detection and Recognition (Reconstruction)



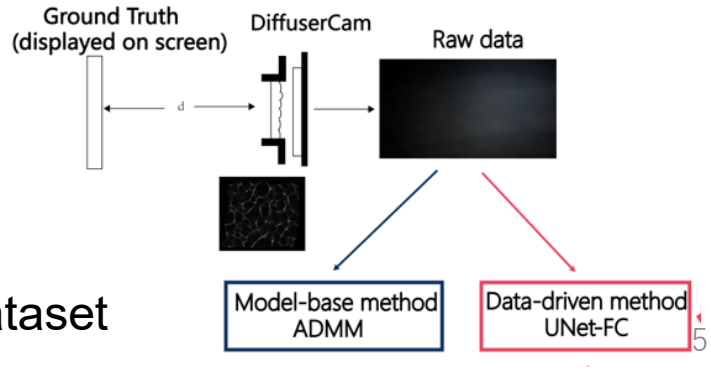
(a) Schematic diagram


(b) Lensless imaging system setup (c) Diffuser and sensor

- Work2: Hand Gestures Recognition in Videos (Reconstruction-free)

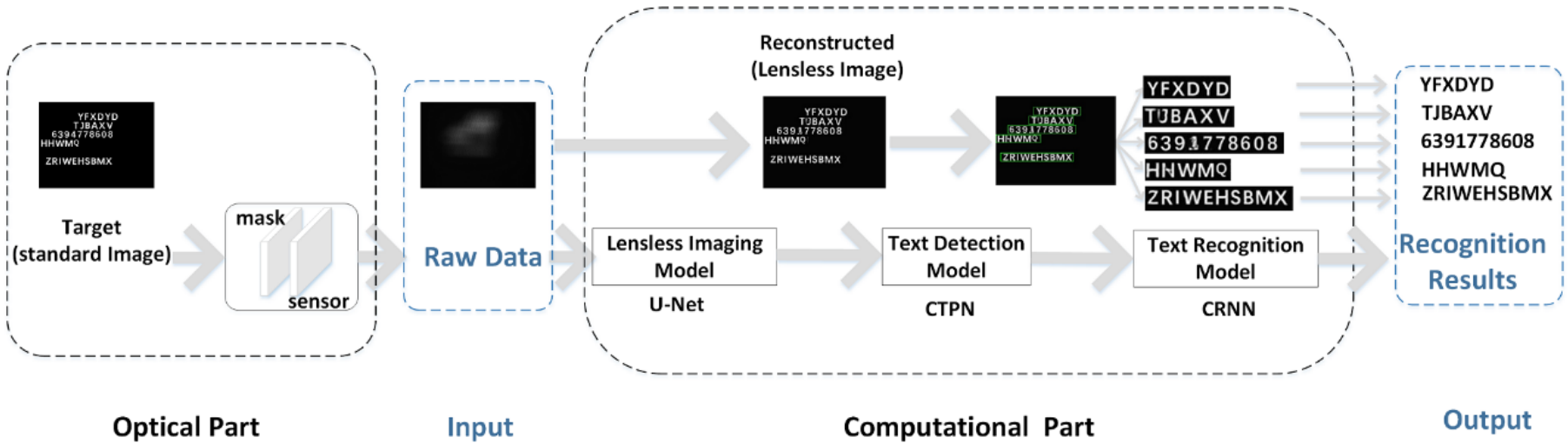


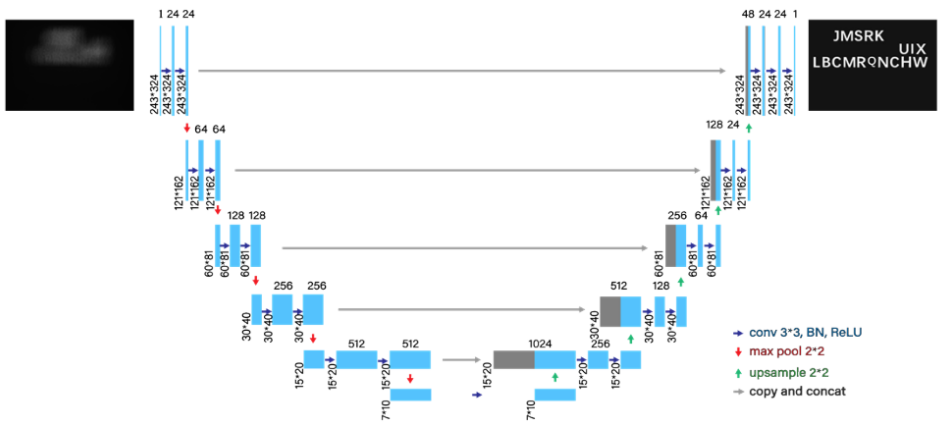
- Work3: Lensless imaging with two-branch fusion model (Reconstruction)



