



Association for  
Computing Machinery

**SenSys** 2024



SIMON FRASER  
UNIVERSITY



Microsoft

# M<sup>3</sup>Cam: Lightweight Super-Resolution via Multi-Modal Optical Flow for Mobile Cameras

Yu Lu<sup>†&</sup>, Dian Ding<sup>†&</sup>, Hao Pan<sup>‡\*</sup>, Yongjian Fu<sup>◆</sup>, Liyun Zhang<sup>†</sup>, Feitong Tan<sup>£</sup>, Ran Wang<sup>†</sup>,  
Yi-Chao Chen<sup>†</sup>, Guangtao Xue<sup>†</sup>, Ju Ren<sup>€</sup>

Shanghai Jiao Tong University<sup>†</sup> Microsoft Research Asia<sup>‡</sup>

Central South University<sup>◆</sup> Simon Fraser University<sup>£</sup> Tsinghua University<sup>€</sup>

# High-resolution photos improve the user's photographic experience



Mobile Photography



Low Resolution Photo

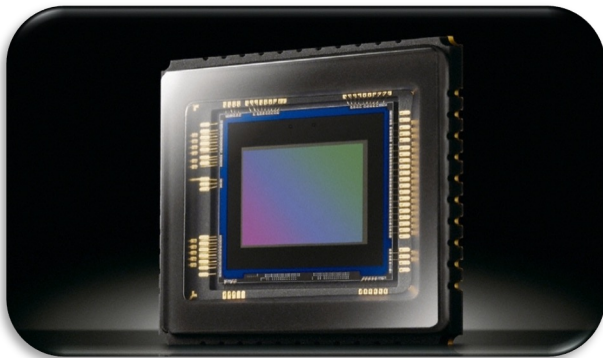


High Resolution Photo

# Smartphone' Camera Boost Resolution through Hardware Upgrades



**The Moto X30 Pro is the first 200MP phone launched on 8/11/2022, while most flagships still use 50MP sensors today.**



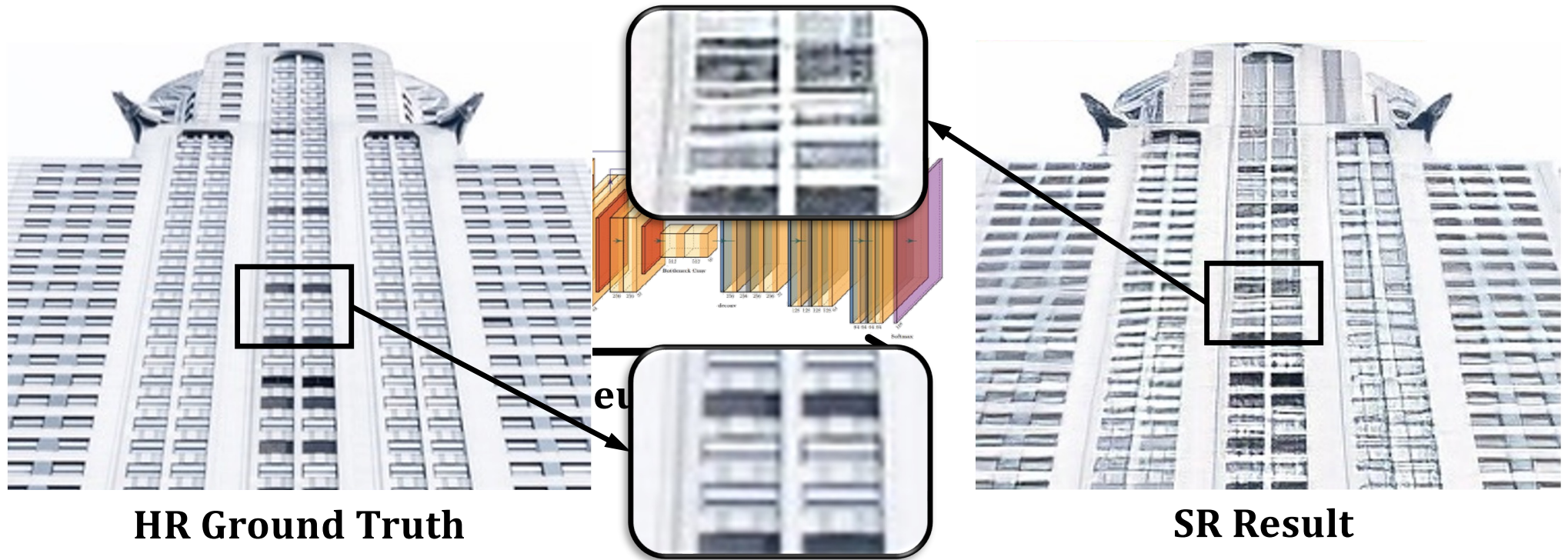
**CMOS sensors face limited pixel account due to **area constraints****



