Association for Computing Machinery

SenSys 2024

M^3Cam: Lightweight Super-Resolution via Multi-Modal Optical Flow for Mobile Cameras

Yu Lu[†&], Dian Ding[†&], Hao Pan[‡*], Yongjian Fu[♦], Liyun Zhang[†], Feitong Tan[₣], Ran Wang[†],
Yi-Chao Chen[†], Guangtao Xue[†], Ju Ren[€]

Shanghai Jiao Tong University[†] Microsoft Research Asia[‡]

Central South University[♦] Simon Fraser University[₣] Tsinghua University[€]

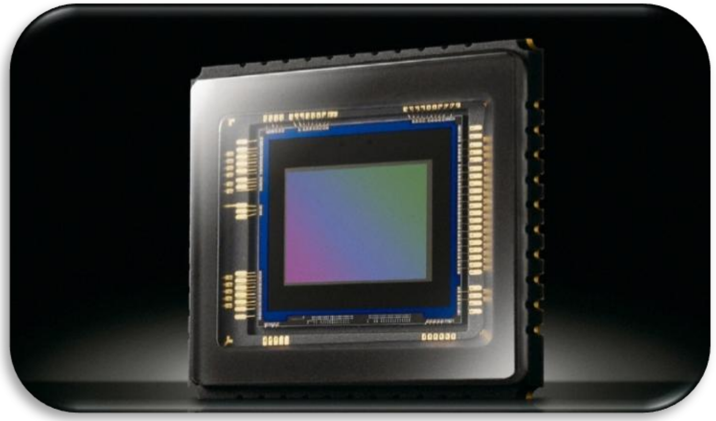
High-resolution photos improve the user's photographic experience

Mobile Photography

Low Resolution Photo → High Resolution Photo

Smartphone' Camera Boost Resolution through Hardware Upgrades

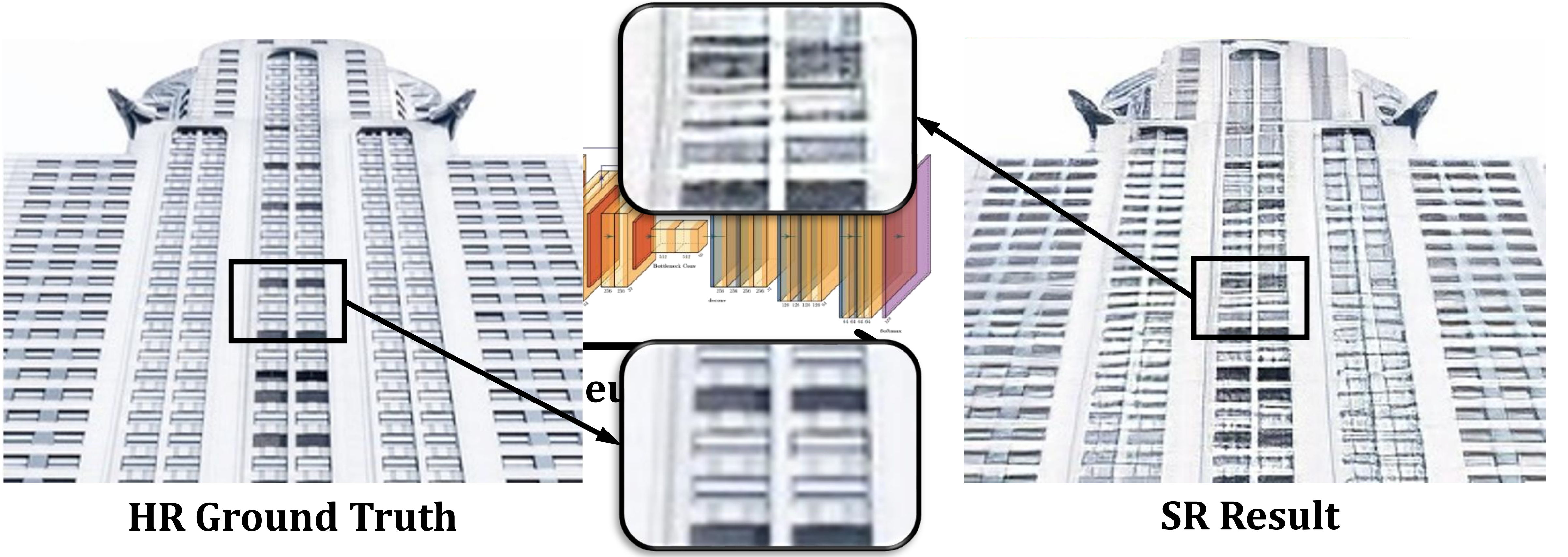
The Moto X30 Pro is the first 200MP phone launched on 8/11/2022, while most flagships still use 50MP sensors today.



