Omnidirectional Image Super-Resolution via Bi-Projection Fusion

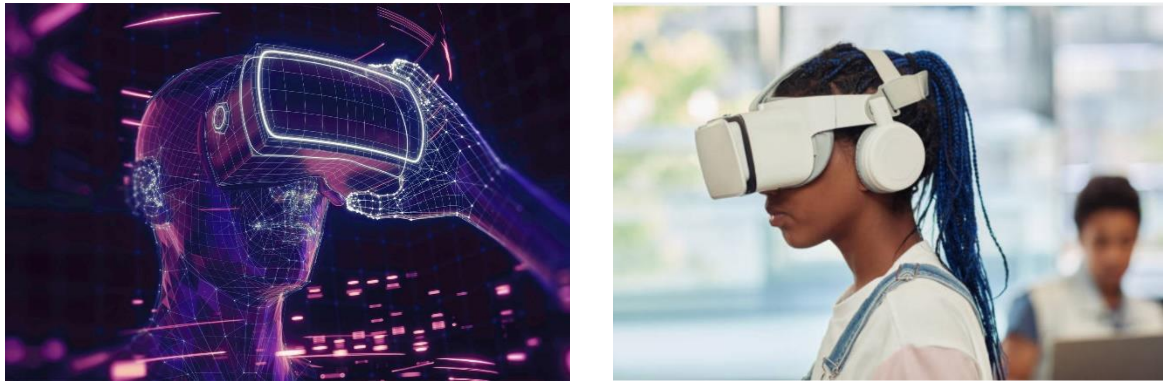
Jiangang Wang[1], Yuning Cui[2], Yawen Li[3], Wenqi Ren[1*], Xiaochun Cao[1]

[1]Shenzhen Campus of Sun Yat-sen University
[2]Technical University of Munich
[3]Beijing University of Posts and Telecommunications

1. Research Background

中山大學
SUN YAT-SEN UNIVERSITY

Medical



Metaverse

Industry and Education

To achieve an immersive experience, omnidirectional video requires ultra-high resolution (8K-16K).

2. Related Work and Motivation

- **Single Image Super-Resolution**

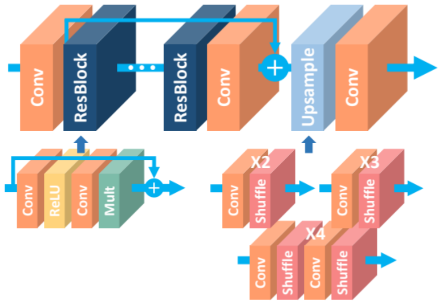


Figure 3: The architecture of the proposed single-scale SR network (EDSR).

EDSR

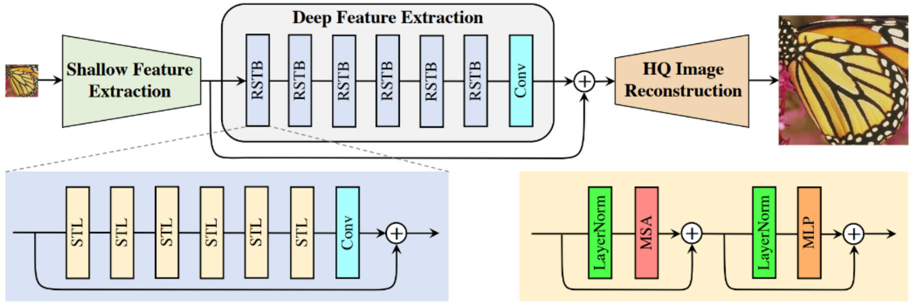
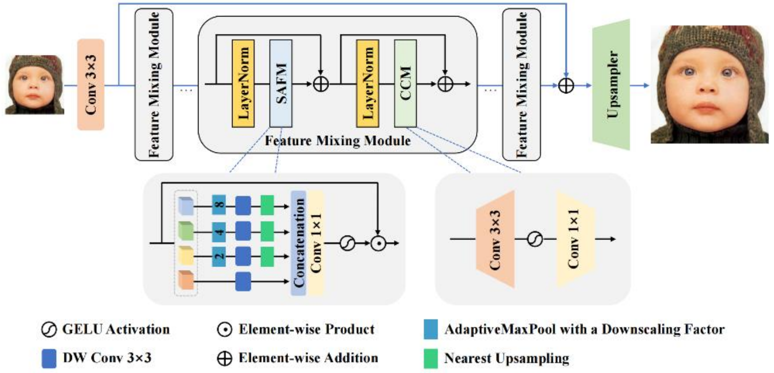


Figure 2: The architecture of the proposed SwinIR for image restoration.

SwinIR



SAFMN

There are differences between the two image domains,
which makes the traditional 2D image super-resolution algorithm not suitable for omnidirectional Images.

Traditional 2D Images

VS

Omnidirectional Images (ODIs)

2. Related Work and Motivation

- **Omnidirectional Image Super-Resolution**

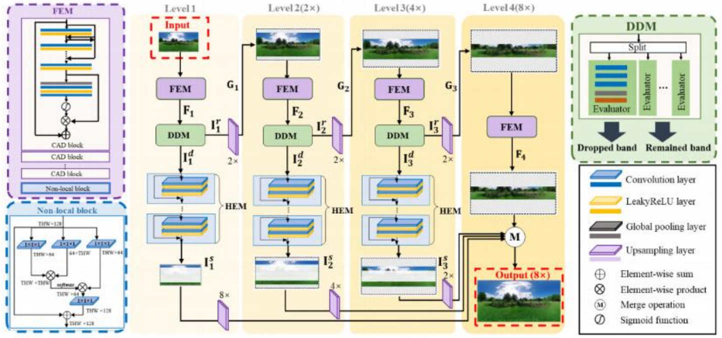


Fig. 6. The progressive architecture of the proposed LAU-Net+. Each level is composed of feature enhancement module (FEM), drop-band decision module (DDM) and high-latitude enhancement module (HEM). The final HR image is obtained by merging the outputs from different levels.

