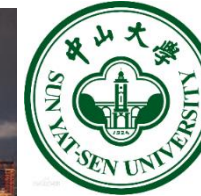




The 38th Annual AAAI Conference on
Artificial Intelligence

FEBRUARY 20-27, 2024 | VANCOUVER, CANADA
VANCOUVER CONVENTION CENTRE – WEST BUILDING



Omnidirectional Image Super-Resolution via Bi-Projection Fusion

Jiangang Wang¹, Yuning Cui², Yawen Li³, Wenqi Ren^{1*}, Xiaochun Cao¹

¹Shenzhen Campus of Sun Yat-sen University

²Technical University of Munich

³Beijing University of Posts and Telecommunications

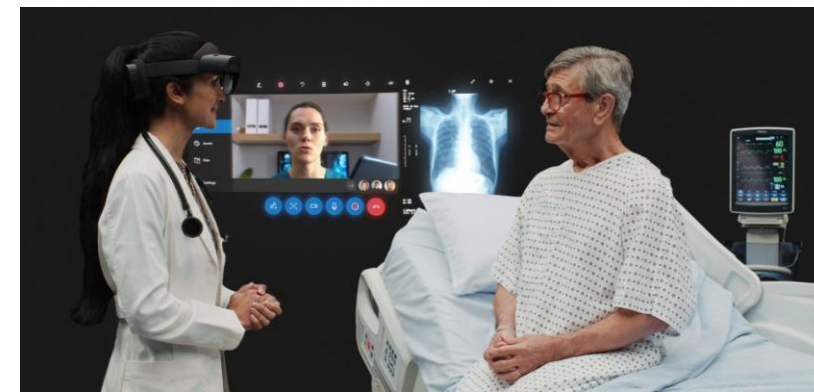
2023年12月23日
深圳



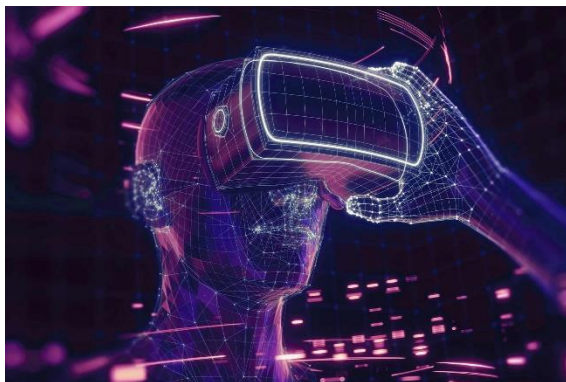
一、研究背景与意义



..... 推动三维图形生成、动态环境建模、实时动作捕捉、快速渲染处理等技术创新，发展虚拟现实整机，感知交互、内容采集制作等设备和开发工具软件、行业解决方案。



医疗



元宇宙



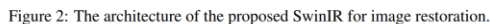
工业 教育

全景视频实现沉浸式的体验，需要超高分辨(8K-16K)的画质

全景视频分析与处理技术已成为多媒体处理领域的研究热点

The diagram illustrates the proposed architecture, which is a deep learning model for image segmentation. It consists of several stages of feature extraction and refinement. The input image is processed by a series of layers: a Convolutional layer (Conv), a Residual Block (ResBlock), another Convolutional layer (Conv), and an Upsample layer. The output of the Upsample layer is then processed by a Convolutional layer (Conv). The architecture also includes a skip connection from the input to the output of the ResBlock. The final output is a segmented image. The diagram uses color-coded blocks to represent different layers: orange for Conv, dark blue for ResBlock, light blue for Upsample, and green for Mult. The flow is indicated by arrows, showing the progression from input to output.

EDSR



SwinIR



VS



传统2D图像



全景图像



二. 相关工作与挑战问题



中山大学
SUN YAT-SEN UNIVERSITY

• Omnidirectional Image Super-Resolution

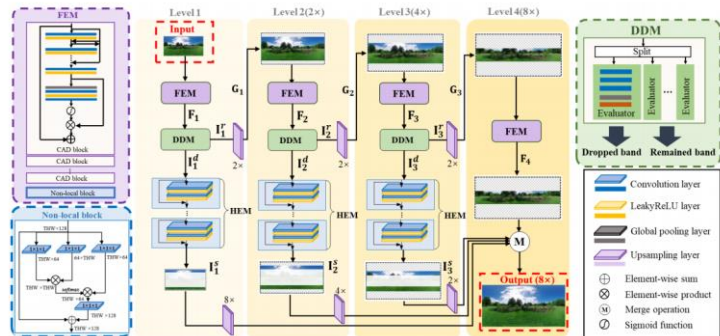


Fig. 6. The progressive architecture of the proposed LAU-Net+. Each level is composed of feature enhancement module (FEM), drop-band decision module (DDM) and high-latitude enhancement module (HEM). The final HR image is obtained by merging the outputs from different levels.