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

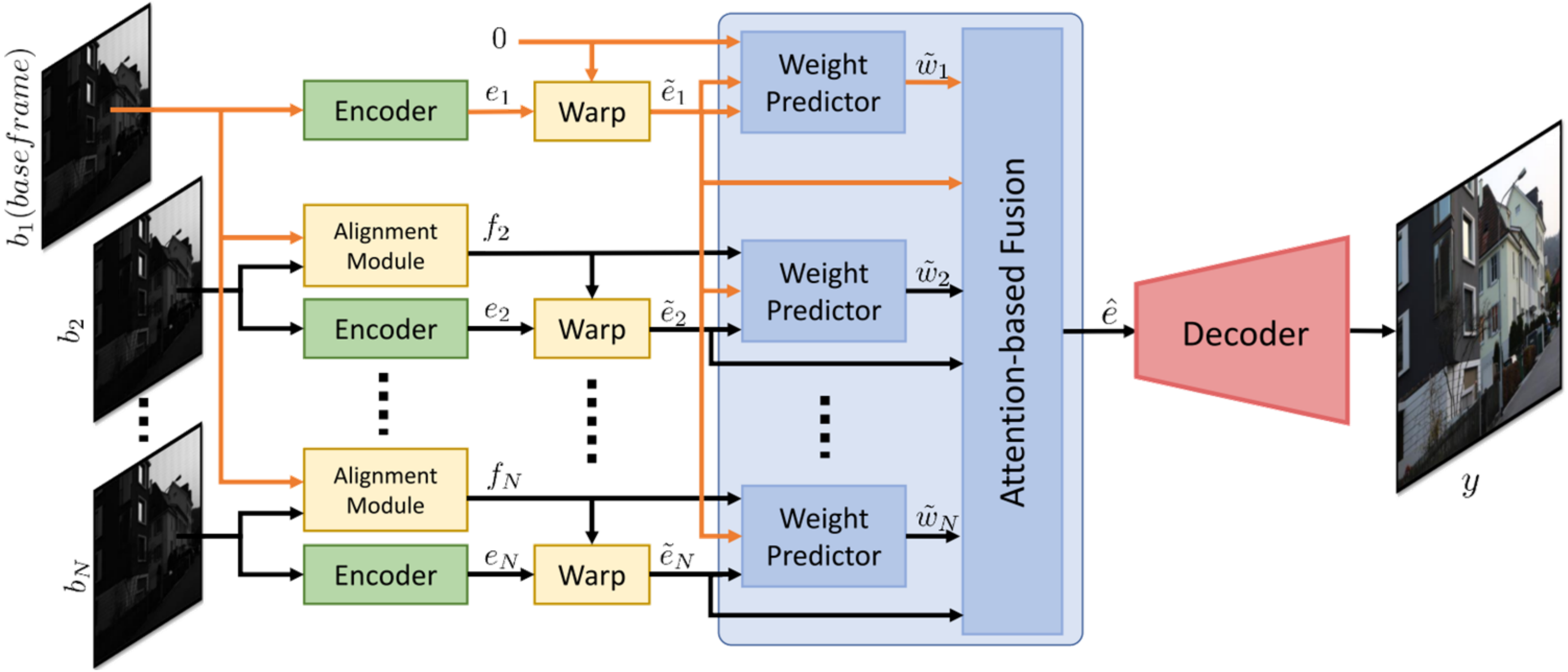


HR Ground Truth

SR Result

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

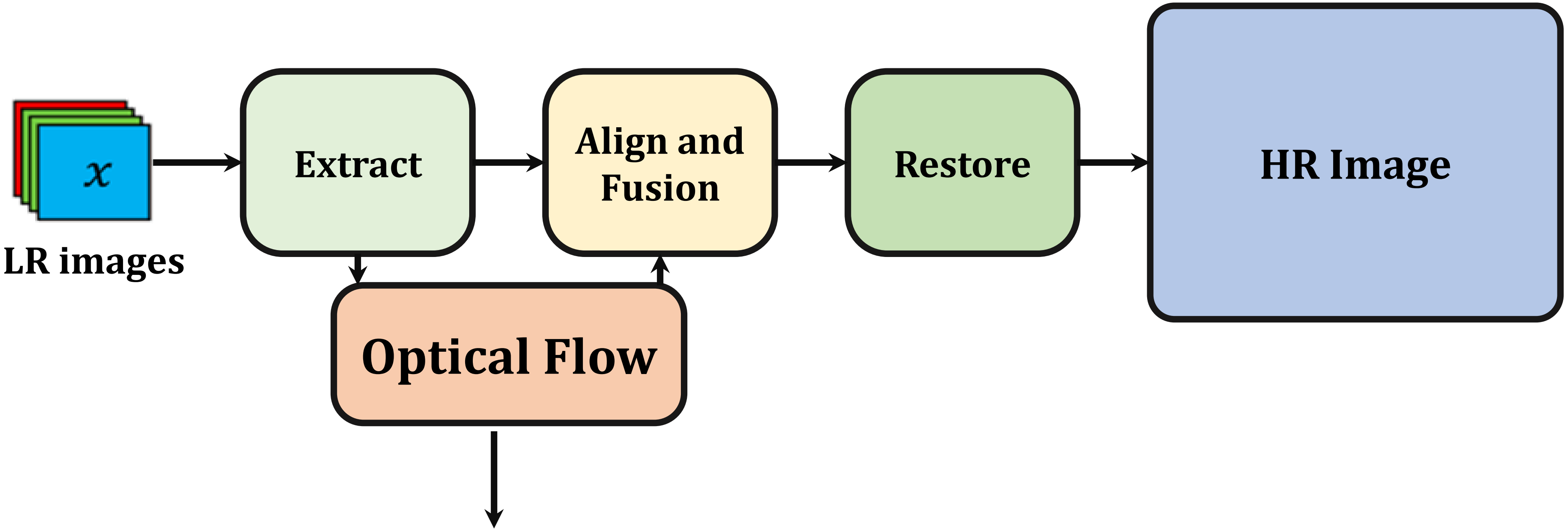
Multi-frame Super-resolution (MFSR)

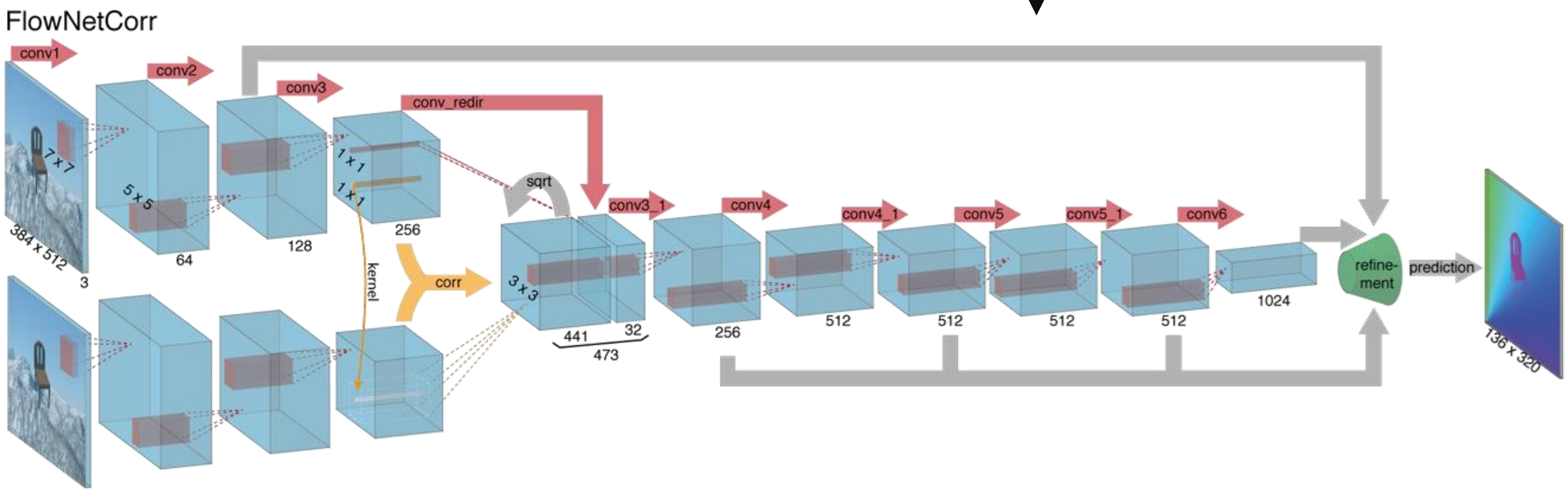


MFSR yields better results than SFSR, because it can restore more information about the real scene through multiple sampling.

MFSR on mobile devices and its challenges

Capture LR images with camera at different positions

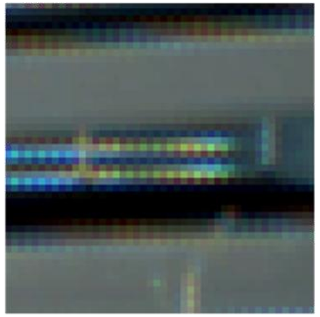




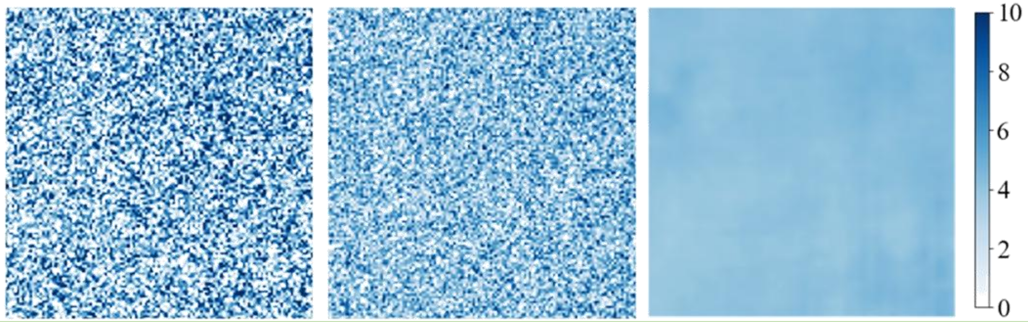
Optical flow-based image alignment is challenging.