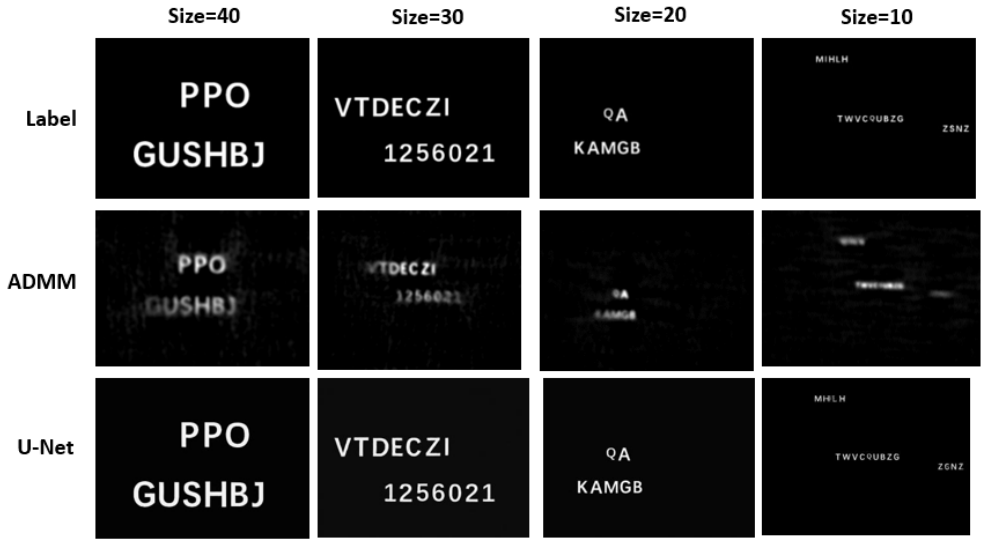
DiffuserCam Dataset

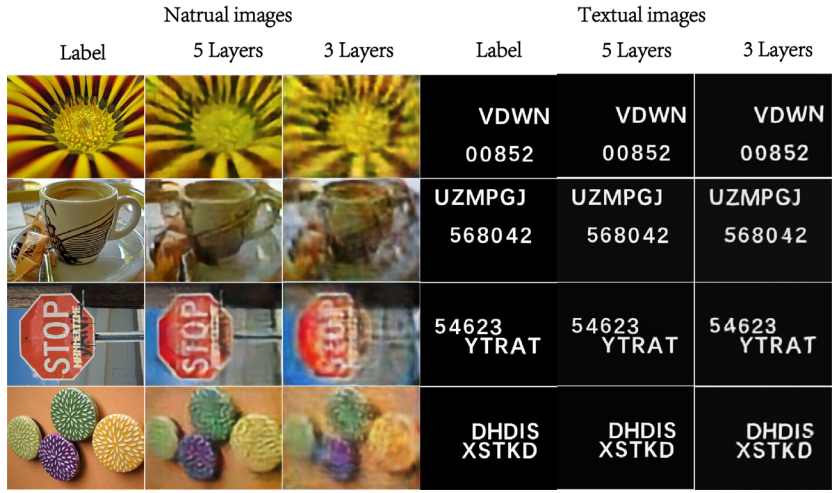
Work1: Framework

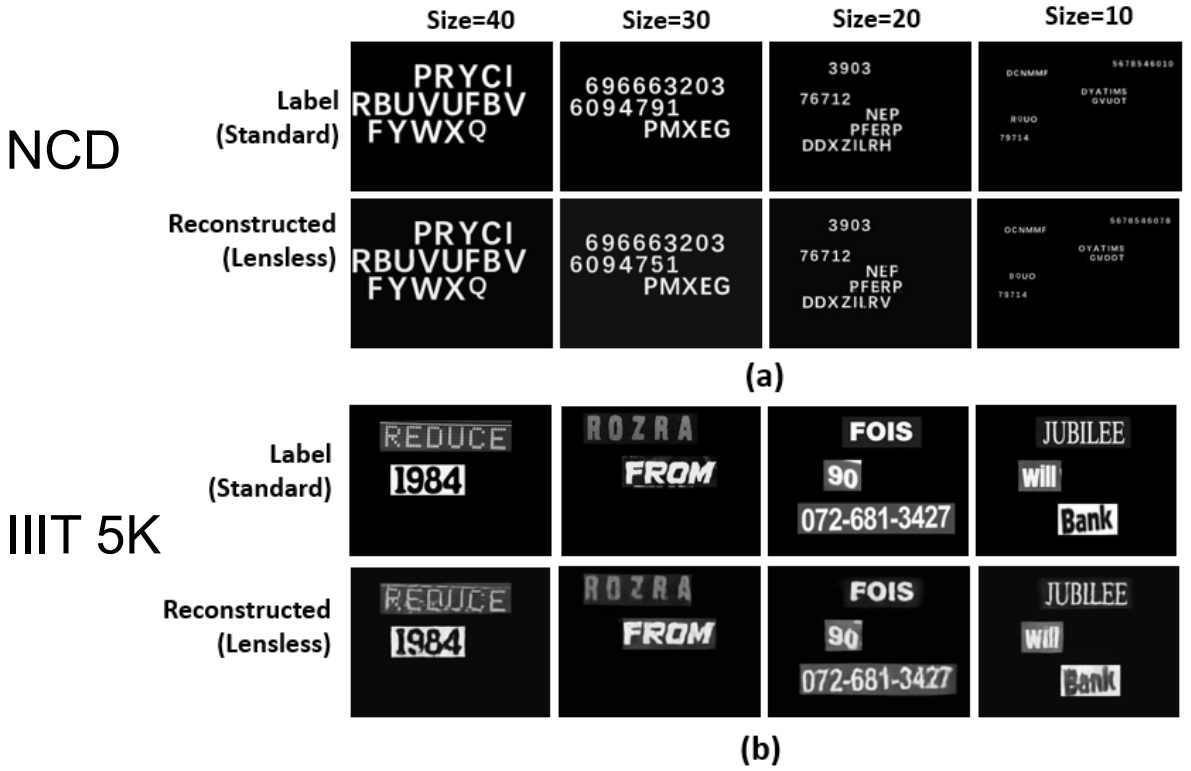




Work1: Reconstruction Quality



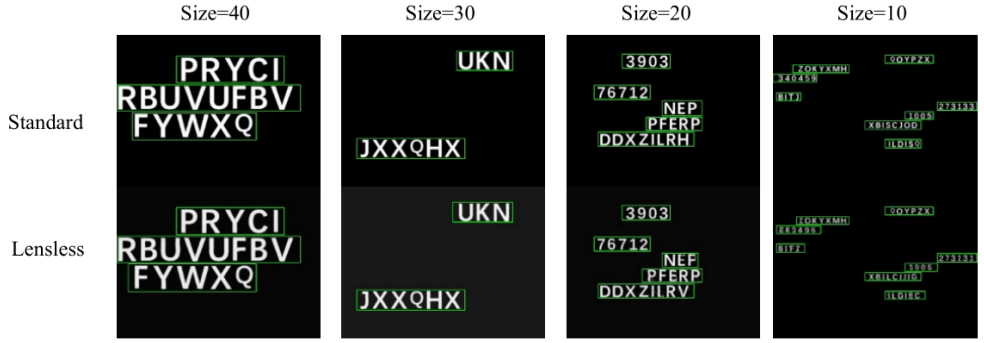




- Break through the limitations of resolution
- Supplemented by the judgment of the category
- Applicable to category type detection

Work1: Text Detection

NCD



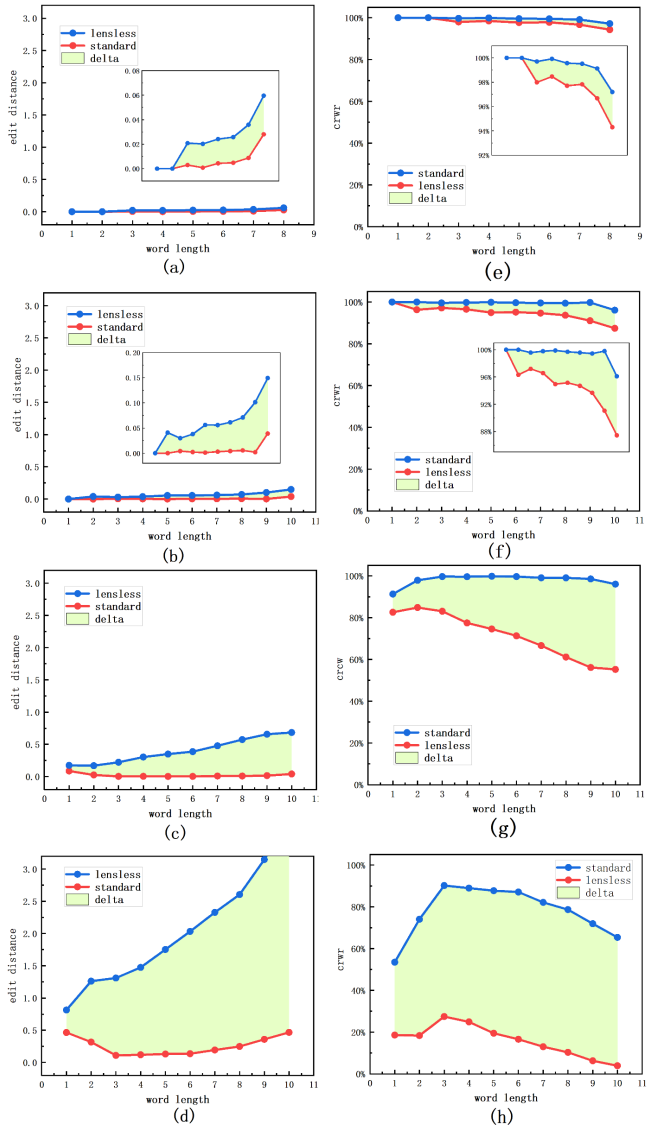
	Precision	Recall	F-score
Size40	1.0000	1.0000	1.0000
Size30	1.0000	1.0000	1.0000