**Periscope lenses extend our view without increasing resolution**

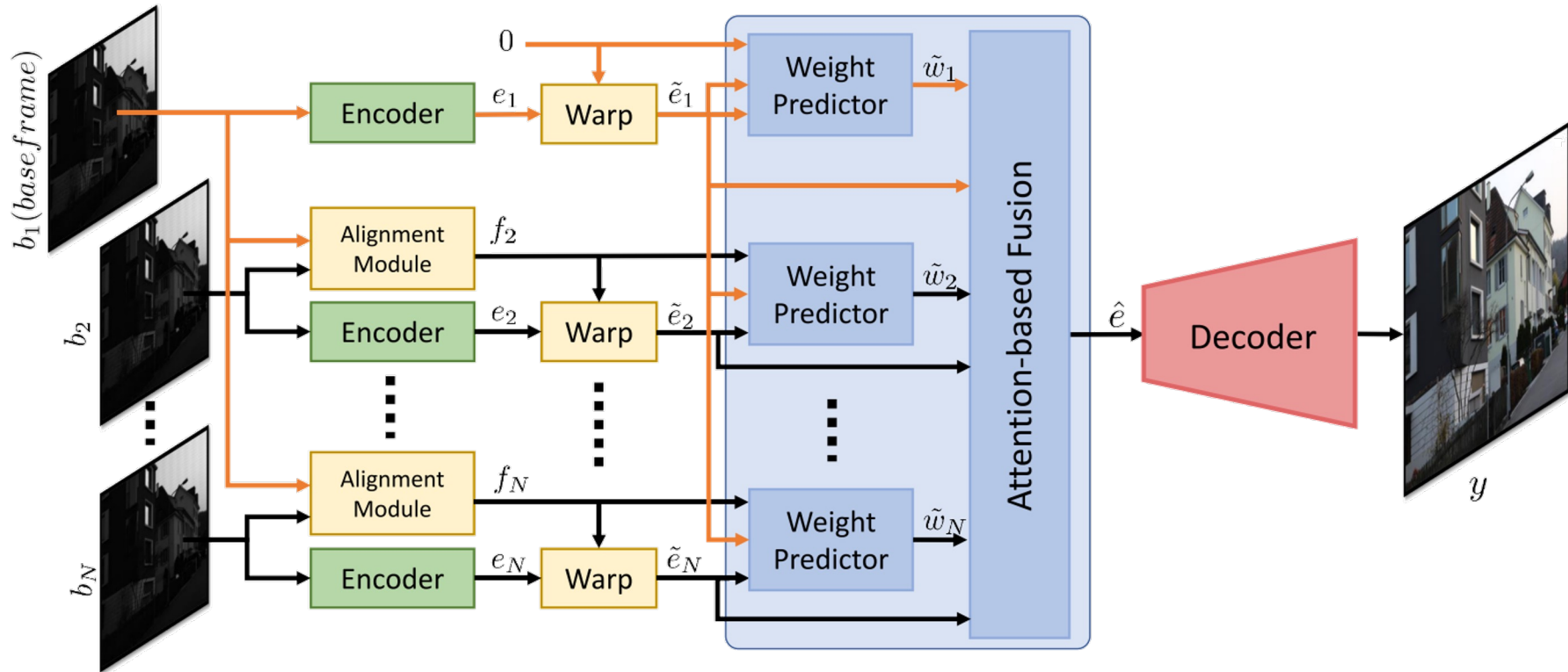


# Single-Frame Super-Resolution (SFSR) with Neural Network



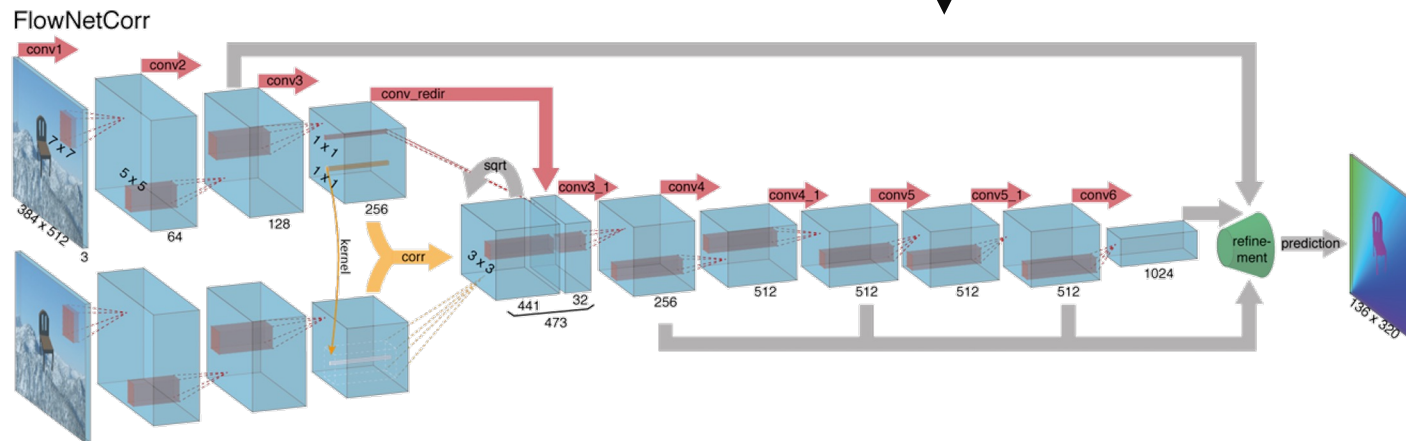
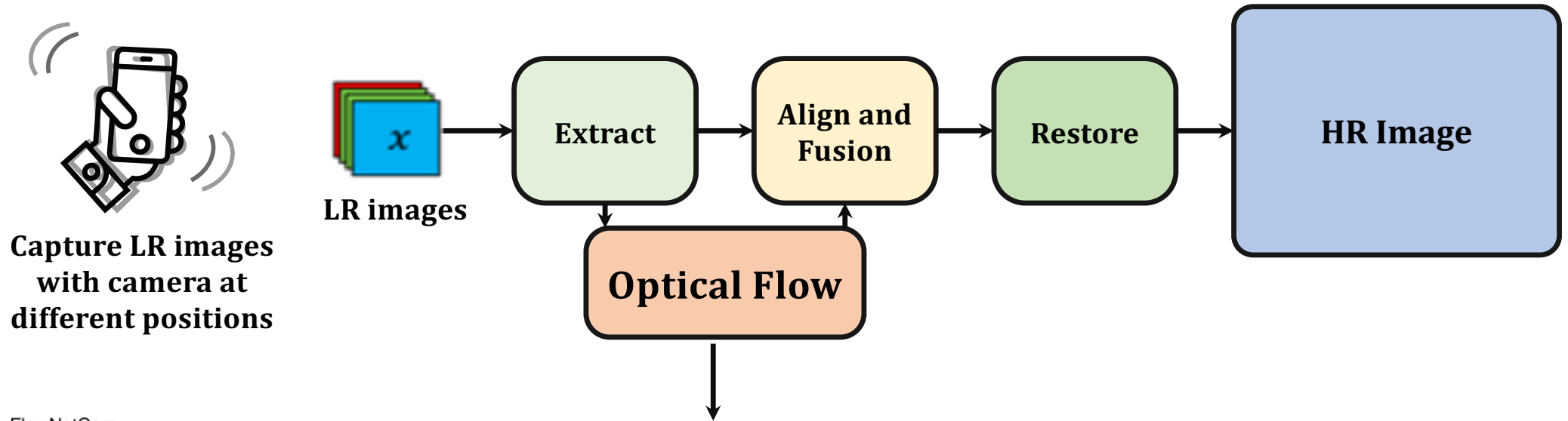
SFSR leverages learned experiences to enhance details in up-sampled images, but may encounter **artifacts** or **excessive smoothing** due to insufficient information.