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

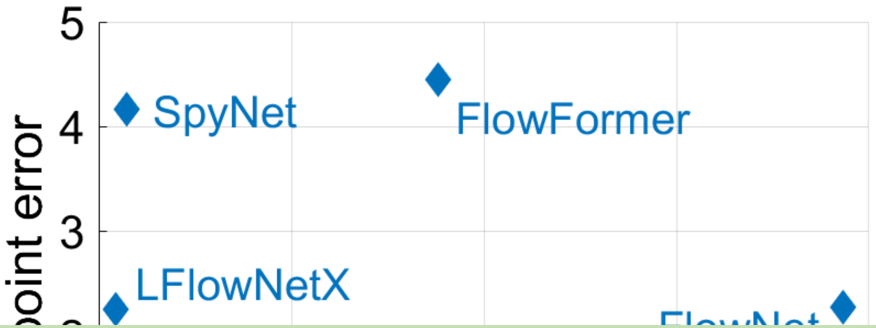
Low-accuracy optical flow impacts the quality of MFSR



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

Parameter (Mb)

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

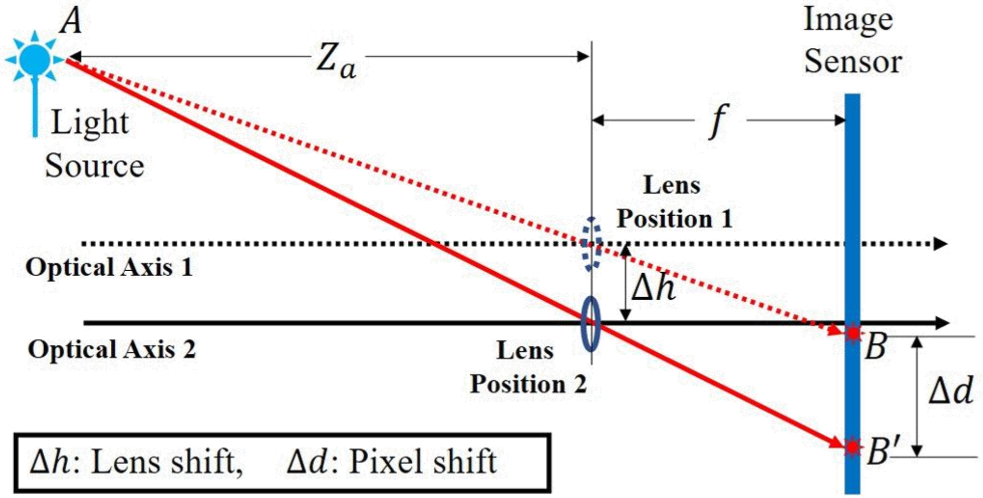
Accuracy of optical flow significantly affects super-resolution imaging results

16-flod Super-Resolution on mobile needs <0.25-pixel error

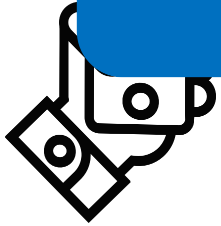
Intuition: Enhance optical flow accuracy with an auxiliary modality

Frame 1

Frame 2



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

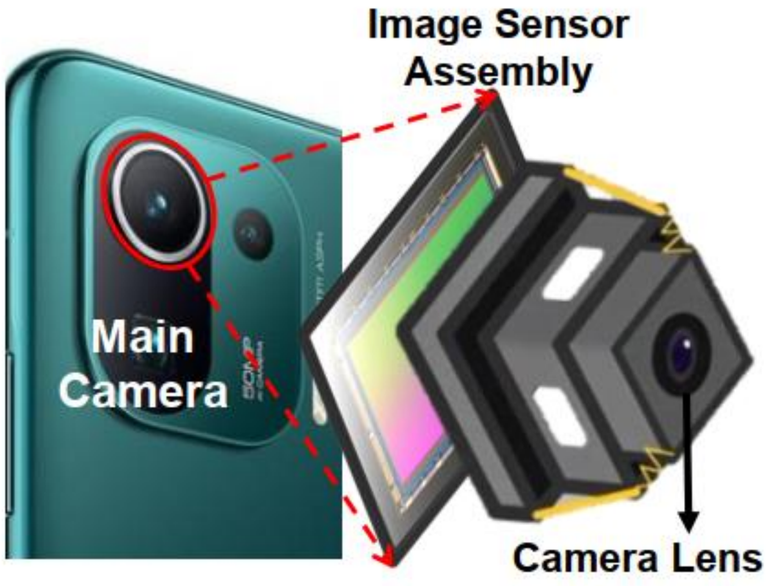


However, steady and regular camera movement is nearly impossible for common users.

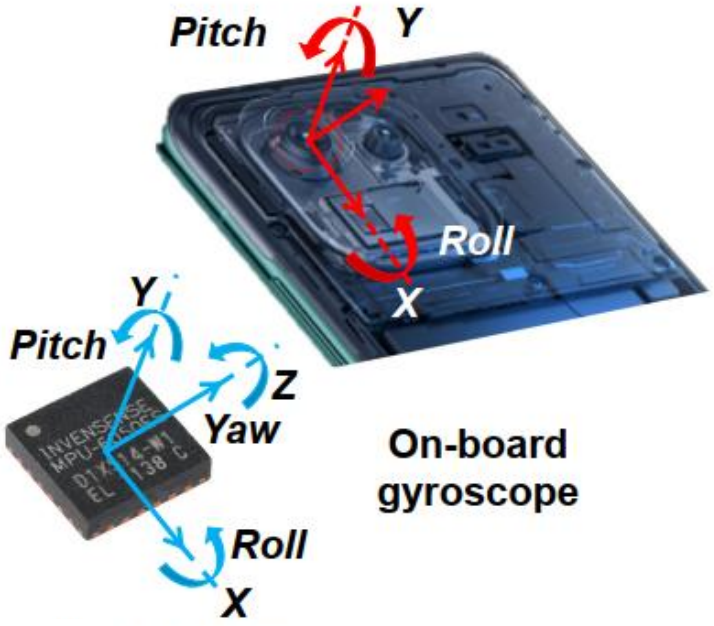
information

a lightweight NN model

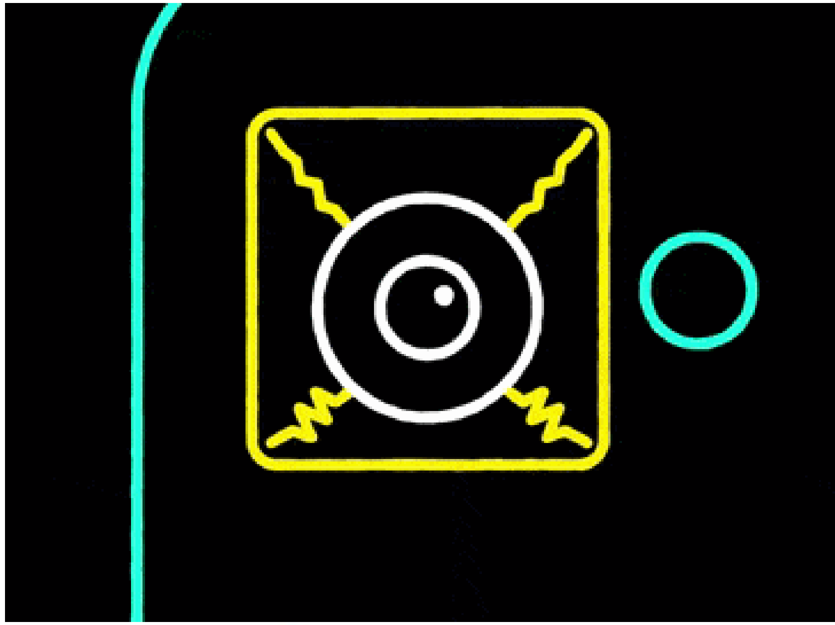
OIS module can control lens motion with the phone stationary



An OIS-supported camera built in the smartphone



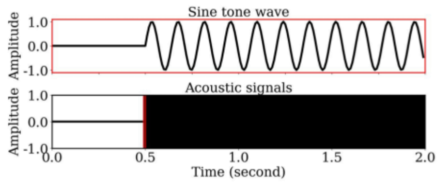
MEMS sensors sense camera movements



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.