Size20	1.0000	1.0000	1.0000
Size10	0.9991	1.0000	0.9996

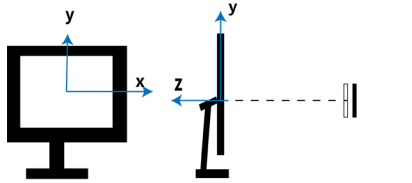
IIIT 5K

Dataset		Total Num	Precision	Recall	F-score
simple	Standard	715	0.8737	0.8769	0.8753
	Lensless	715	0.8574	0.8283	0.8426
complex	Standard	976	0.8599	0.8640	0.8620
	Lensless	976	0.8337	0.7704	0.8008

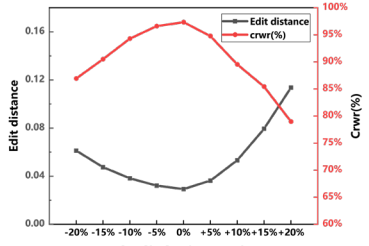
Work1: Text Recognition

NCD

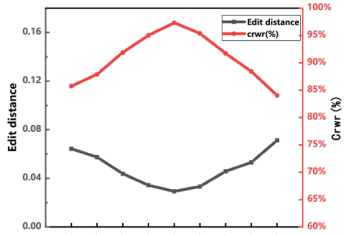




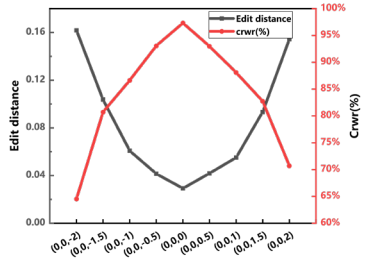
(a) coordinate system



(b) light intensity



(c) target in (x,y,0)