# Multi-frame Super-resolution (MFSR)



MFSR yields better results than SFSR, because it can restore more information about the real scene through **multiple sampling**.

# MFSR on mobile devices and its challenges

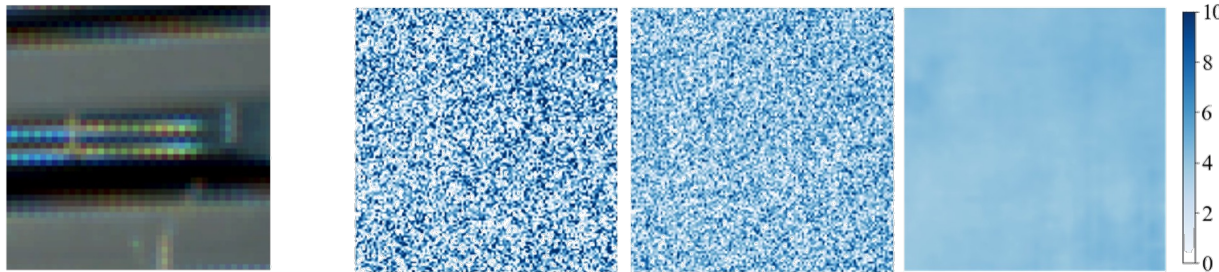


Optical flow-based image alignment is **challenging**.

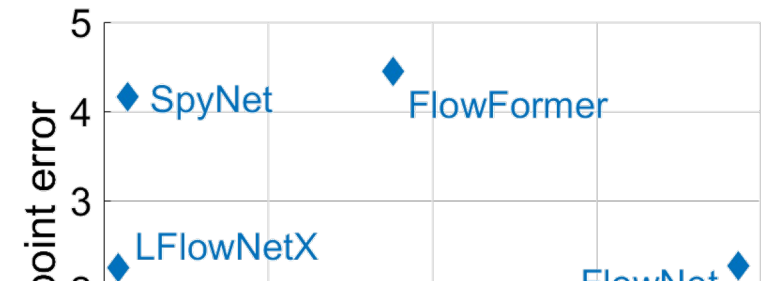
**N-fold Super-resolution** requires optical flow estimation with  $\frac{1}{\sqrt{N}}$  precision

# Low-accuracy optical flow impacts the quality of MFSR

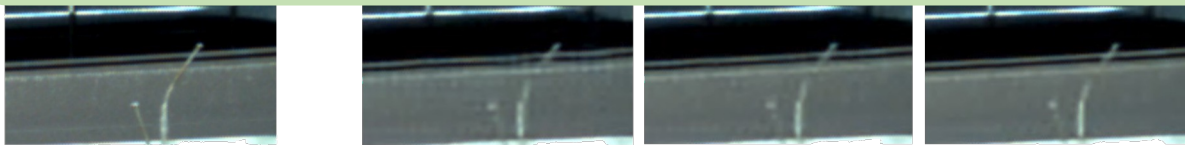
Optical Flow Maps with *Varying Error Levels*



Existing Optical Flow Models



## How to Design **Extremely Accurate** and **Lightweight** Optical Flow Modules for Mobile 16-fold Super-Resolution Imaging?



HR image

SR results generated using the above optical flow

Accuracy of optical flow **significantly affects** super-resolution imaging results

Parameter (Mb)

RAFT (SOTA) achieves **~0.5-pixel** error, supporting **4-fold** super-resolution

**16-fold** Super-Resolution on mobile needs **<0.25-pixel** error





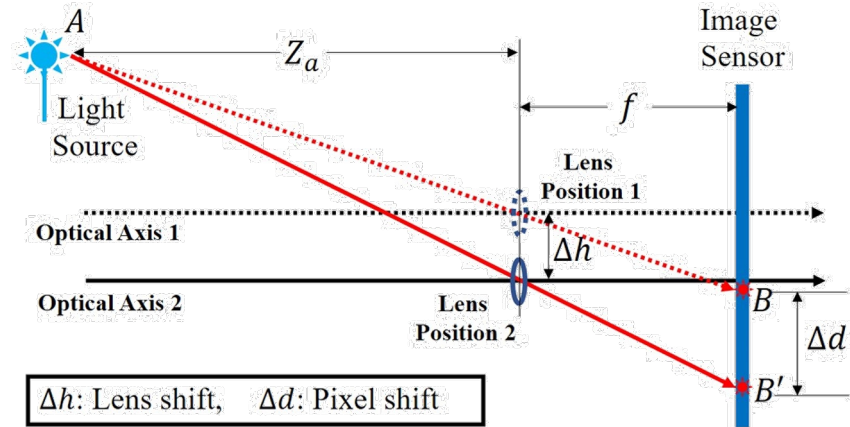
# Intuition: Enhance optical flow accuracy with an auxiliary modality



Frame 1

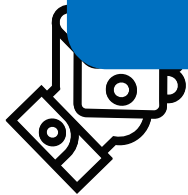


Frame 2



Lens motion yields optical flow results that negatively correlate with depth information :

However, steady and regular camera movement is nearly **impossible** for common users.