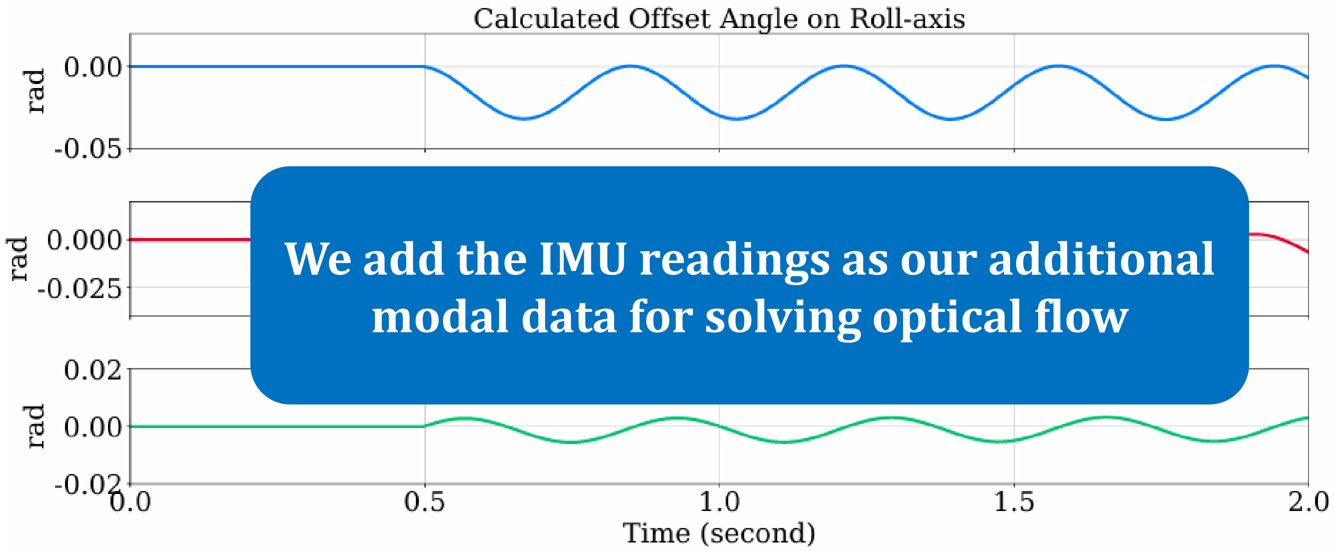
Additional Modality : Controlled IMU readings for lens movement