(d) target in (0,0,z)

Factors :

- Word length
- Character size
- Light intensity
- Position
- Background complexity

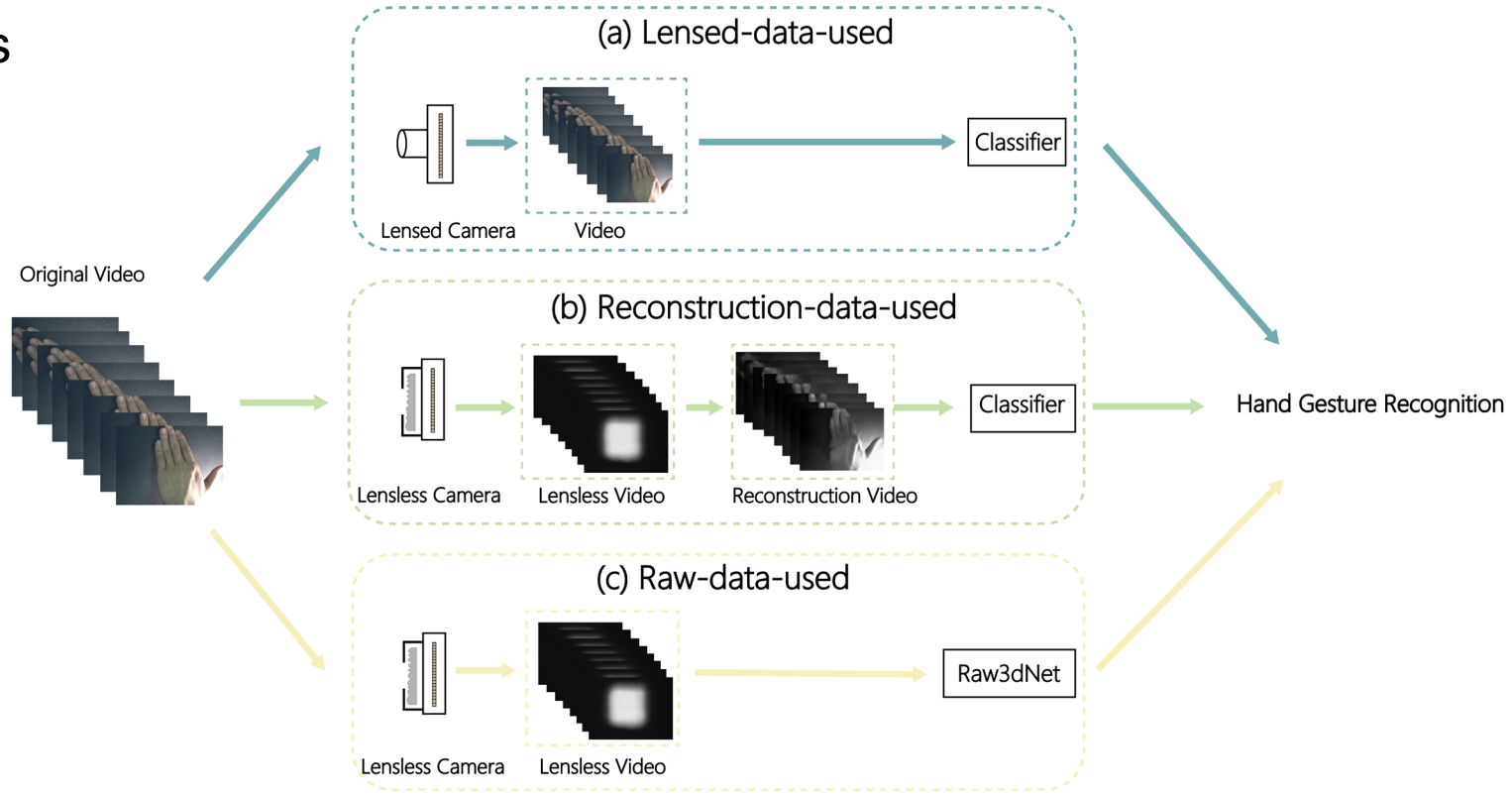
IIIT 5K

Mode		Total	Edit Distance	Crwr
Simple	Standard	1272	255	88.80%
	Lensless	1272	704	71.78%
Complex	Standard	1857	705	78.72%
	Lensless	1857	2315	51.23%

Work2: Framework

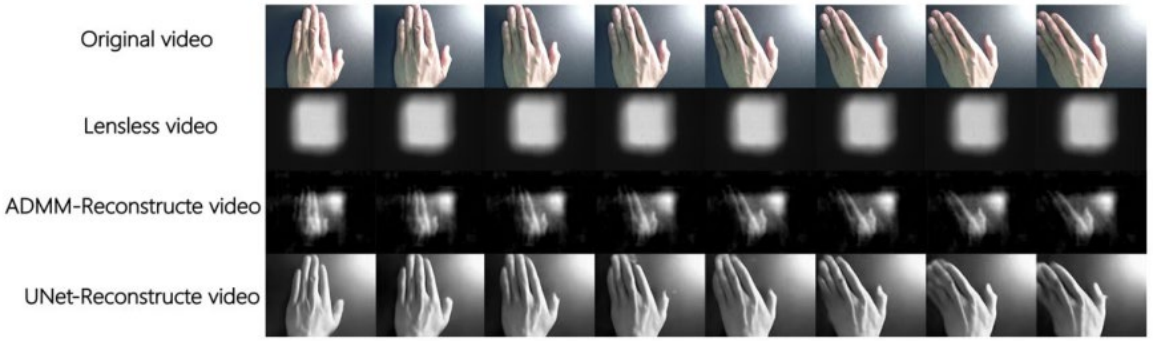
Advantage :