LAU-Net

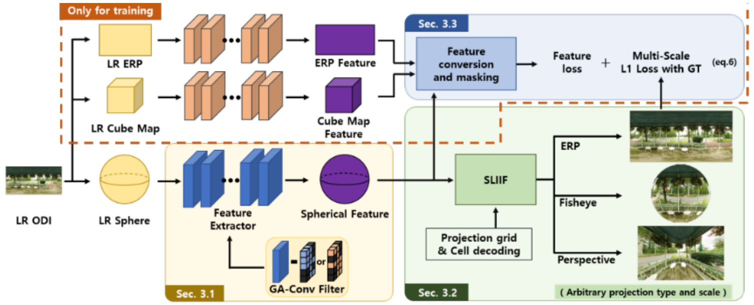


Figure 2. Overall framework of the proposed SphereSR.

SphereSR

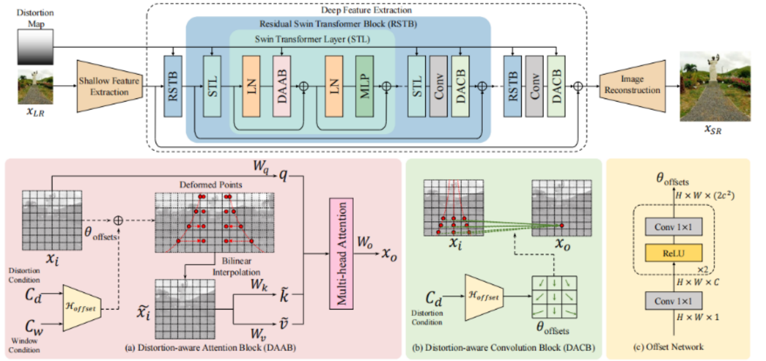
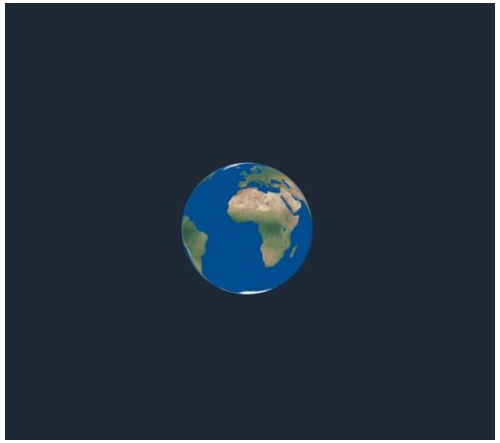


Figure 4. Overall illustration of OSRT. From SwinIR [24], we replace the standard multi-head self-attention block with DAAB and insert DACB behind the end of the RSTB. Channel dimensions of $\theta_{offsets}$ in DAAB and DACB are 2 and 18, respectively.

OSRT

Existing omnidirectional image super-resolution only uses ERP,
and the geometric features of omnidirectional images are not fully utilized.



Earth Projection

ERP

CMP

ISP

SSP

EAC

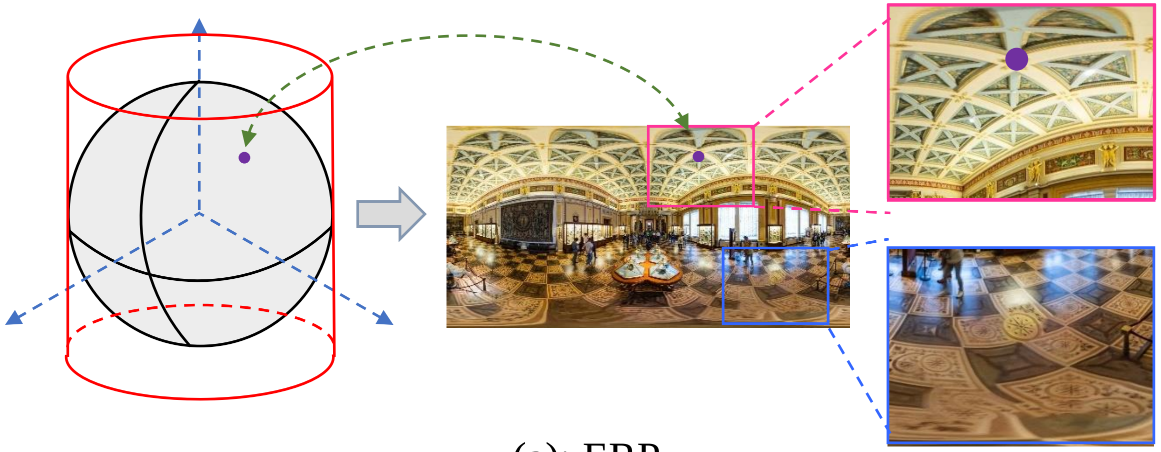
TSP

OHP

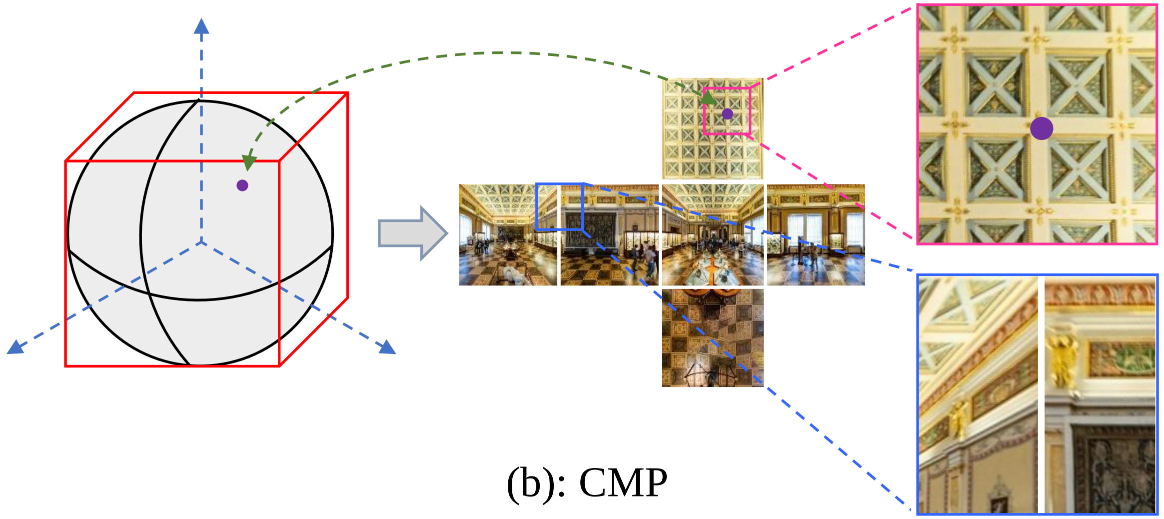
Different omnidirectional image projection formats