LAU-Net

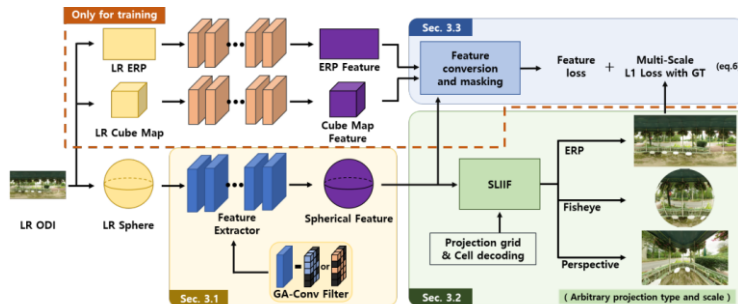


Figure 2. Overall framework of the proposed SphereSR.

SphereSR

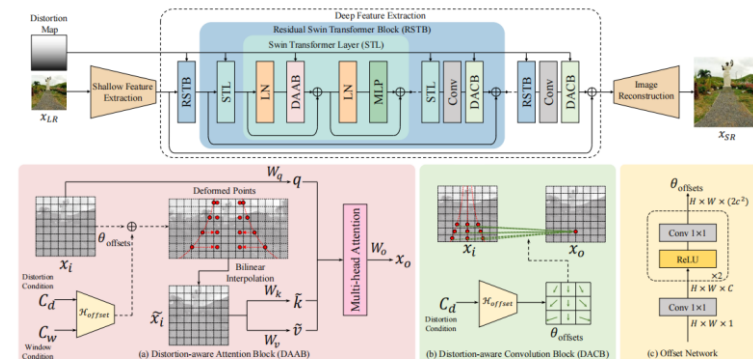
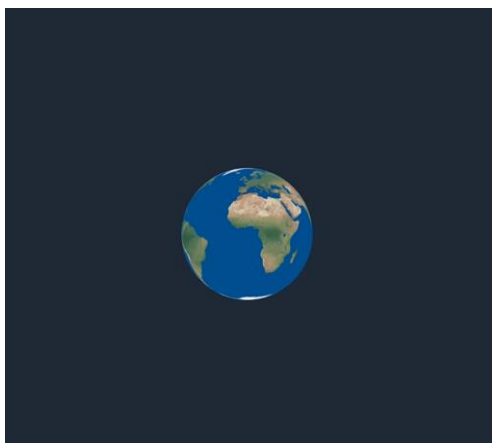


Figure 4. Overall illustration of OSRT. From SwinIR [24], we replace the standard multi-head self-attention block with DAAB and insert DACB behind the end of the RSTB. Channel dimensions of $\theta_{offsets}$ in DAAB and DACB are 2 and 18, respectively.