- Reduce computational burdens
- Protect privacy
- Sample data, small data traffic

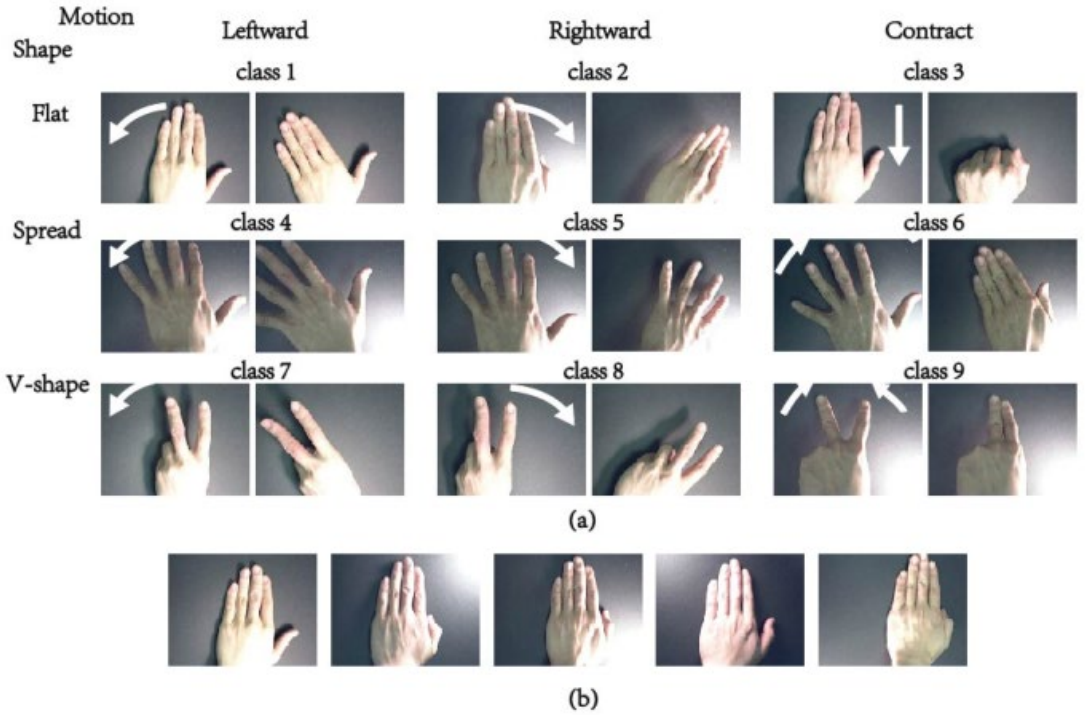


Work2: Dataset

Definition of dataset

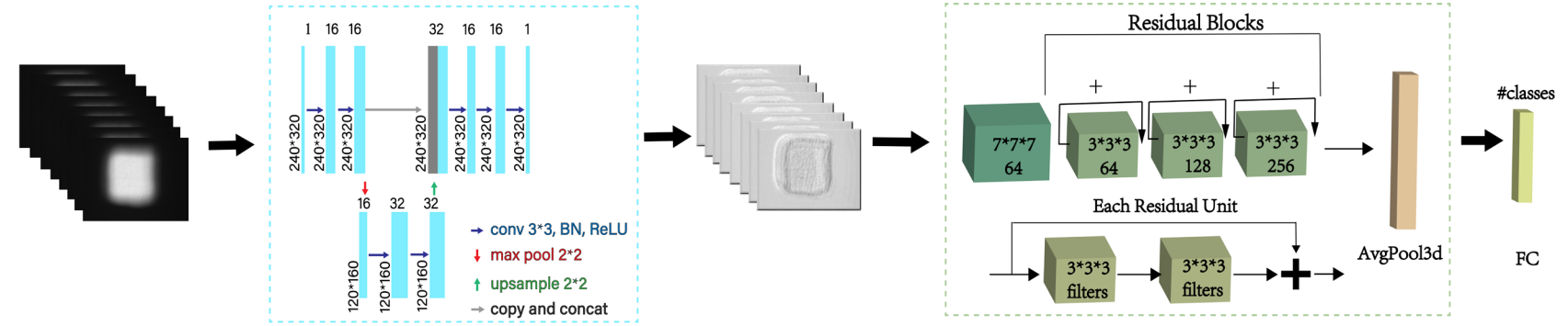


Cambridge Hand Gesture



Train set: 2832 video
Test set: 780 video

Work2: Method



(a) Measurement Frames (b) Spatial Feature Extractor (c) Feature maps (d) 3D-ResNet

Index	Dataset	Model	Accuracy on Test Dataset
Exp1	Original video	3d-ResNet	99.36%
Exp2	ADMM-Reconstructed video	3d-ResNet	93.33%
Exp3	UNet-Reconstructed video	3d-ResNet	95.64%
Exp4	Lensless video	3d-ResNet	78.97%
Exp5	Lensless video	Raw3dNet	98.59%

Work2: Method

Why SFE?