3. ODIs Analysis

- **Two common omnidirectional image projections:**
  - Equirectangular projection (ERP)
  - Cubemap projection (CMP)



(a): ERP



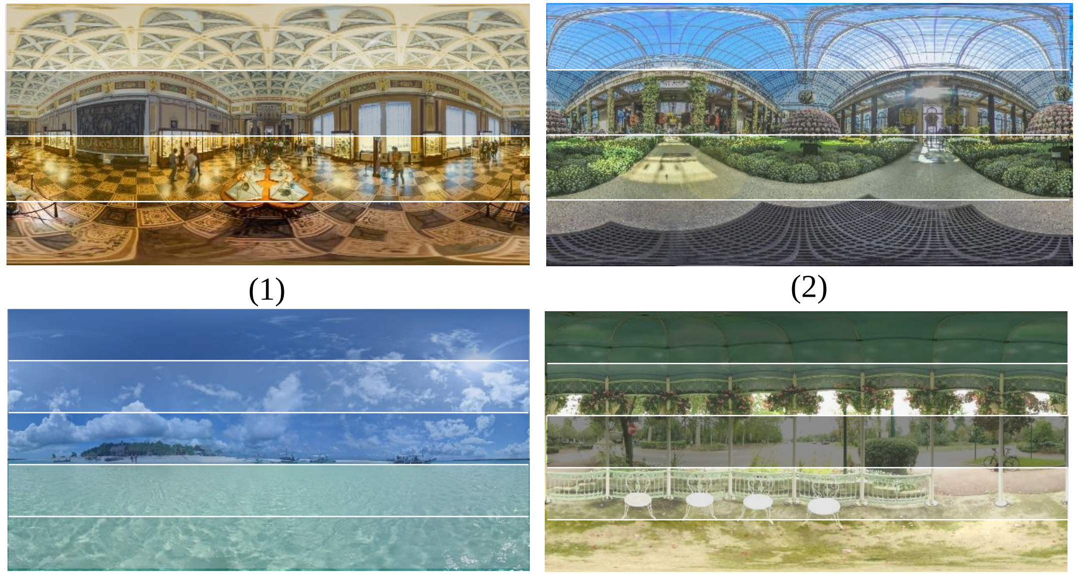
(b): CMP

ERP and CMP are complementary

Projection	Advantage	Disadvantage
ERP	Global perspective	High distortions, particularly at high latitudes
CMP	Low distortions	Local perspective and discontinuities between the individual surfaces

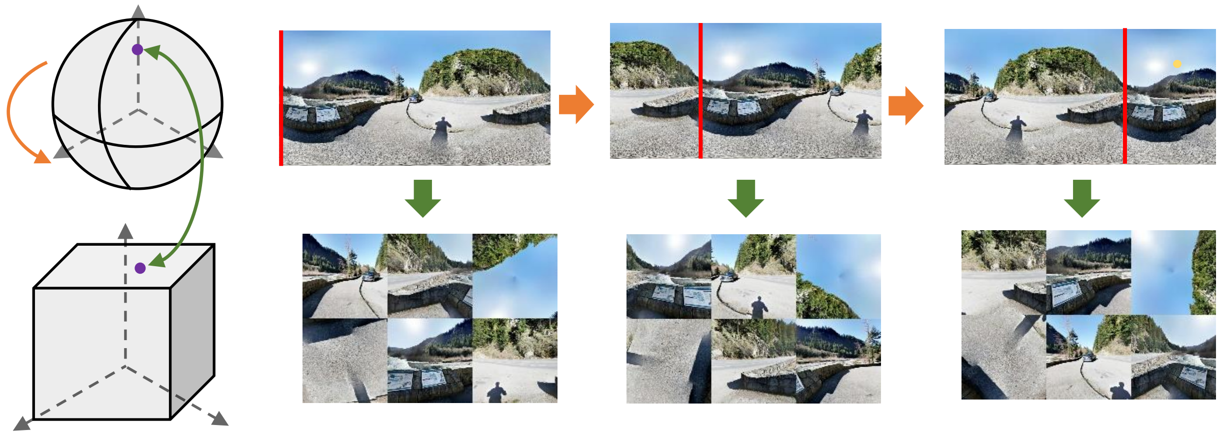
3. ODIs Analysis

- **ERP and CMP geometric properties**



(1) (2) (3) (4)