information

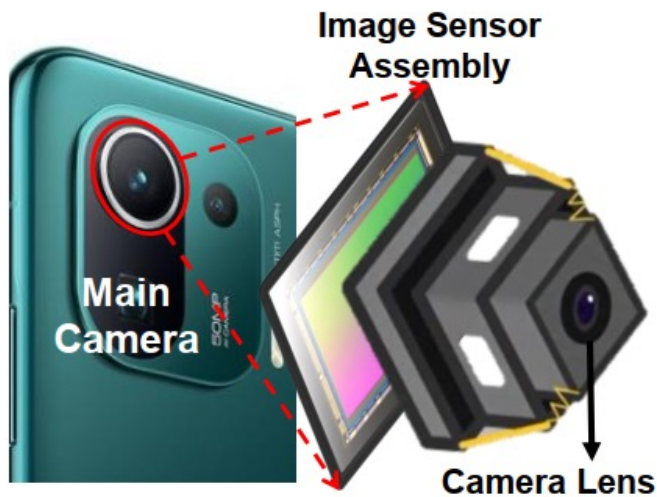
a lightweight NN model



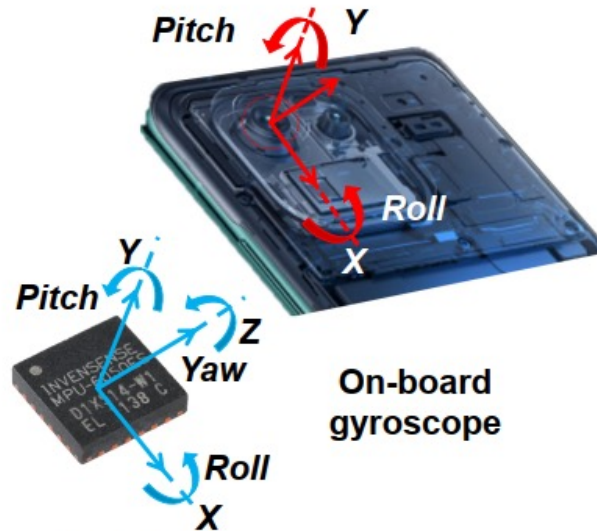




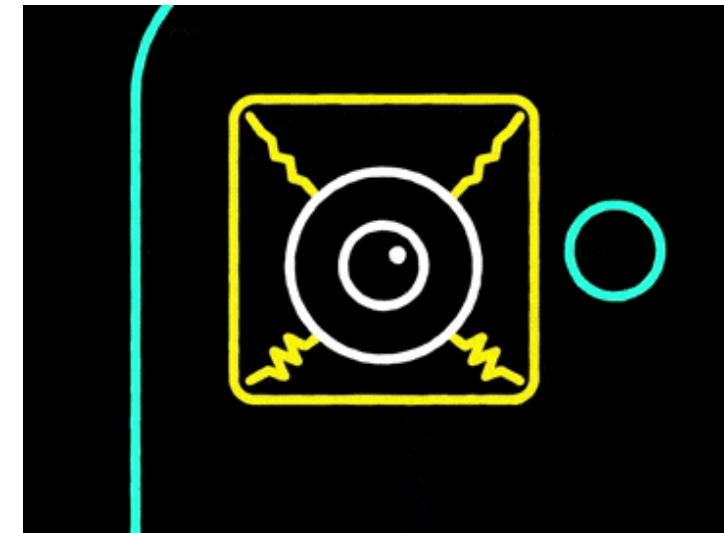
## OIS module can control lens motion with the phone stationary



An OIS-supported camera built in the smartphone



MEMS sensors sense camera movements

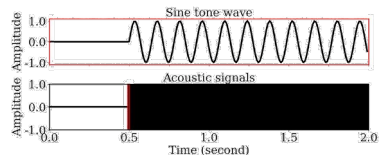


OIS actuator controls lens on the  $X/Y$ -plane

Controlling the OIS module enables steady and regular lens movement for handheld shooting by common users.



# Additional Modality : Controlled IMU readings for lens movement

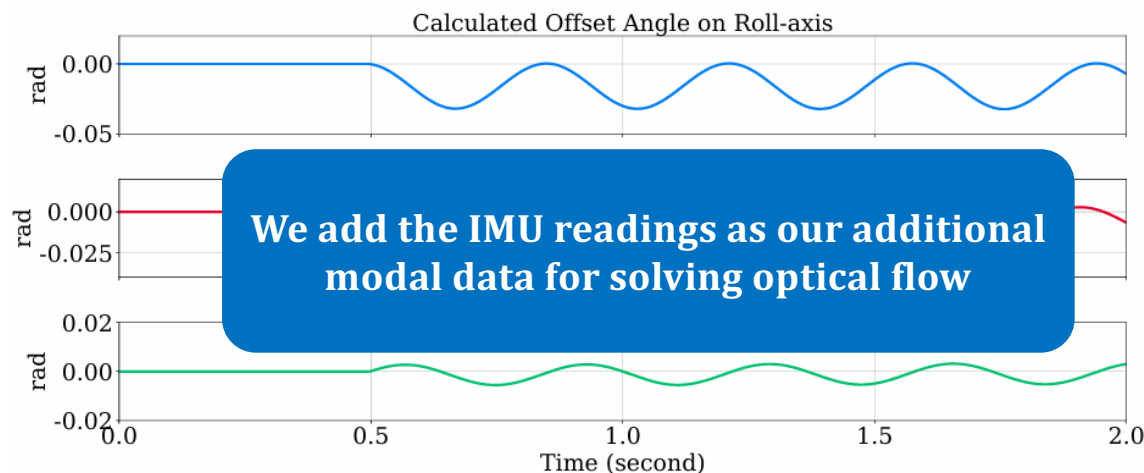


high-frequency  
cosine wave  
(e.g., 19KHz)