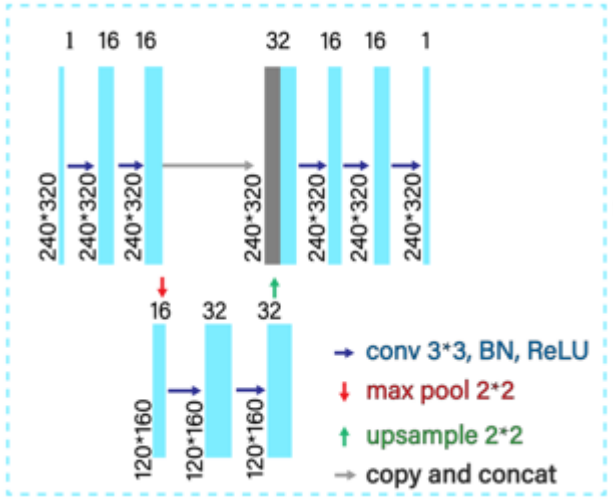


Table 1. Confuse matrix when using 3D-ResNet for lensless video classification.

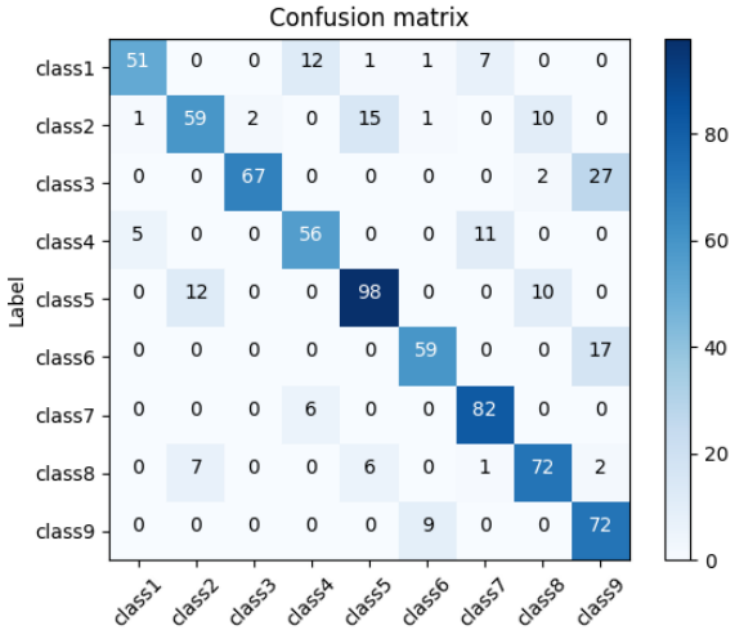


Table 2. The distribution of the most pertinent category for class 1. Row1 represents images of raw data, and Row2 represents feature maps produced by SFE.

Dataset	Class 1	Class 4	Class 7
Raw data	25	37	10
Feature map	60	6	6

Work2: Result

Table 3. Comparison of performances for 3D-ResNet/ Raw3dNet for lensless video; comparison for lensless video/reconstruction video/lensed video.

Index	Dataset	Model	Accuracy on Test Dataset
Exp1	Original video	3d-ResNet	99.36%
Exp2	ADMM-Reconstructed video	3d-ResNet	93.33%
Exp3	UNet-Reconstructed video	3d-ResNet	95.64%
Exp4	Lensless video	3d-ResNet	78.97%
Exp5	Lensless video	Raw3dNet	98.59%