OSRT

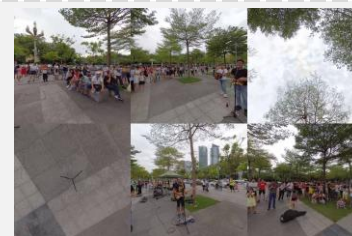
现有全景图像超分辨率仅使用ERP投影，全景图像特征利用不充分



地图投影



ERP



CMP



EAC



TSP



ISP



OHP

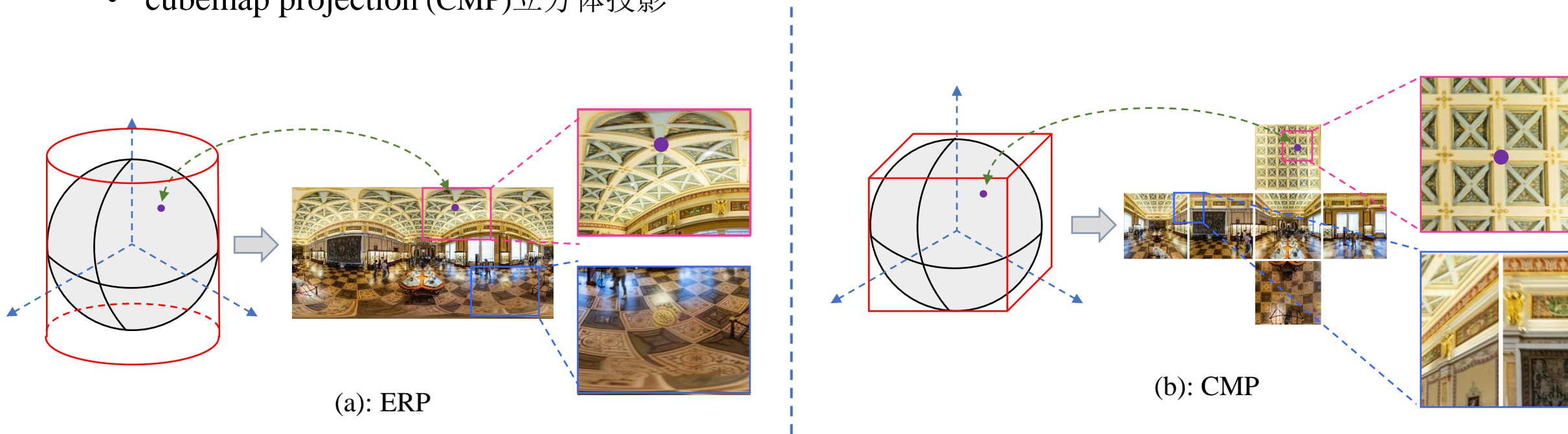


SSP

不同全景投影格式

三、全景图像投影的特征分析

- 全景图像最常用的两种投影：
 - equirectangular projection (ERP) 圆柱形投影
 - cubemap projection (CMP) 立方体投影