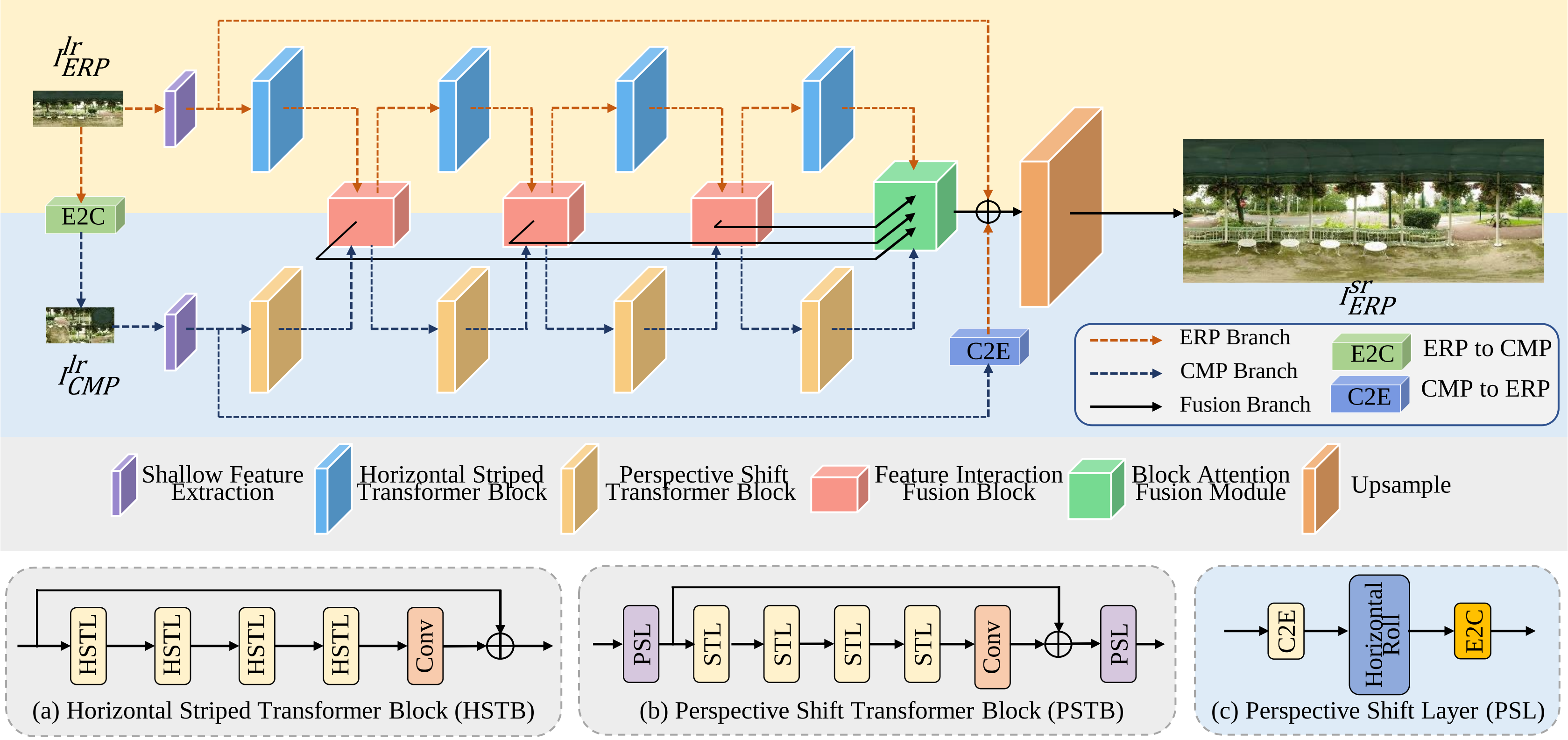
(a): ERP Horizontal Similarity



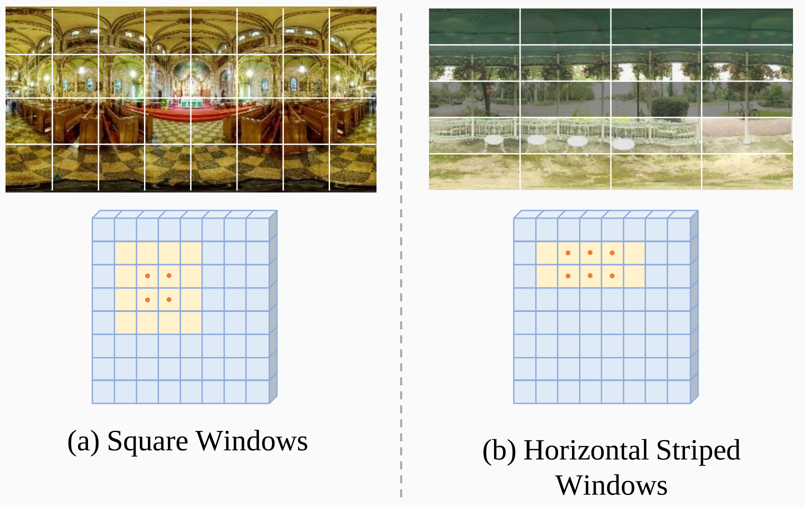
(b): CMP Perspective Variability

Projection	**Geometric properties**	**Model structure design**
ERP	Horizontal Similarity	Local similarity modeling
CMP	Perspective Variability	Multi-perspective information fusion