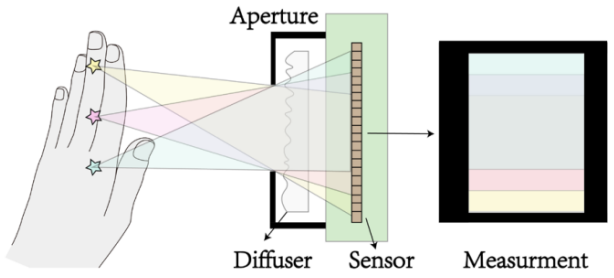
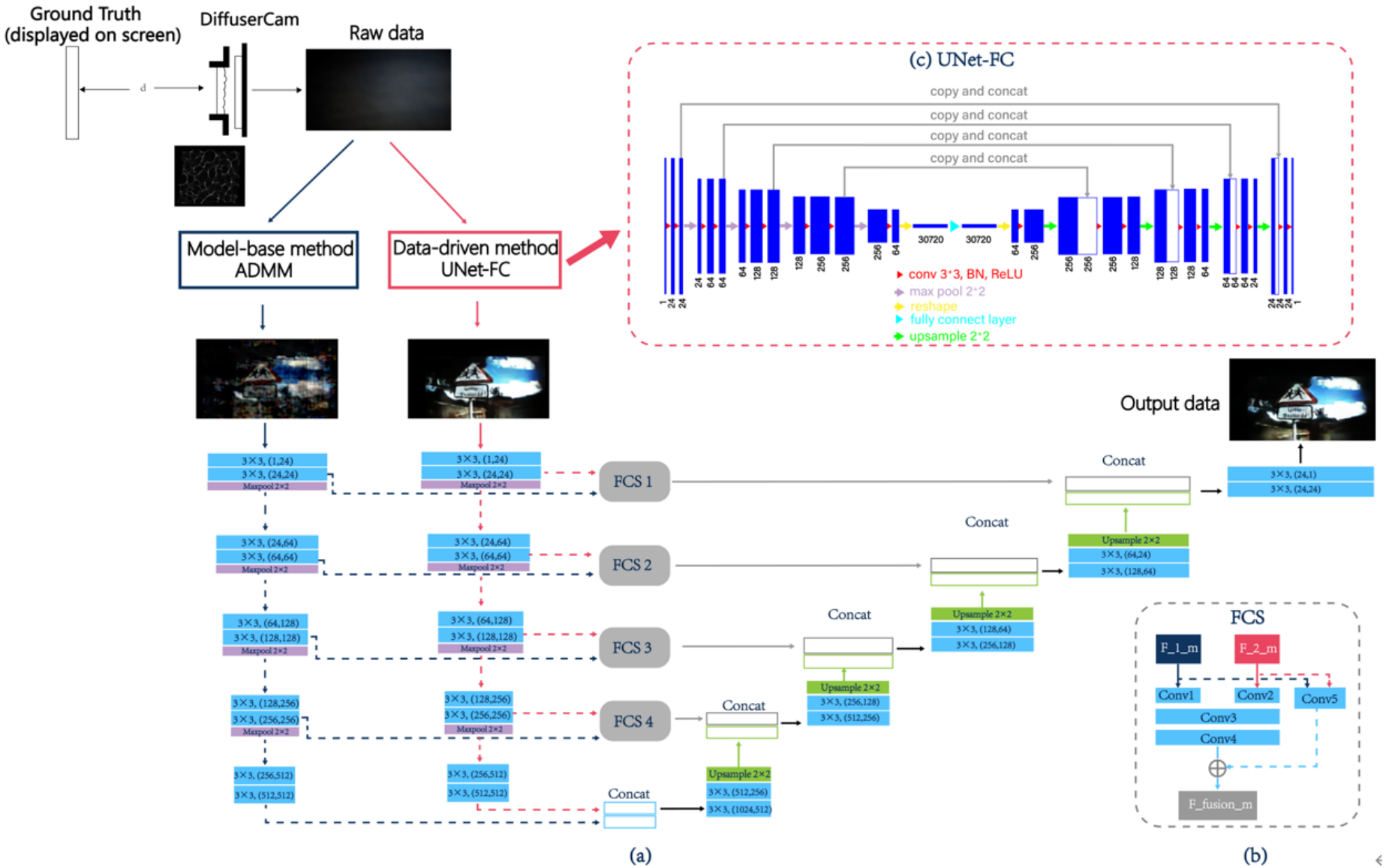


Table 4. Assessment for various down-sampling techniques and ratios.

Pixel Size	Compress Method	Accuracy on Test Dataset
(320,240)	None	98.59%
(100,75)	Resize	98.46%
(100,75)	Uniform sample	96.92%
(100,75)	Random sample	79.74%
(200,150)	Erase (25% reserved)	91.54%
(50,37)	Resize	90.13%

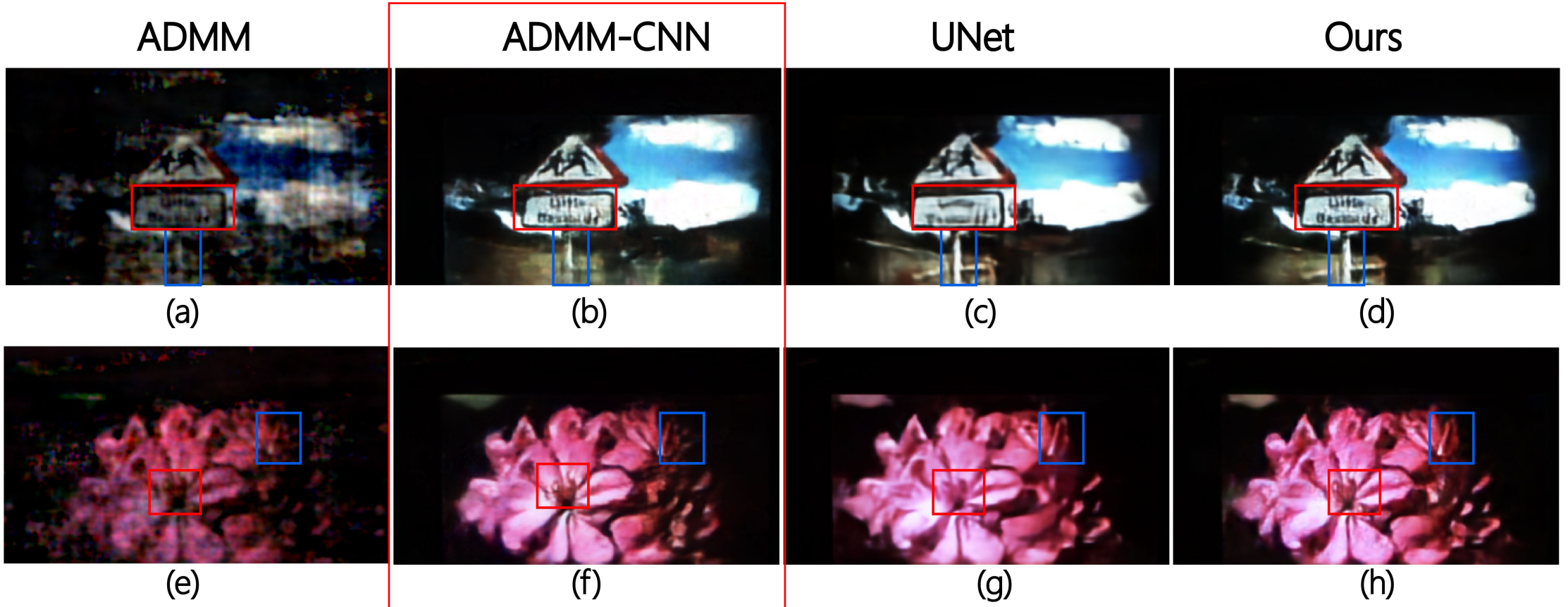
- Reconstruction-free method achives acc comparable to that of a lensed camera
- Reconstruction-free method outperforms reconstruction method
- Hand gesture recognition is possible with a small amount of raw data

Work3: Framework



Work3： Why fusion?