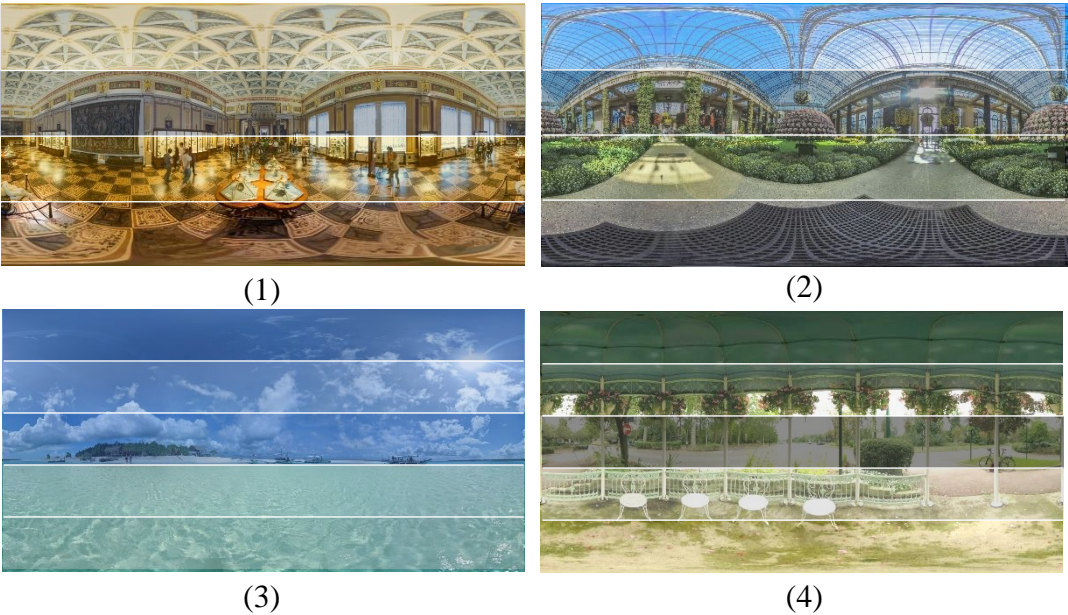


ERP和CMP投影具有互补性

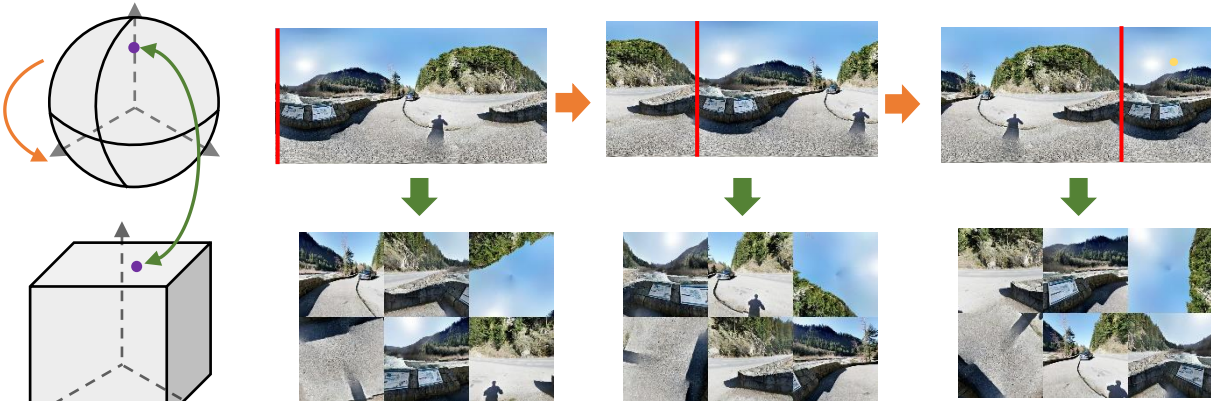
投影类型	优点	缺点
ERP	全局视角	画面扭曲失真大，特别是高纬度地区
CMP	扭曲失真小	画面只有局部视角、画面边缘不连续