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

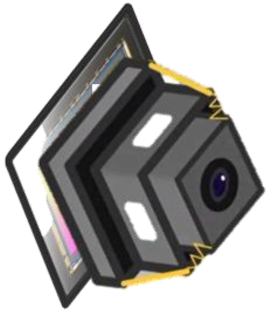
Acoustic injection

Built-in MEMS IMU

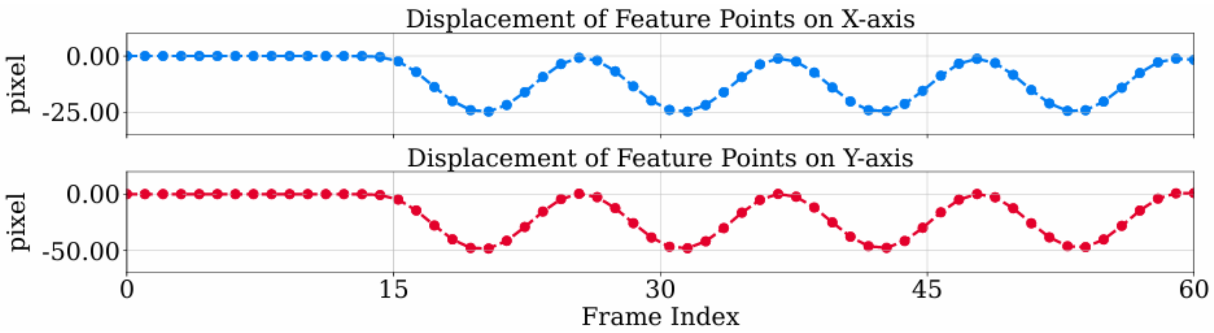


We add the IMU readings as our additional modal data for solving optical flow

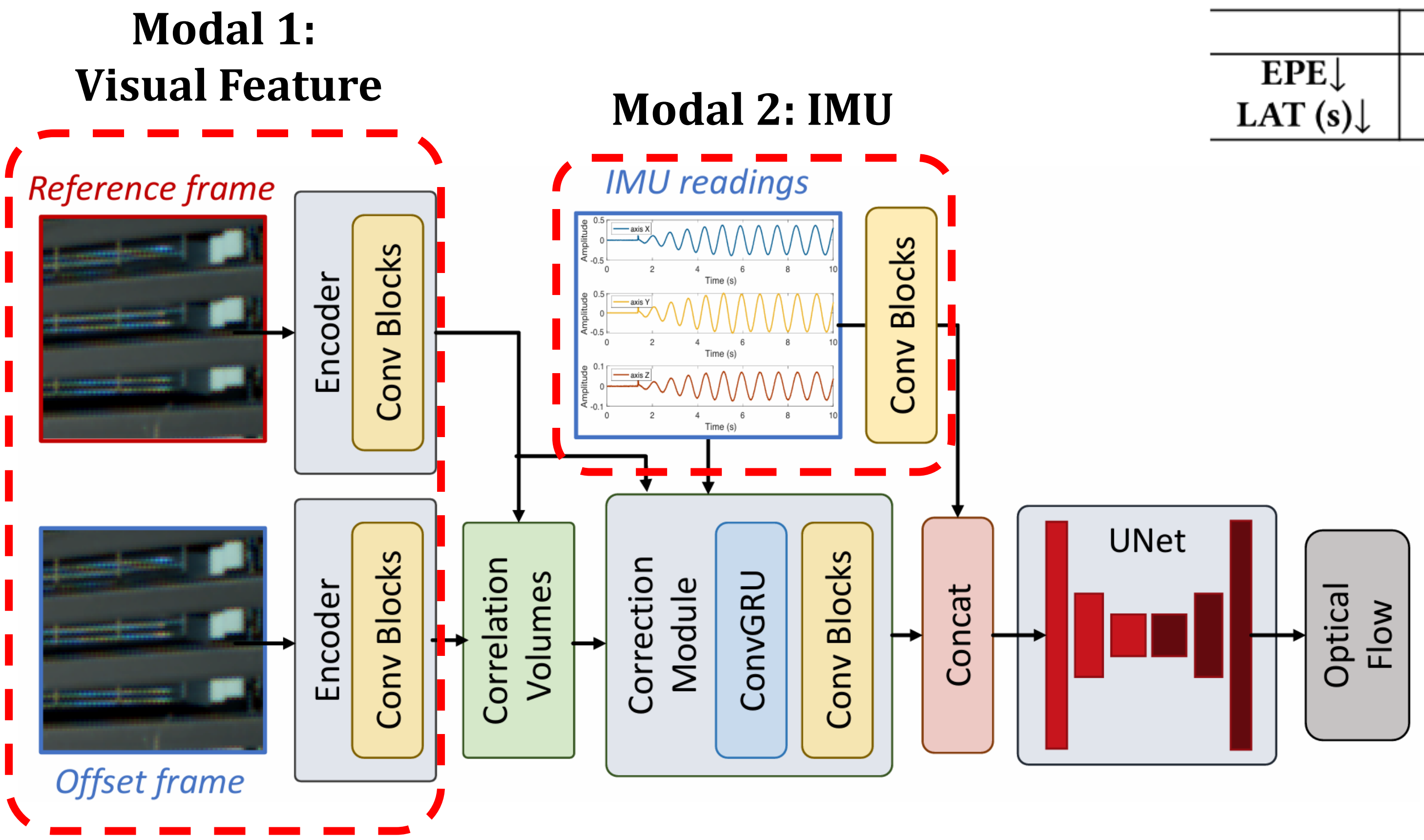
OIS module controls regular lens shift



Regular optical flow pattern

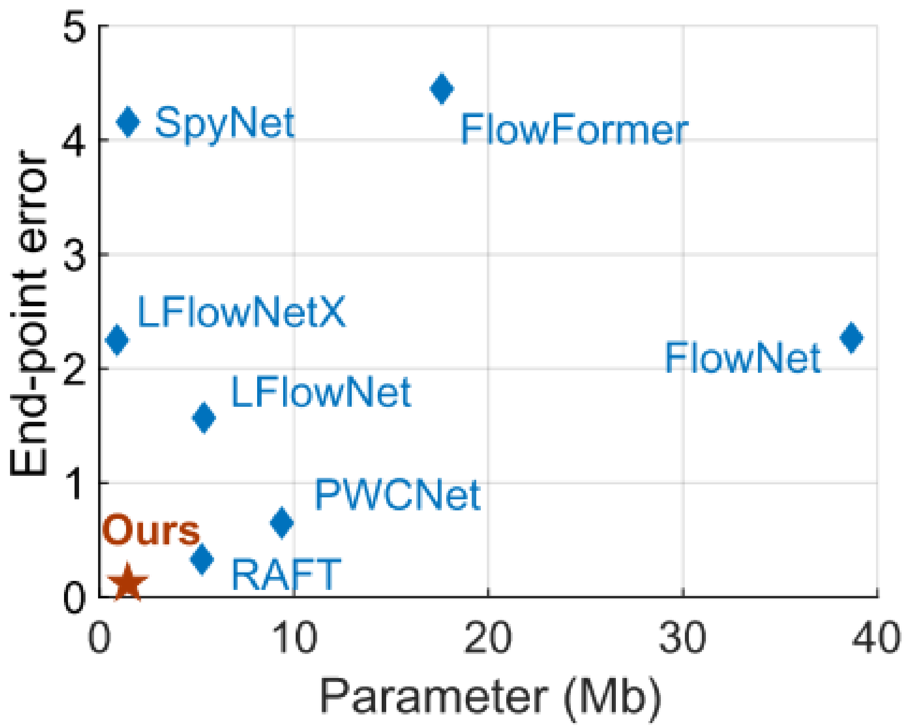


Our proposed multi-modal optical flow estimation module



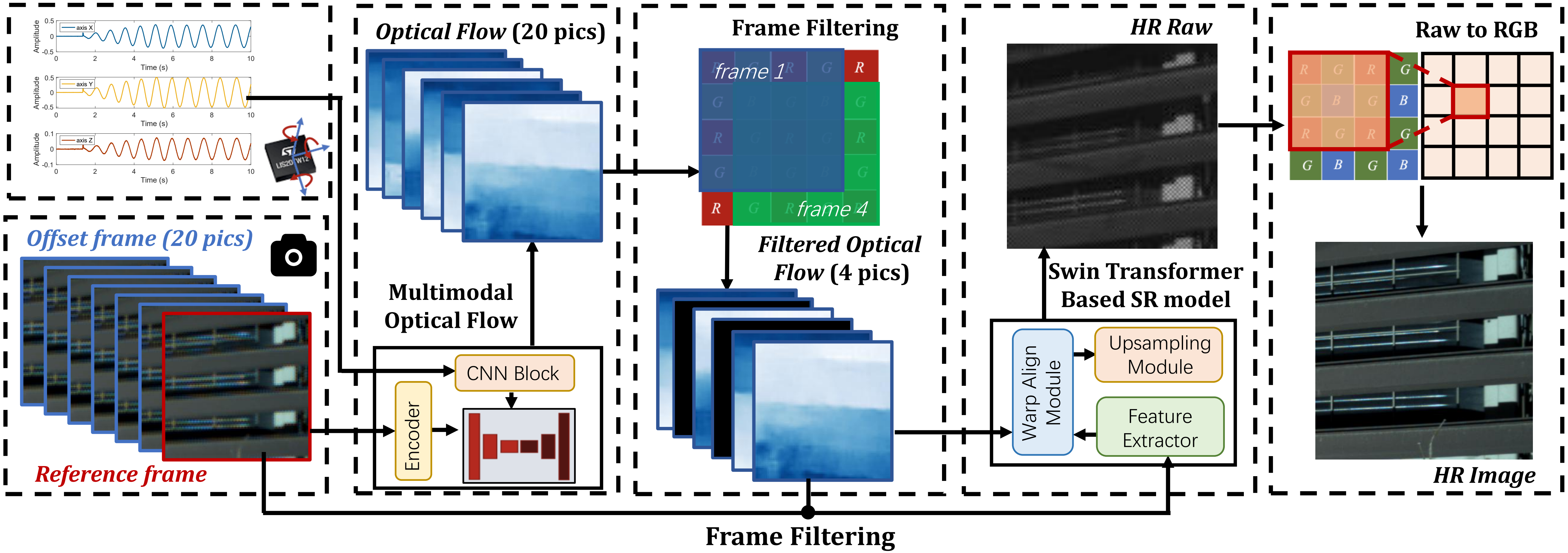
Multi-modal optical flow estimation model

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



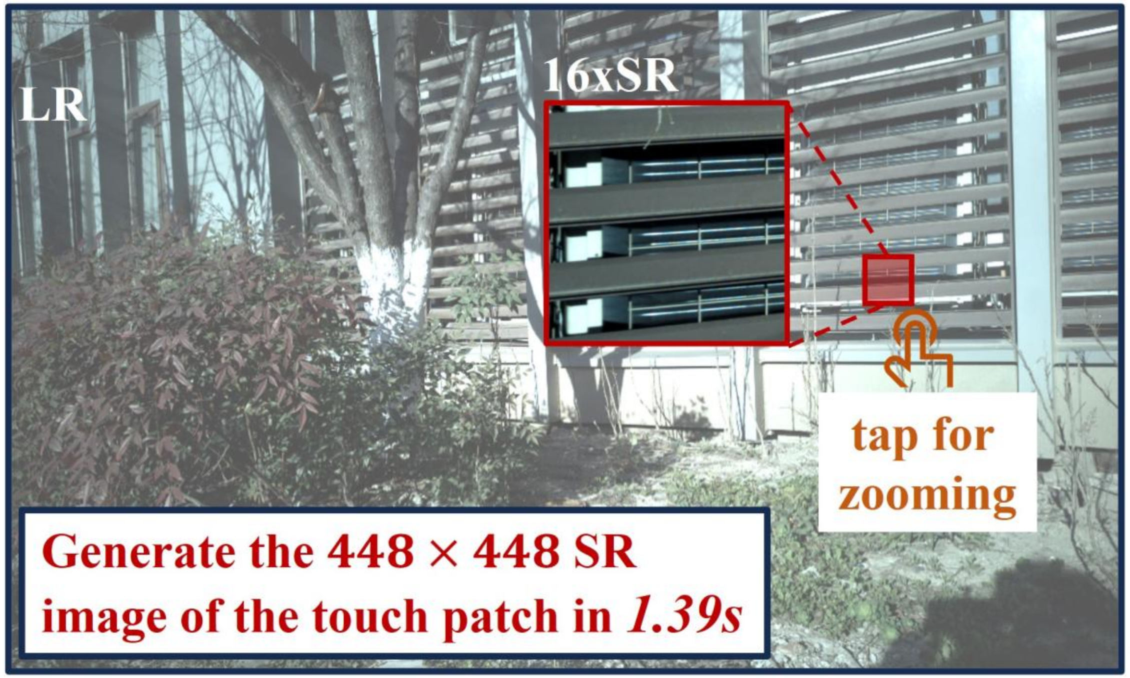
Our model achieves the minimal computational overhead and best performance (<0.25 pixel)!