State-of-the-art

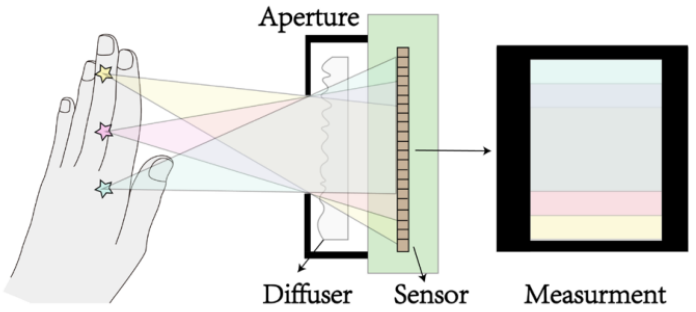


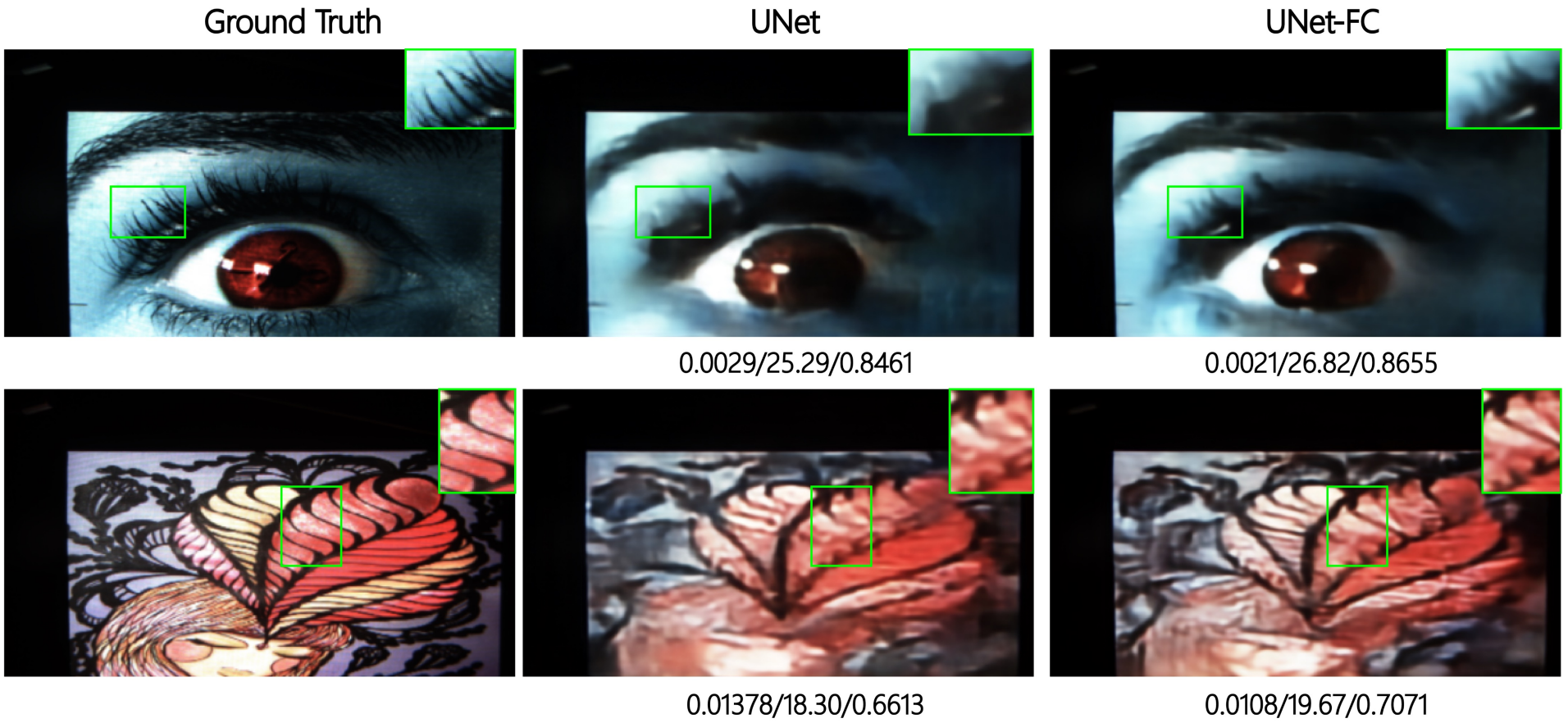
Higher resolution
Less details in edges

More details in edges
Lower resolution

Work3：UNet-FC

Adapt to Multiplexing property





Work3：Result

Table 1. Average MSE, PSNR and SSIM metrics for each method on the test dataset.

Reconstruction	MSE	PSNR	SSIM
Le-ADMM	0.0312	12.89	0.6102
Le-ADMM-U	0.0065	22.88	0.8354
UNet	0.0081	20.20	0.7791
Ours	**0.0035**	**25.61**	**0.8665**