Acoustic injection

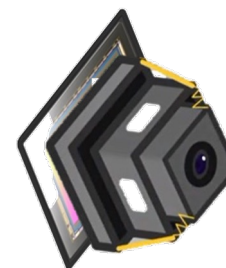


Built-in MEMS IMU

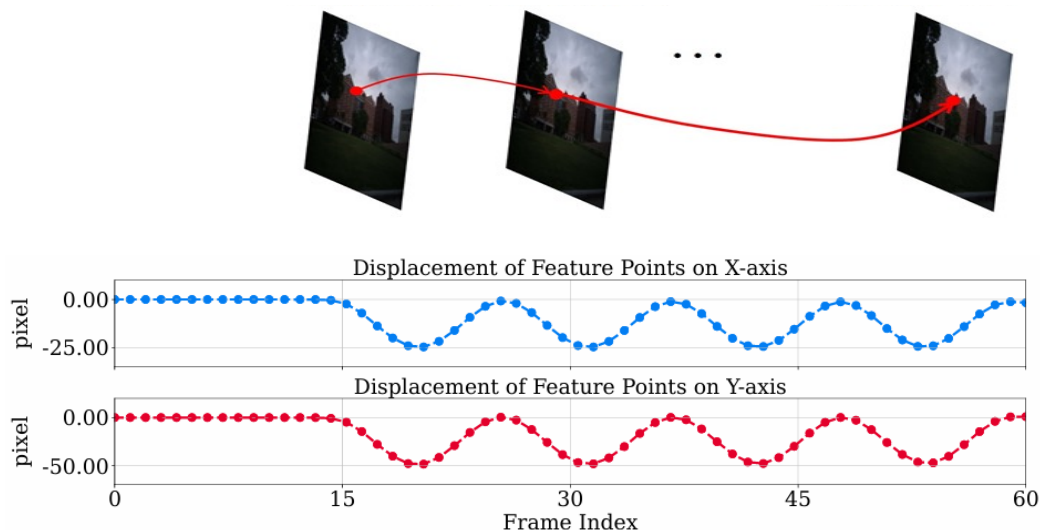


We add the IMU readings as our additional modal data for solving optical flow

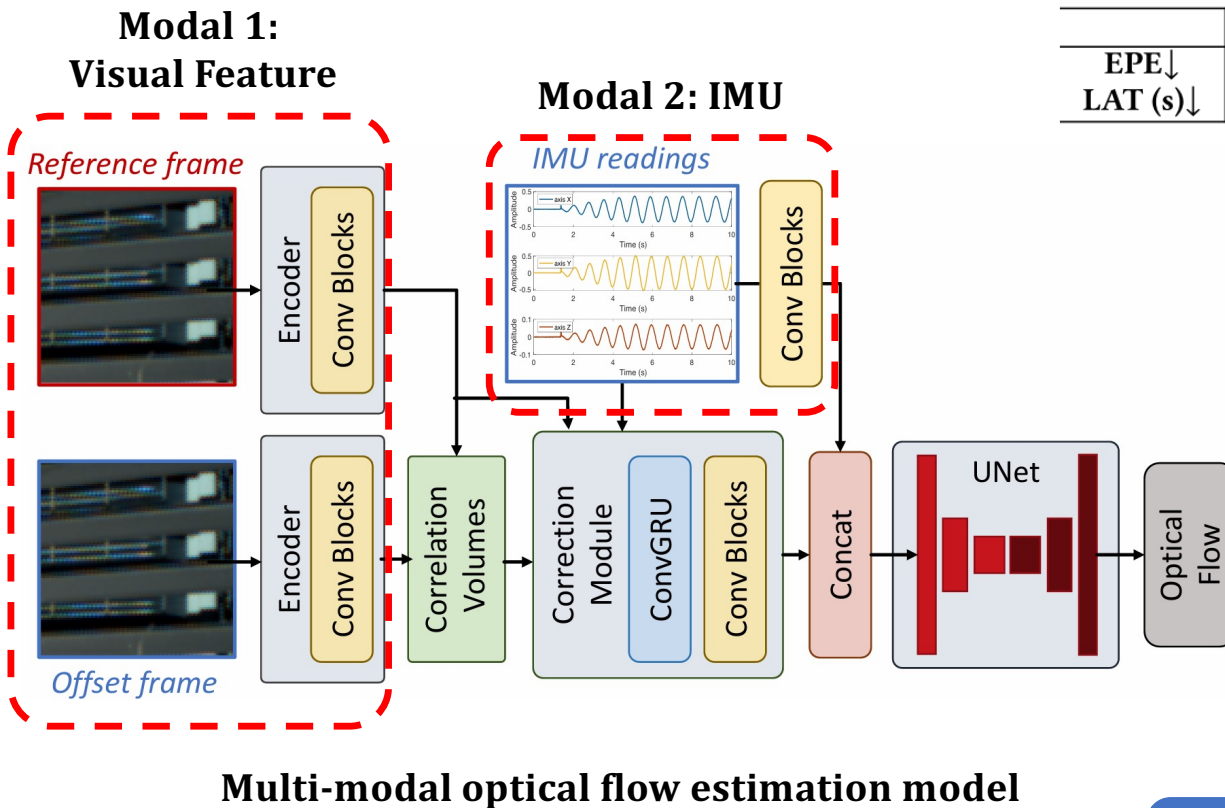
OIS module controls  
regular lens shift