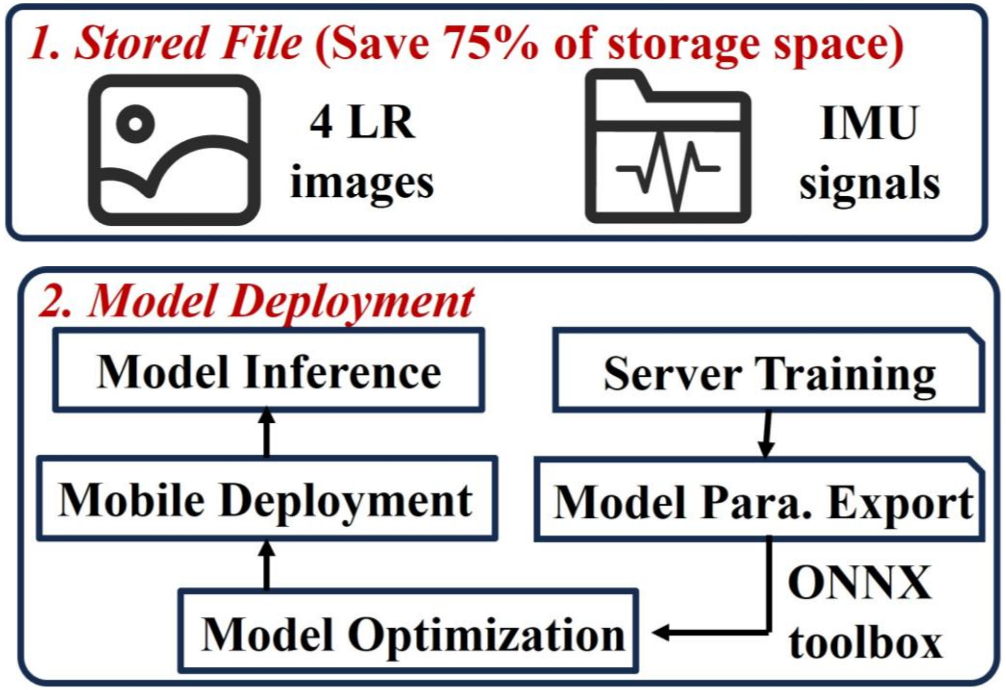
M^{3}Cam system : a lightweight mobile 16× SR system

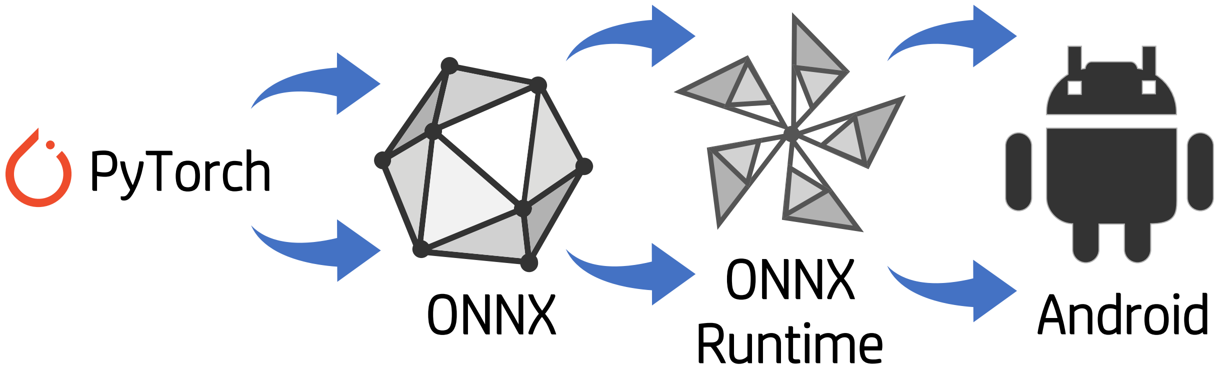


Overview of our designed M3Cam, a lightweight mobile 16×SR system begin with multi-frame images based on acoustic injection

Mobile Deployment and “How to use M3Cam”