4. Methodology



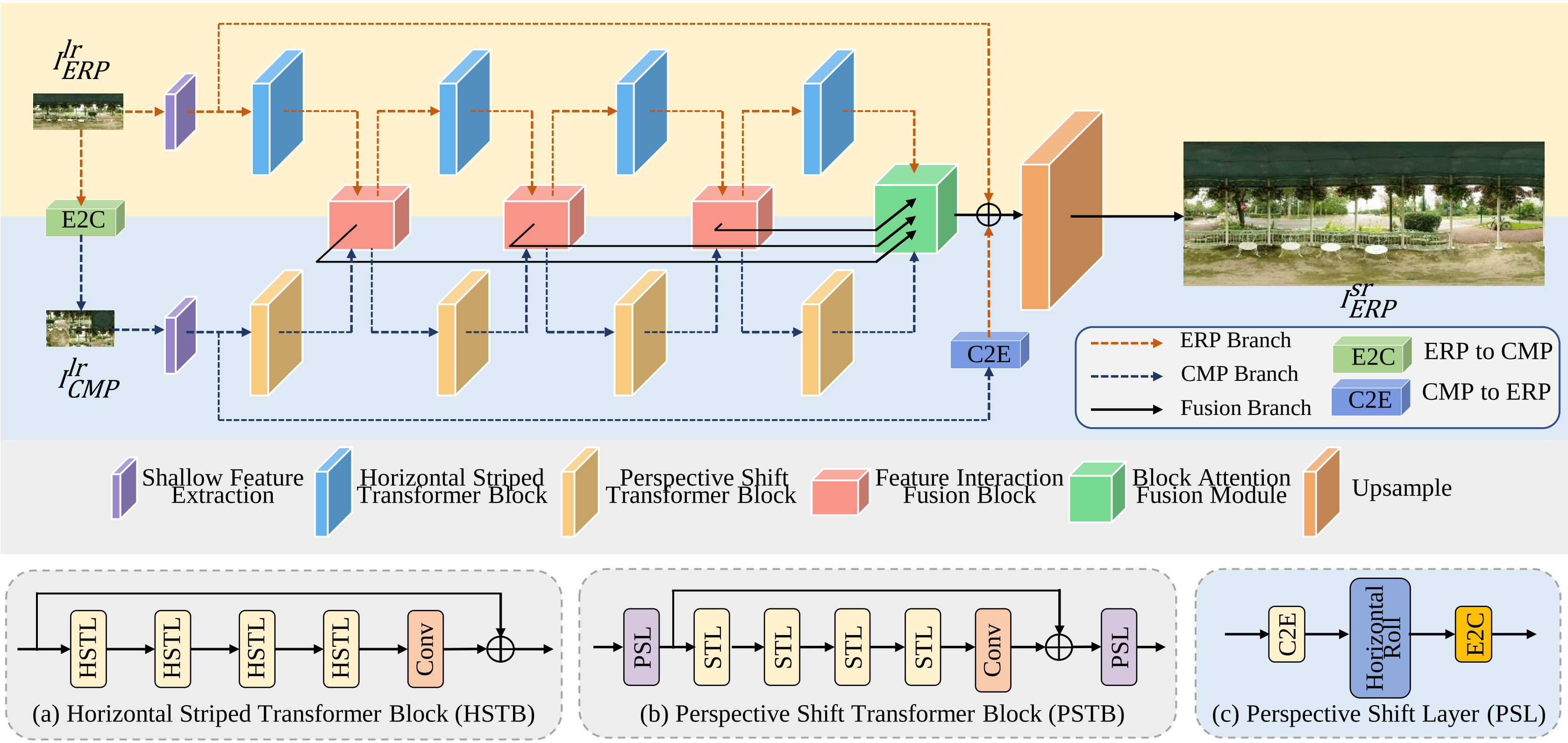
The overall diagram illustrates the architecture of BPOSR



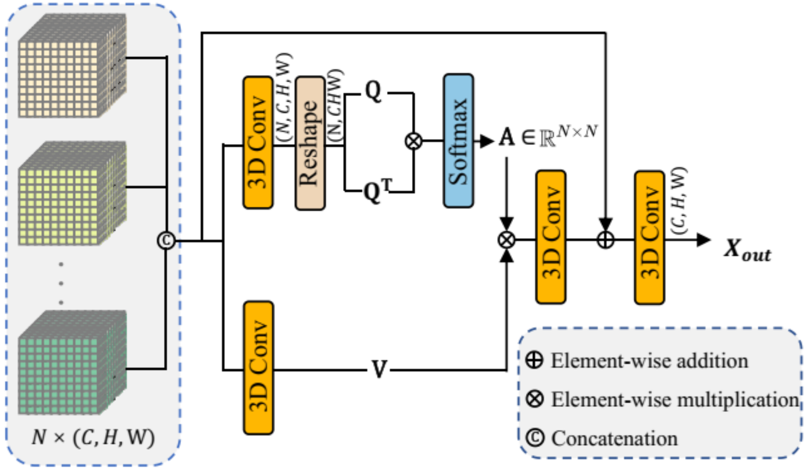
Different self-attention windows

(a) Square Windows (b) Horizontal Striped Windows. As can be seen, Horizontal Striped Windows are more effective in capturing the similarity within ERP compared to Square Windows.

Projection	Geometric properties	Model structure design
ERP	Horizontal Similarity	Local similarity modeling
CMP	Perspective Variability	Multi-perspective information fusion



The overall diagram illustrates the architecture of BPOSR



Block Attention Fusion Module

BAFM receives input from different projections and depths, employing a 3D self-attention mechanism to fuse all the features.