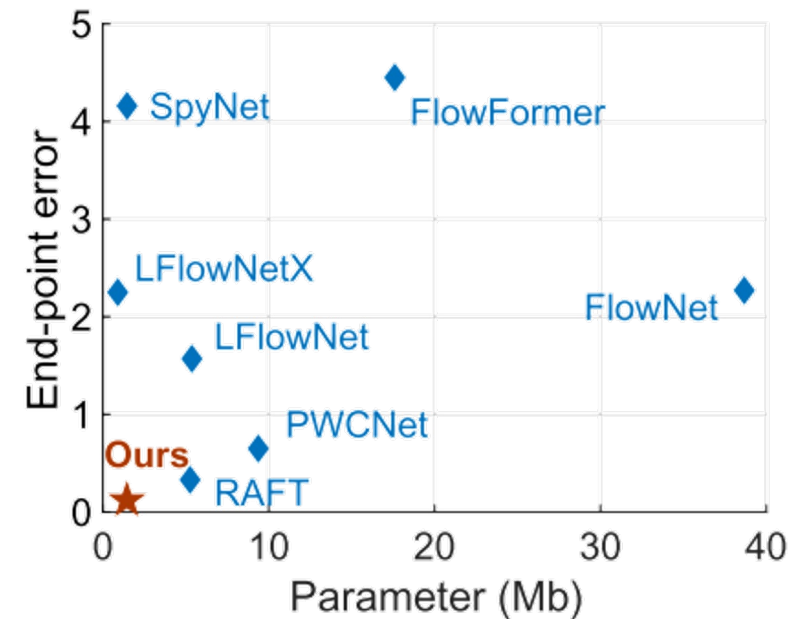
Regular  
optical flow  
pattern



# Our proposed multi-modal optical flow estimation module

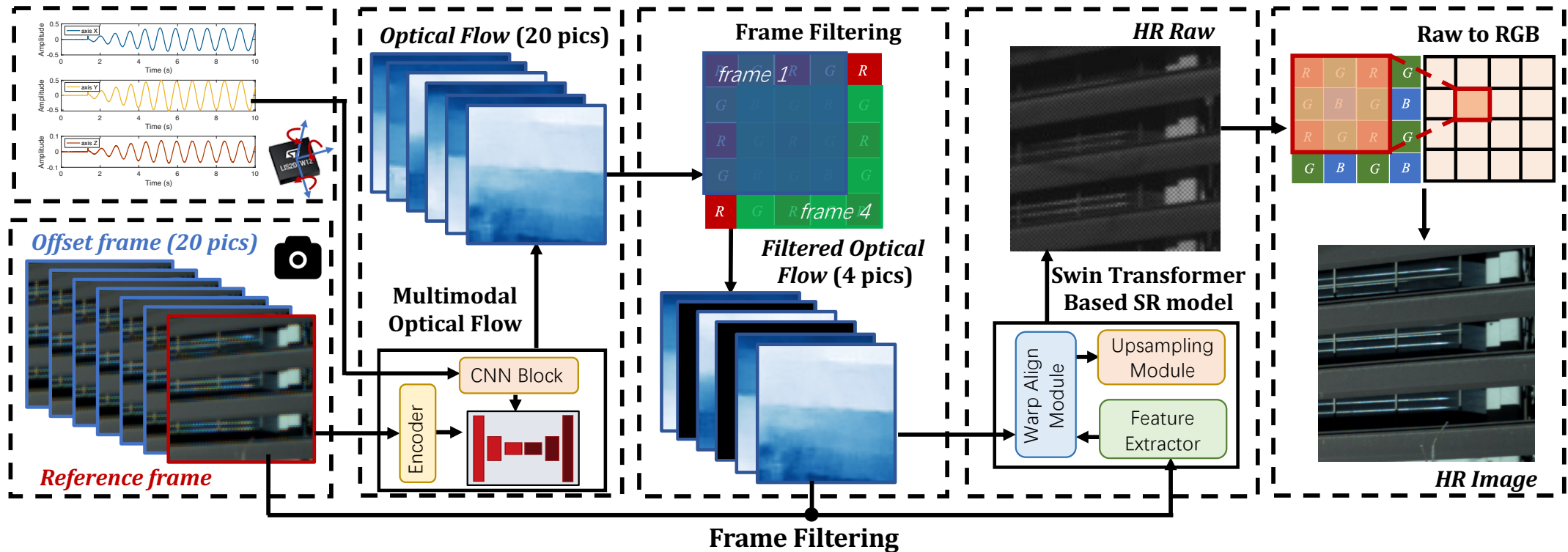


	RAFT[43]	PWCNet[42]	SpyNet[40]	FlowNet[21]	Ours
EPE↓	0.33	0.65	4.16	2.27	<b>0.12</b>
LAT (s)↓	0.76	0.84	0.42	1.35	<b>0.19</b>



Our model achieves the **minimal computational overhead** and **best performance (<0.25 pixel)!**

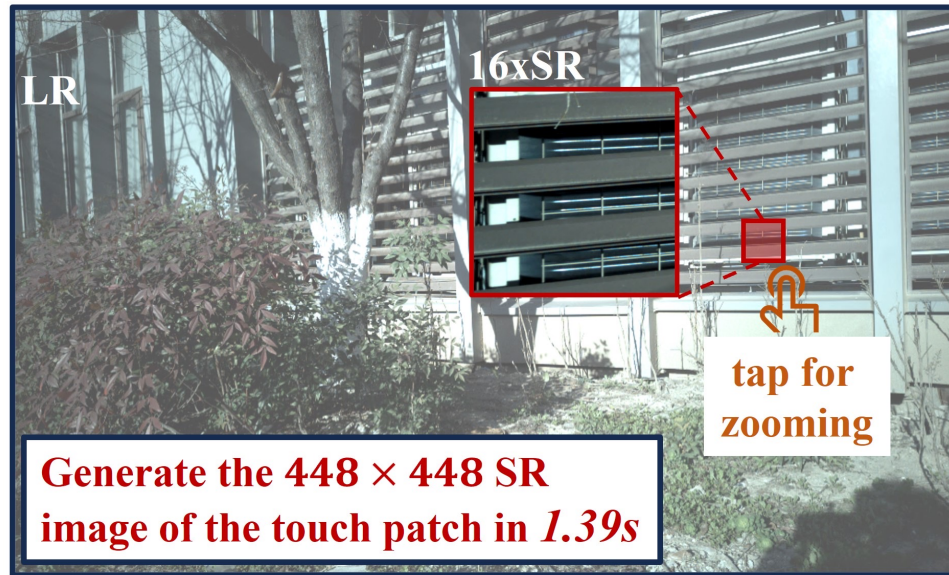
# M<sup>3</sup>Cam system : a lightweight mobile 16× SR system



Overview of our designed M3Cam, a **lightweight** mobile **16×SR** system begin with multi-frame images based on acoustic injection



# Mobile Deployment and “How to use M3Cam”



## 1. Stored File (Save 75% of storage space)

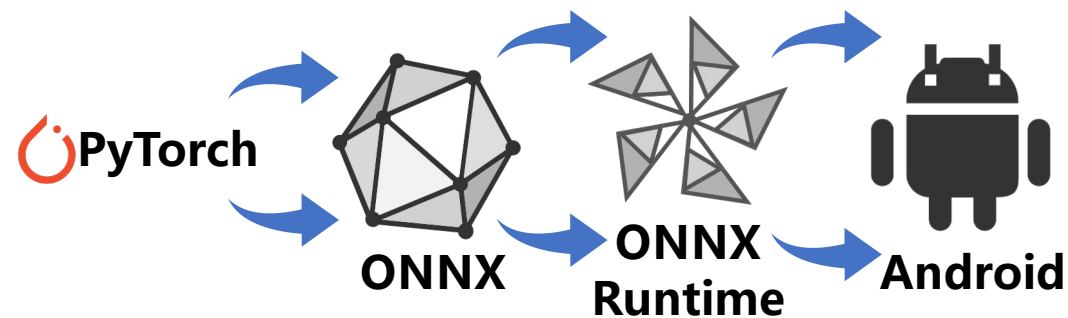
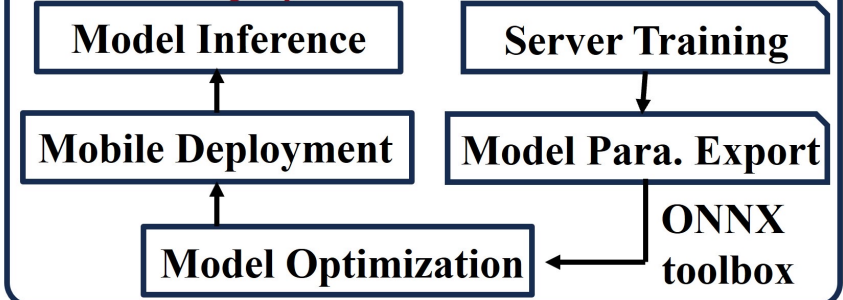