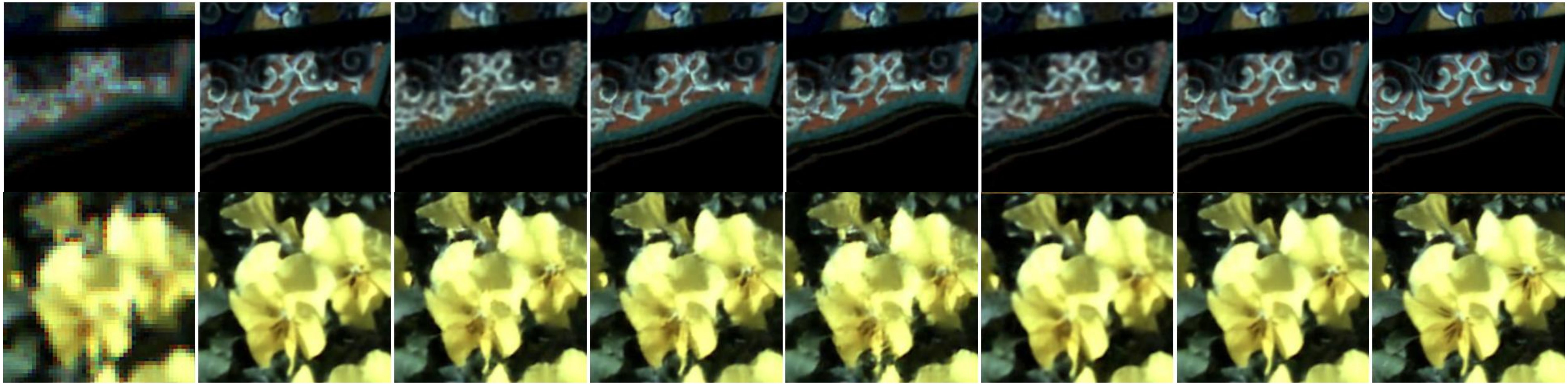




Evaluation : Comparative analysis for various RAW-format MFSR systems

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

End-to-end imaging visualization comparison

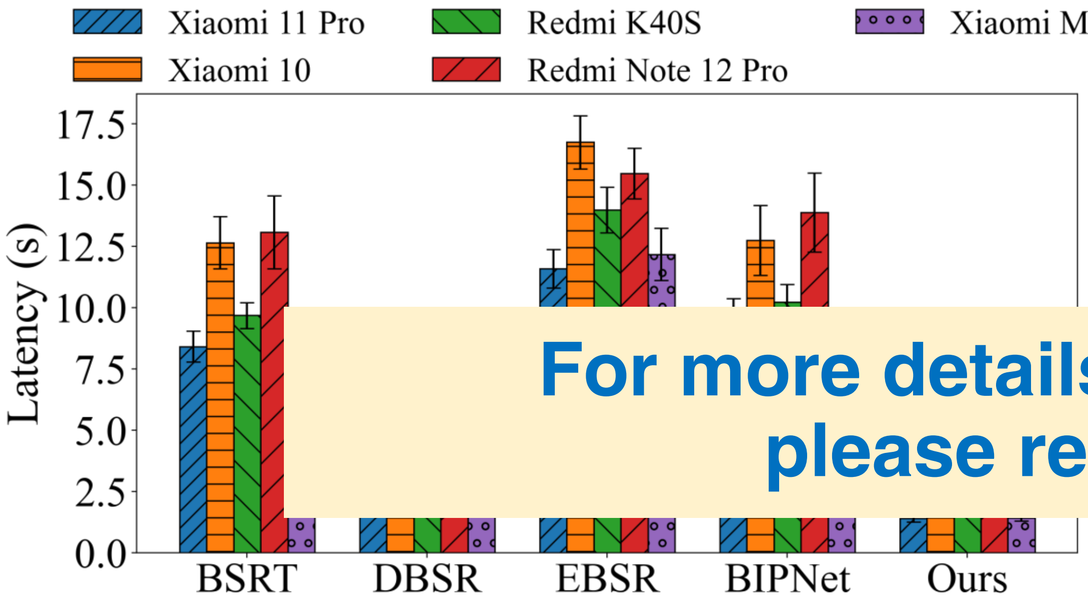


(a) LR image (b) BSRT (c) DBSR (d) EBSR (e) BIPNet (f) BurstFormer (g) Ours (h) HR image

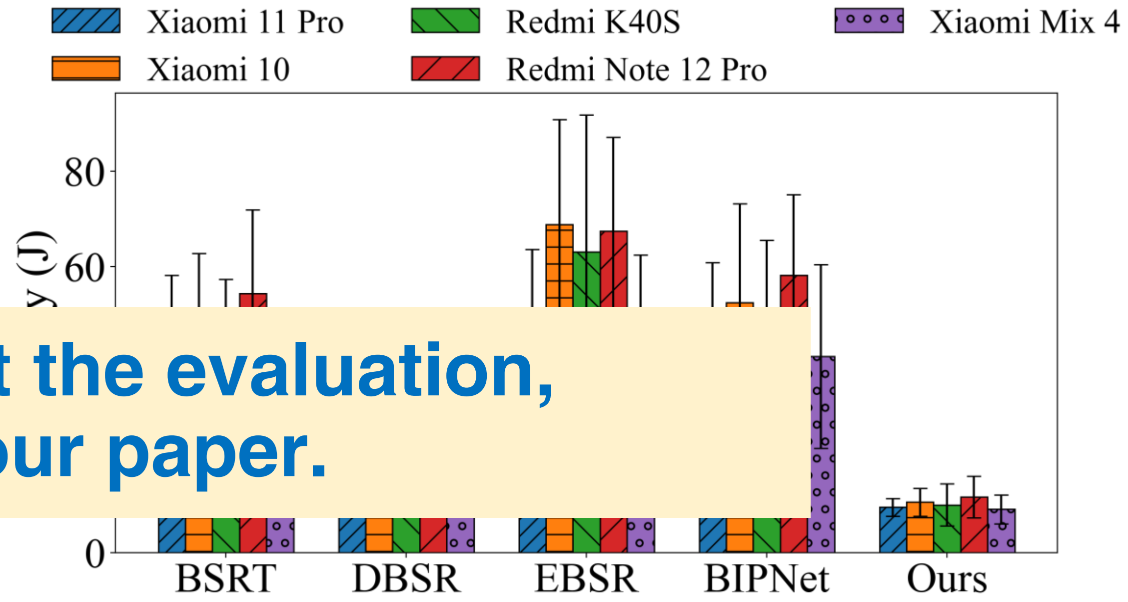
Evaluation : the on-device SR inference performance