5. Experiments and Results

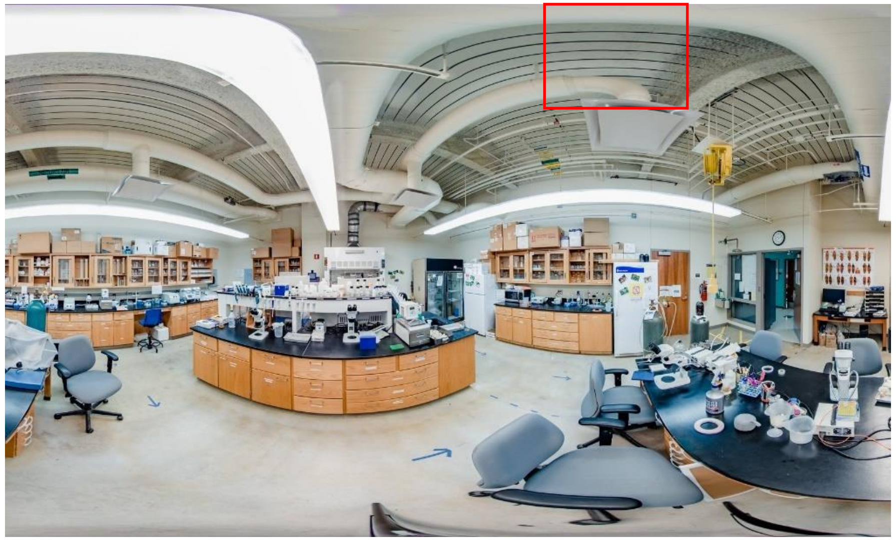
中山大學 SUN YAT-SEN UNIVERSITY

	Dataset	ODI-SR						SUN360					
	Scale	×4		×8		×16		×4		×8		×16	
	Method	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM	WS-PSNR	WS-SSIM
SISR	Bicubic	24.62	0.6555	19.64	0.5908	17.12	0.4332	24.61	0.6459	19.72	0.5403	17.56	0.4638
	SRCNN	25.02	0.6904	20.08	0.6112	18.08	0.4501	26.30	0.7012	19.46	0.5701	17.95	0.4684
	VDSR	25.92	0.7009	21.19	0.6334	19.22	0.5903	26.36	0.7057	21.60	0.6091	18.91	0.5935
	LapSRN	25.87	0.6945	20.72	0.6214	18.45	0.5161	26.31	0.7000	20.05	0.5998	18.46	0.5068
	MemNet	25.39	0.6967	21.73	0.6284	20.03	0.6015	25.69	0.6999	21.08	0.6015	19.88	0.5759
	MSRN	25.51	0.7003	23.34	0.6496	21.73	0.6115	25.91	0.7051	23.19	0.6477	21.18	0.5996
	EDSR	25.69	0.6954	23.97	0.6483	22.24	0.6090	26.18	0.7012	23.79	0.6472	21.83	0.5974
	D-DBPN	25.50	0.6932	24.15	0.6573	22.43	0.6059	25.92	0.6987	23.70	0.6421	21.98	0.5958
	RCAN	26.23	0.6995	24.26	0.6554	22.49	0.6176	26.61	0.7065	23.88	0.6542	21.86	0.5938
	DRN	26.24	0.6996	24.32	0.6571	22.52	0.6212	26.65	0.7079	24.25	0.6602	22.11	0.6092
ODISR	360-SS	25.98	0.6973	21.65	0.6417	19.65	0.5431	26.38	0.7015	21.48	0.6352	19.62	0.5308
	LAU-Net	26.34	0.7052	24.36	0.6602	22.52	0.6284	26.48	0.7062	24.24	0.6708	22.05	0.6058
	SphereSR	—	—	24.37	0.6777	22.51	0.6370	—	—	24.17	0.6820	21.95	0.6342
	OSRT	26.89	0.7581	24.53	0.6780	22.69	0.6261	27.47	0.7985	24.38	0.7072	22.13	0.6388
	BPOSR	**26.95**	**0.7598**	**24.61**	**0.6782**	**22.72**	**0.6285**	**27.59**	**0.7997**	**24.47**	**0.7084**	**22.16**	**0.6433**

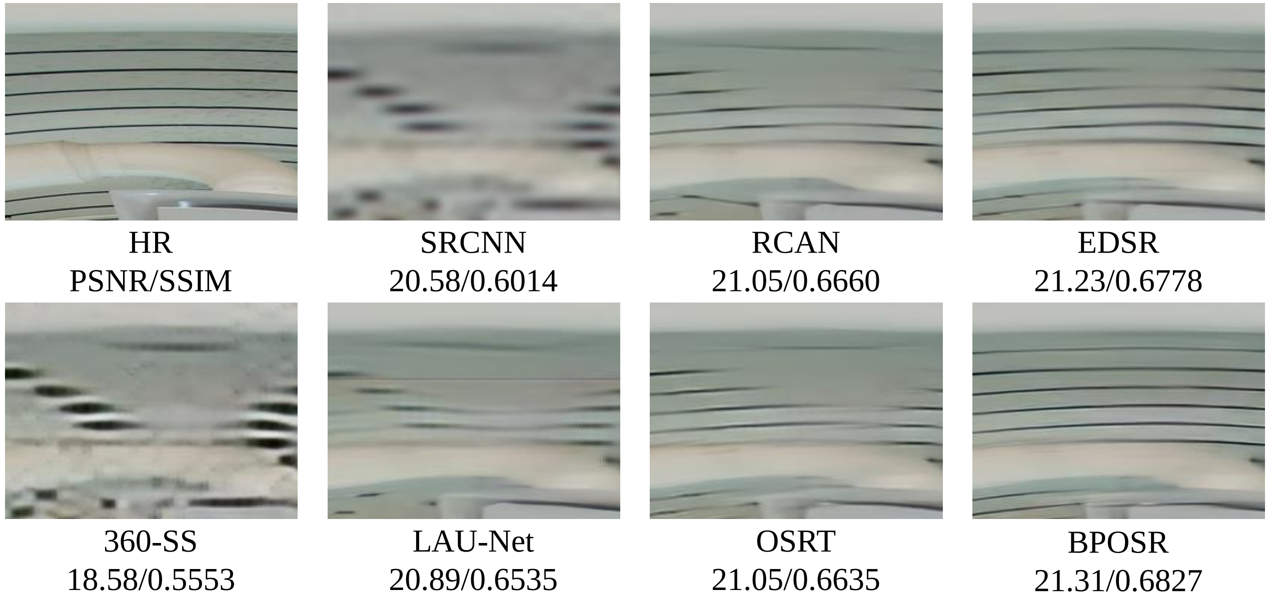
Quantitative comparisons (WS-PSNR/WS-SSIM) with SISR and ODISR algorithms on benchmark datasets. The best results are highlighted in **bold.**

[1] Deng, X.; Wang, H.; Xu, M.; Guo, Y.; Song, Y.; and Yang, L. 2021. **LAU-Net**: Latitude Adaptive Upscaling Network for Omnidirectional Image Super-Resolution. CVPR 2021, 9189–9198.
[2] Yoon, Y.; Chung, I.; Wang, L.; and Yoon, K.-J. **SphereSR**: 360deg Image Super-Resolution With Arbitrary Projection via Continuous Spherical Image Representation. CVPR 2022, 5677–5686.
[3] Yu, F.; Wang, X.; Cao, M.; Li, G.; Shan, Y.; and Dong, C. 2023. **OSRT**: Omnidirectional Image Super-Resolution With Distortion-Aware Transformer. CVPR 2023, 13283–13292