4 LR  
images



IMU  
signals

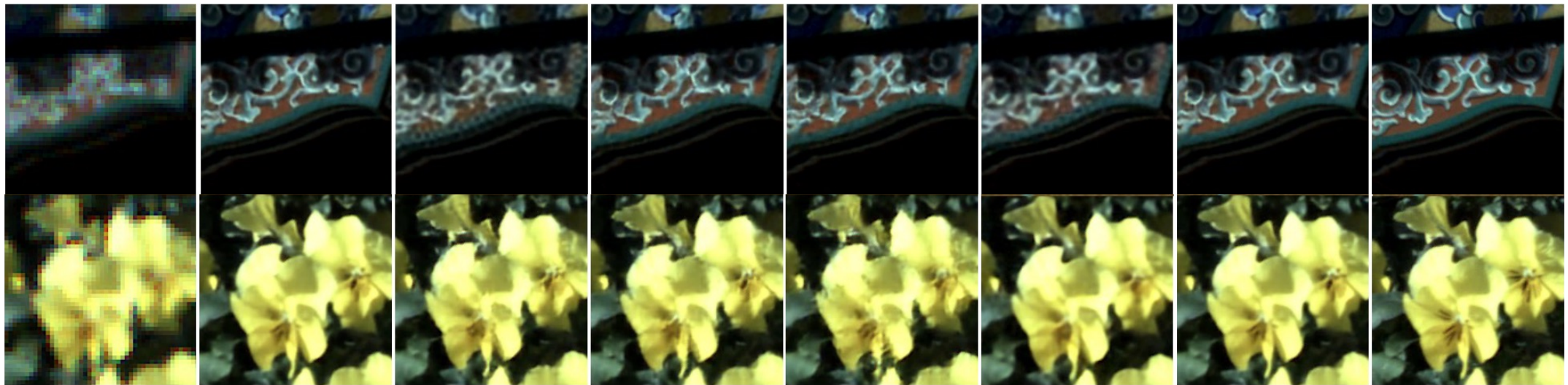
## 2. Model Deployment



# Evaluation : Comparative analysis for various RAW-format MFSR systems

Metrics	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	Para. # ( $10^6$ ) $\downarrow$	Latency (s) $\downarrow$	RAM (MB) $\downarrow$	onnx. (MB) $\downarrow$	Frame # $\downarrow$	Power (J) $\downarrow$
<b>BSRT</b> [31]	35.89	0.8812	0.0847	7.06	8.41	721.3	27.1	12	39.535
<b>DBSR</b> [4]	35.23	0.8876	0.0989	12.94	3.96	827.2	49.3	14	22.703
<b>EBSR</b> [32]	34.96	0.8629	0.0945	9.52	11.58	736	36.7	8	51.068
<b>BIPNet</b> [10]	35.26	0.8603	0.0934	6.67	9.23	753.7	25.6	8	43.337
<b>Burstormer</b> [11]	34.88	0.8610	0.1248	2.49	N/A	N/A	N/A	8	N/A
<b>Ours</b>	<b>36.49</b>	<b>0.8917</b>	<b>0.0687</b>	<b>2.17</b>	<b>1.39</b>	<b>479.4</b>	<b>9.33</b>	<b>4</b>	<b>9.495</b>