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



On-device inference latency

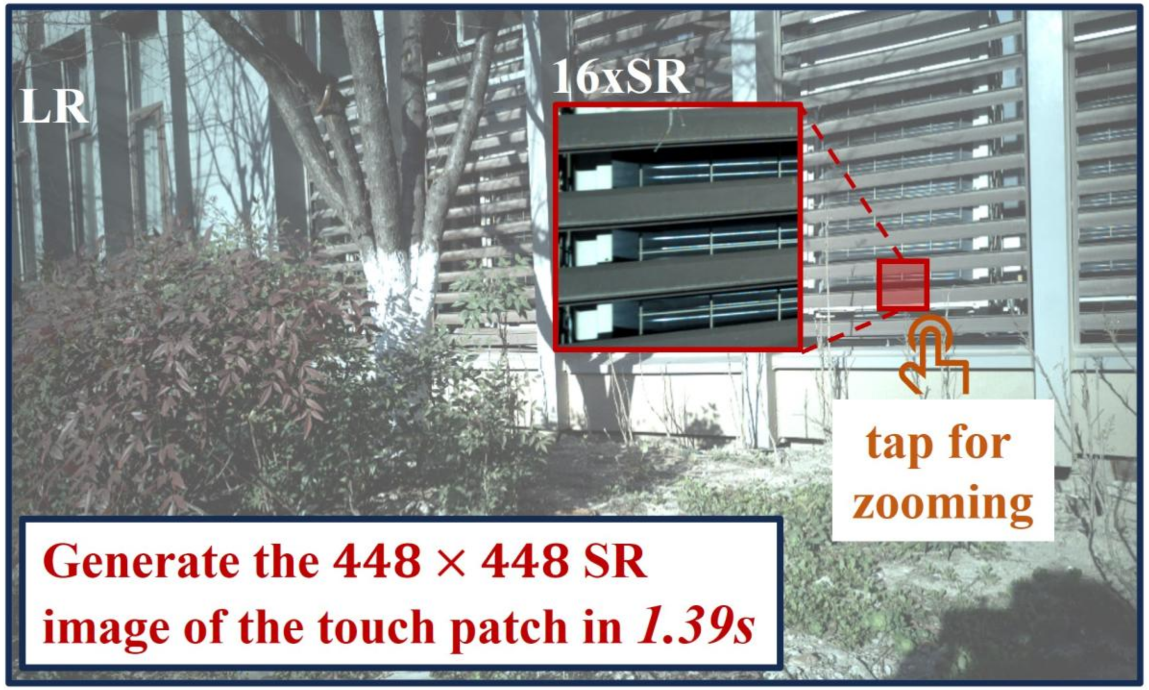


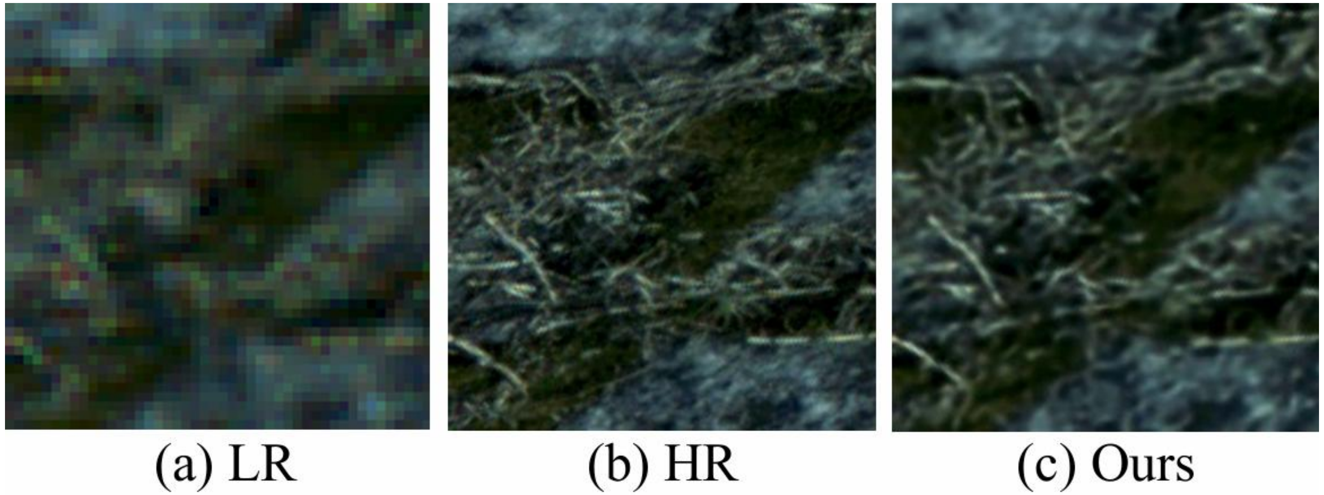
On-device inference energy consumption

For more details about the evaluation, please refer to our paper.

Conclusion

- We propose a novel multi-modal optical flow estimation module.
- We propose M³Cam, a lightweight SR network based on the Swin Transformer.
- We implement a prototype of M³Cam and deploy it on various Android smartphones.
- M³Cam outperforms other systems in both image quality and inference overhead.





Night shot SR performance comparison

Association for Computing Machinery

SenSys 2024

SHANGHAI JIAO TONG UNIVERSITY

CENTRAL SOUTH UNIVERSITY

TSINGHUA UNIVERSITY

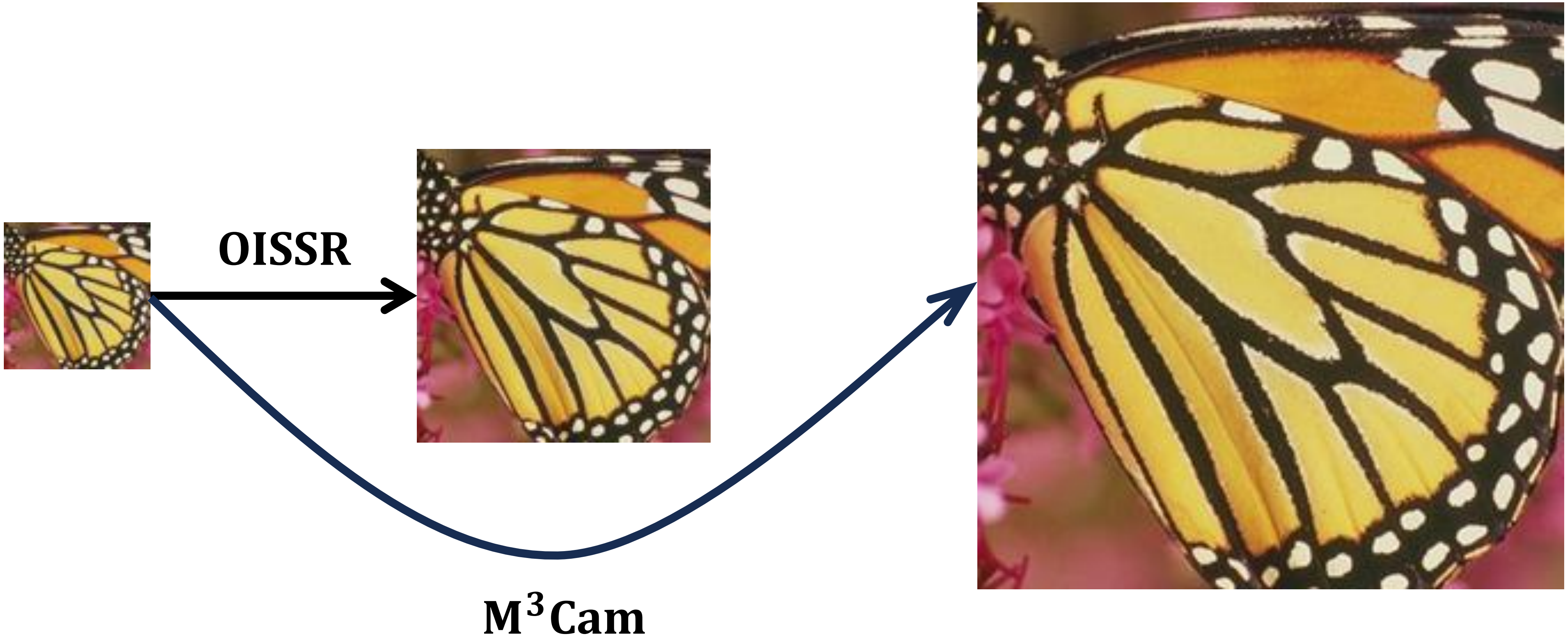
SFU SIMON FRASER UNIVERSITY

Microsoft

Thanks for listening!

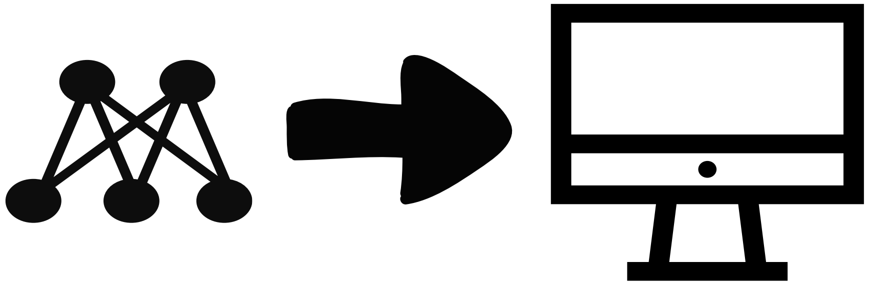

Q
A

M^{3}Cam vs OISSR

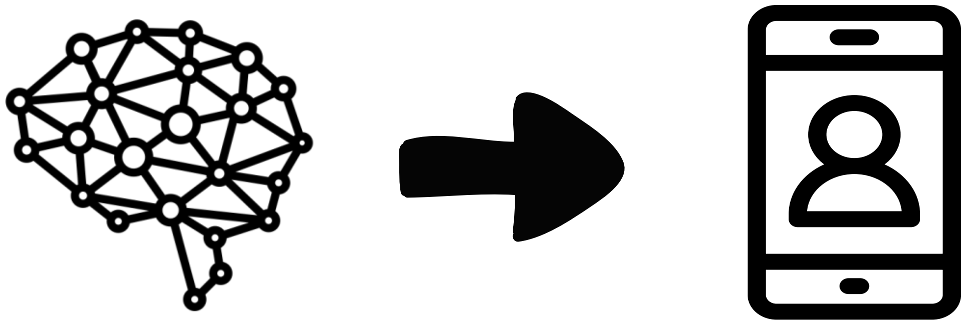


- **OISSR achieves high-quality results (PSNR > 35) only in 4x up-sampling (2x length and 2x width), while M3CAM delivers high-quality results in 16x up-sampling tasks.**

M^3Cam vs OISSR



OISSR



M^3Cam

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