三、全景图像投影的特征分析

• ERP 和 CMP 几何特征挖掘



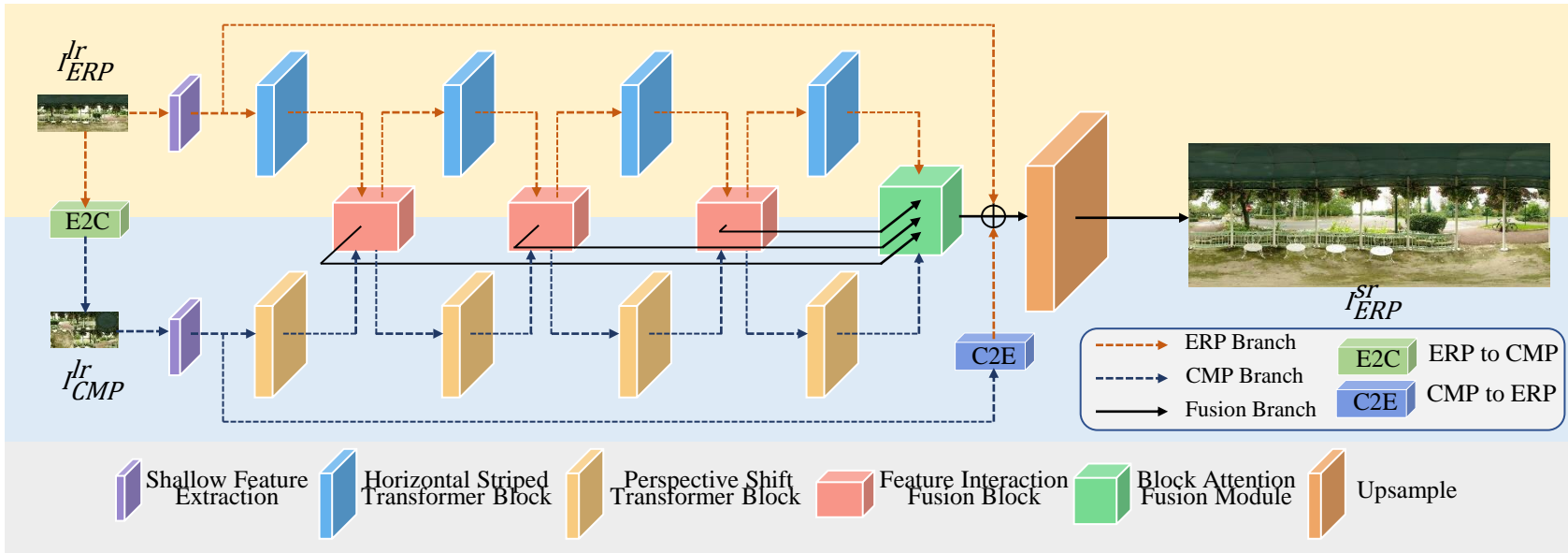
(a): ERP Horizontal Similarity

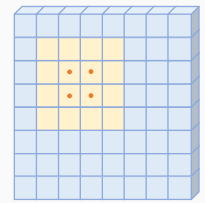
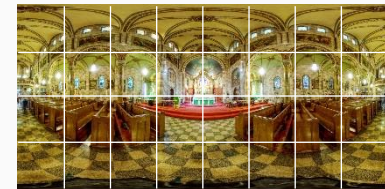


(b): CMP Perspective Variability

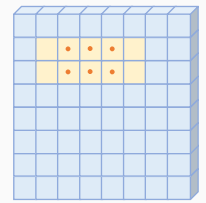
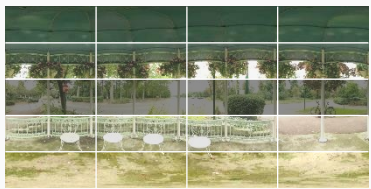
投影类型	几何特征	模型结构设计
ERP	水平相似性	局部相似性建模
CMP	视角多变性	多视角信息融合

四、融合全景投影几何特征的模型设计





(a) Square Windows



(b) Horizontal Striped Windows

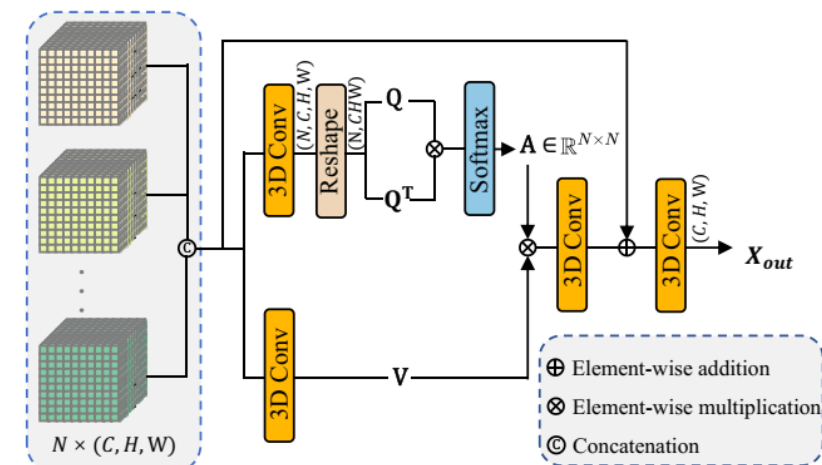
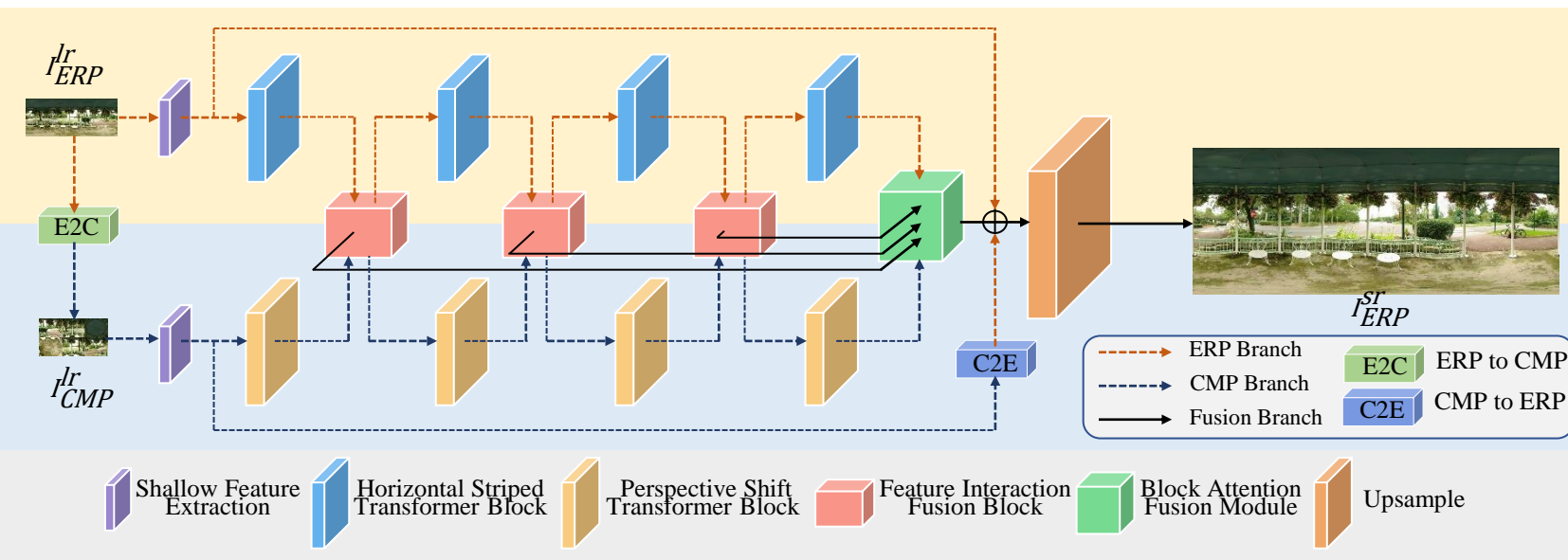
Different self-attention windows