End-to-end imaging visualization comparison



(a) LR image

(b) BSRT

(c) DBSR

(d) EBSR

(e) BIPNet

(f) BurstFormer

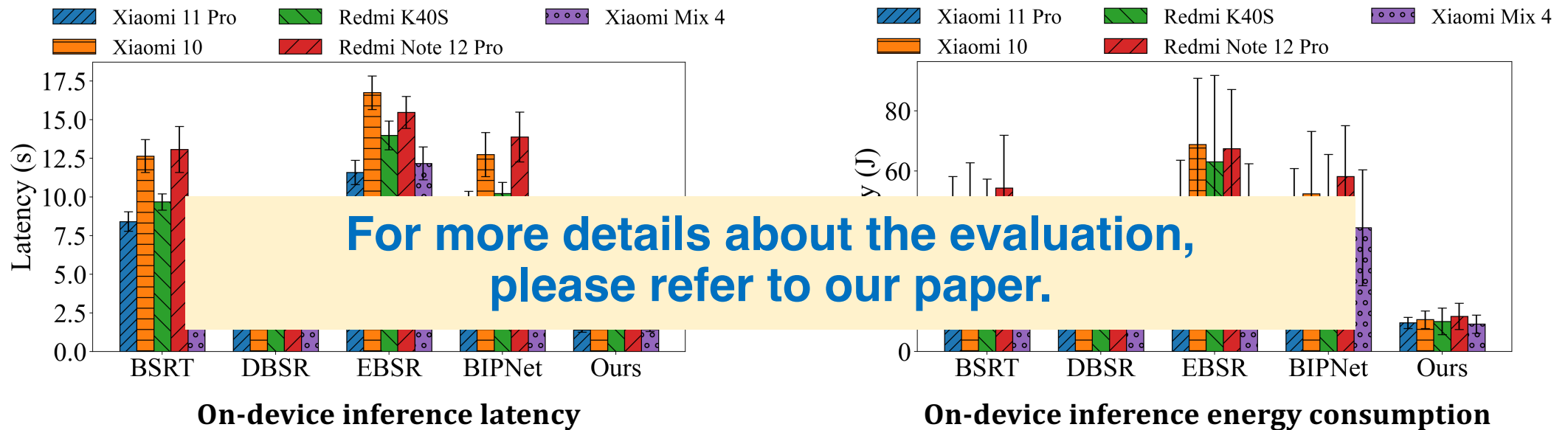
(g) Ours

(h) HR image

# Evaluation : the on-device SR inference performance

Comparison of computation micro benchmark tests on the test smartphones

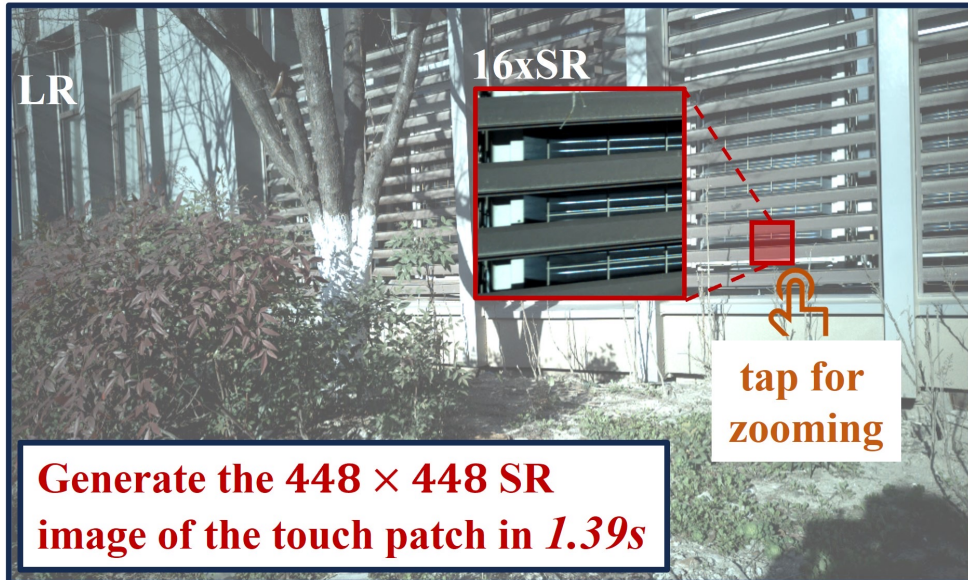
Smartphones	CPU	GPU	MEM	UX	SUM
Xiaomi 11 Pro	177112	198164	138905	167166	681347
Redmi K40S	186205	172761	111705	152525	623196
Xiaomi Mix 4	171285	243559	114380	117454	646678
Xiaomi 10	164215	202188	108087	91243	565733
Redmi Note 12 Pro	147656	184272	89794	134706	556428





# Conclusion

- We propose a novel **multi-modal optical flow** estimation module.
- We propose M<sup>3</sup>Cam, a **lightweight** SR network based on the Swin Transformer.
- We implement a prototype of M<sup>3</sup>Cam and deploy it on various **Android smartphones**.
- M<sup>3</sup>Cam outperforms other systems in both **image quality** and **inference overhead**.



(a) LR

(b) HR

(c) Ours

**Night shot** SR performance comparison





Association for  
Computing Machinery

**SenSys** 2024



SIMON FRASER  
UNIVERSITY

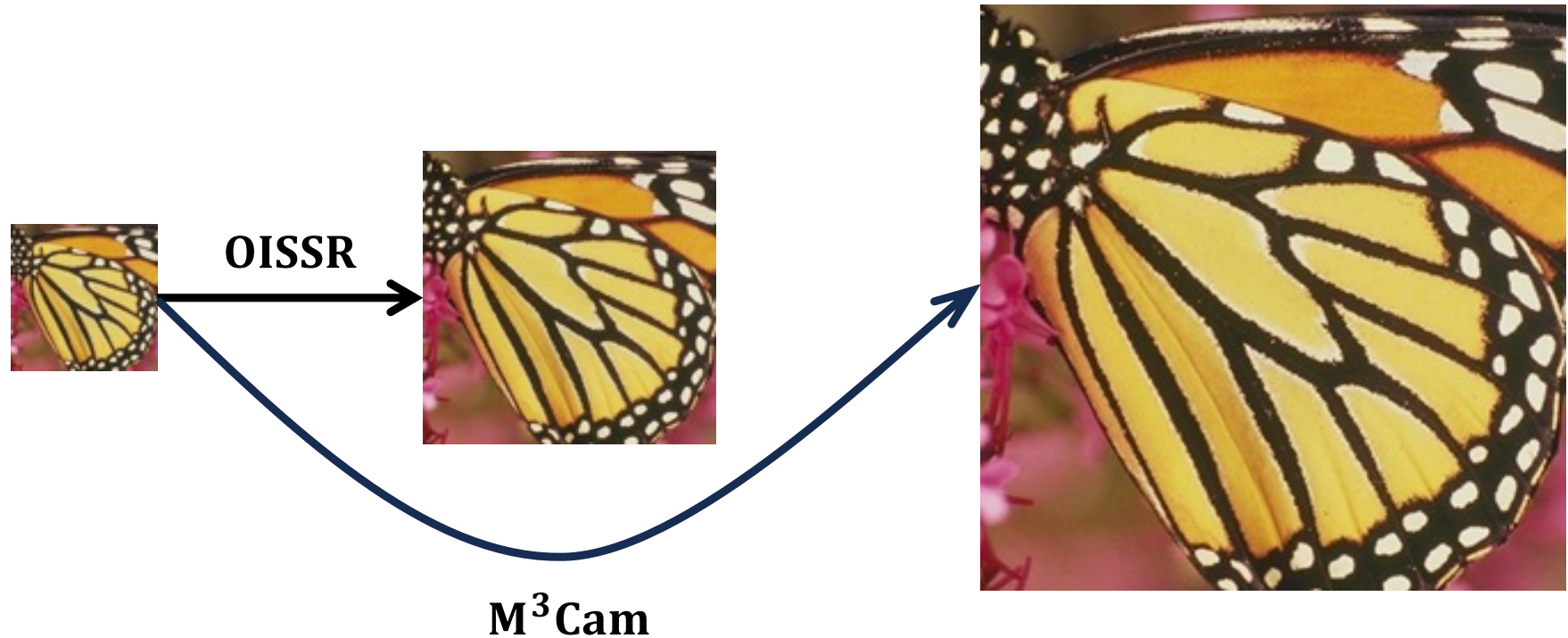


**Microsoft**

**Thanks for listening!**

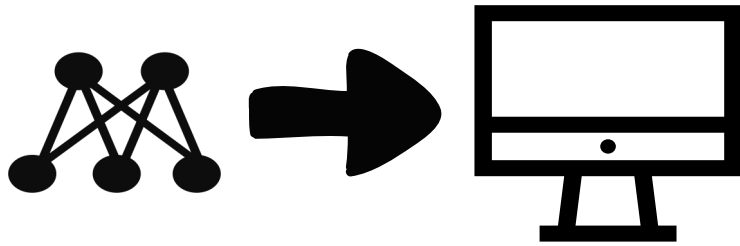


## M<sup>3</sup>Cam vs OISSR

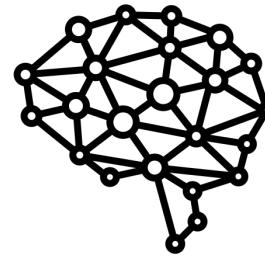


- OISSR achieves high-quality results (PSNR > 35) only in 4x up-sampling (2x length and 2x width), while M3CAM delivers high-quality results in 16x up-sampling tasks.

## M<sup>3</sup>Cam vs OISSR



OISSR



M<sup>3</sup>Cam

- **OISSR lacks validation for real-time deployment on mobile devices, whereas M3CAM successfully implements this capability.**