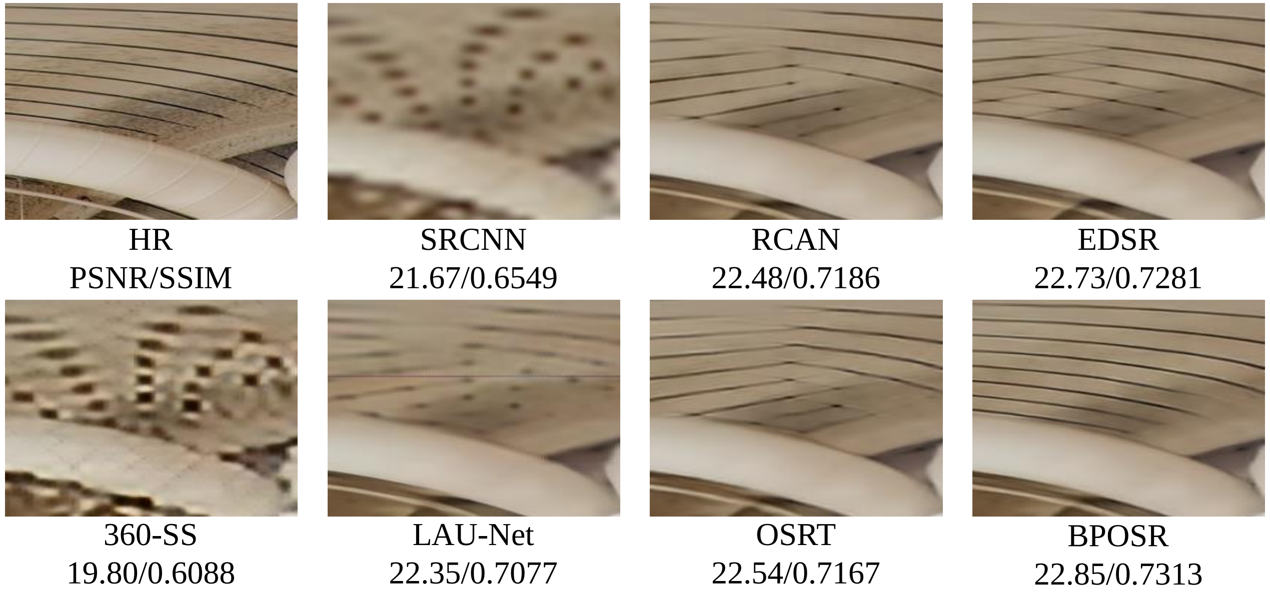
5. Experiments and Results



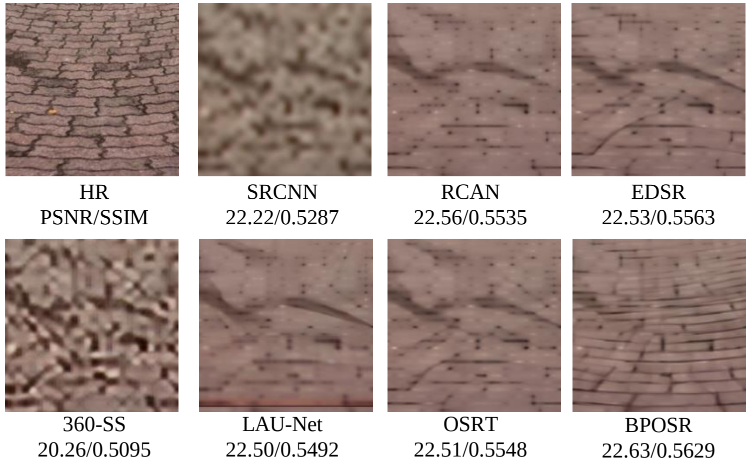
SUN360 (× 8): 060



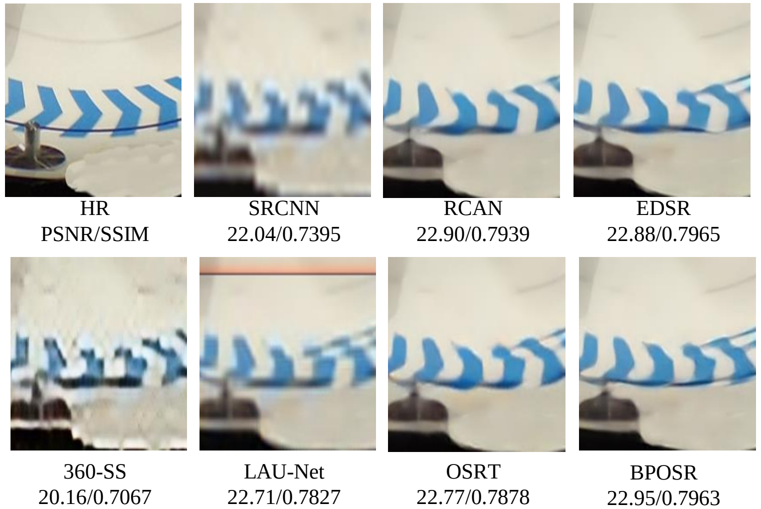
SUN360 (× 8): 049



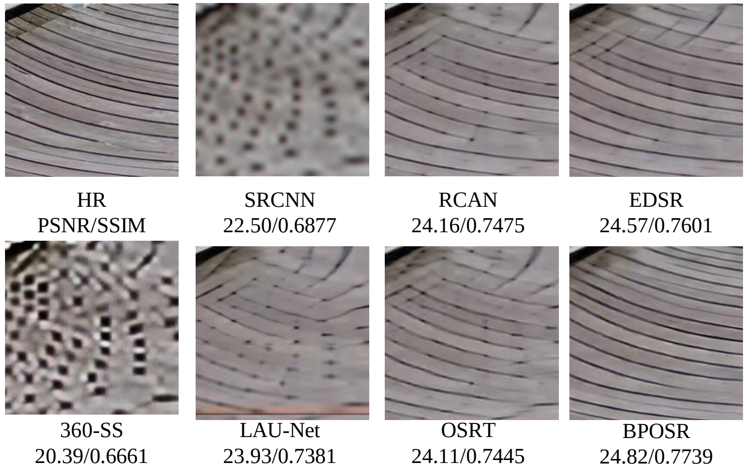
(a) ODI-SR (× 8): 046



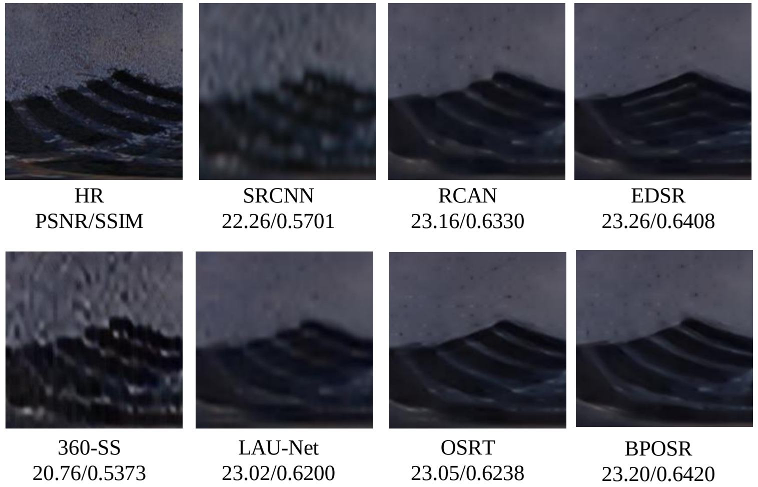
(c) ODI-SR (× 8): 005



(b) ODI-SR (× 8): 064



(d) ODI-SR (× 8): 091



6. Ablation Study

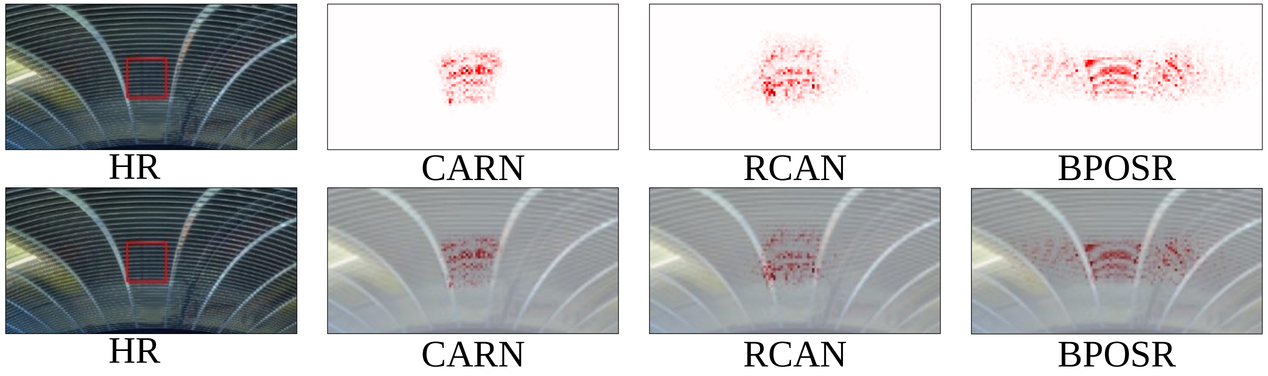


Figure: Local Attribution Maps (LAM) results for different networks. The LAM attribution reflects the importance of each pixel in the input LR image when reconstructing the patch marked with a box.

Method	WS-PSNR	WS-SSIM
BPOSR	24.61	0.6782
Variant-CMP	24.30	0.6620
Variant-ERP	24.47	0.6716

Table: Ablation studies for Bi-Projection

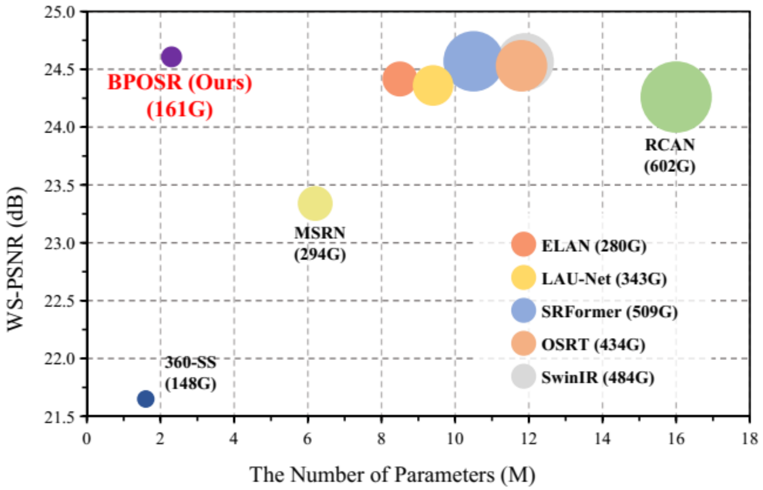


Figure: WS-PSNR vs. the number of parameters. The circle size indicates MACs.

Method	Venue	MACs	Params	WS-PSNR
LapSRN	CVPR'17	23.0G	1.3M	20.72
EDSR	CVPRW'17	2894.5G	45.5M	23.97
MSRN	ECCV'18	294.4G	6.2M	23.34
RCAN	ECCV'18	602.0G	16.0M	24.26
360-SS	MMSP'19	148.2G	1.6M	21.65
SwinIR	ICCVW'21	484.4G	11.9M	24.56
LAU-Net	CVPR'21	342.8G	9.4M	24.36
ELAN	ECCV'22	279.6G	8.5M	24.42
SRFormer	ICCV'23	509.8G	10.5M	24.57
OSRT	CVPR'23	434.9G	11.8M	24.53
BPOSR	-	160.7G	2.3M	24.61

Table: Numerical comparisons with other state-of-the-art algorithms in terms of complexity, parameters, and accuracy.

Thanks