(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.

The overall diagram illustrates the architecture of BPOSR

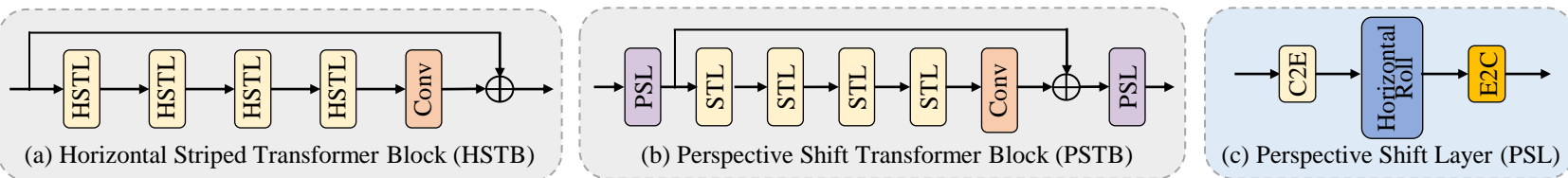
投影类型	几何特征	模型结构设计
ERP	水平相似性	局部相似性建模
CMP	视角多变性	多视角信息融合

四、融合全景投影几何特征的模型设计



Block Attention Fusion Module

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



The overall diagram illustrates the architecture of BPOSR

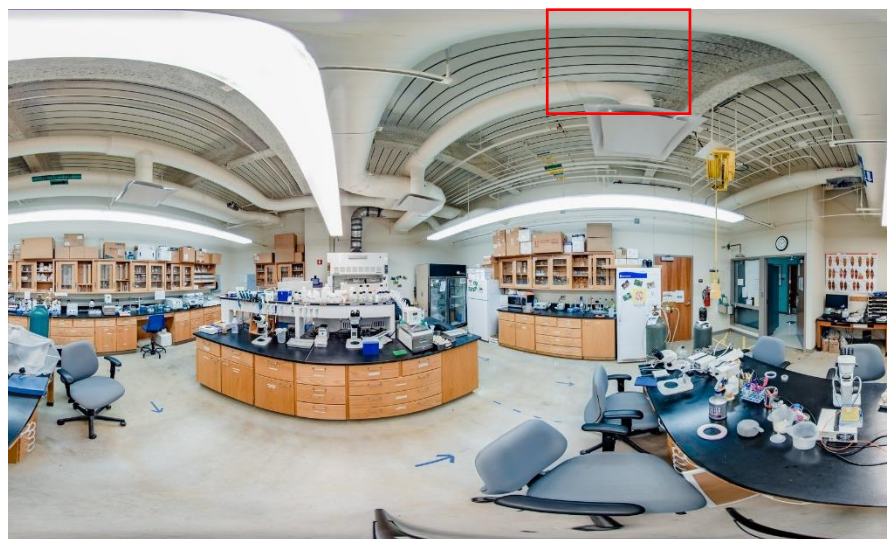
五、实验与结果

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

五、实验与结果



SUN360 ($\times 8$): 060



HR
PSNR/SSIM



SRCNN
20.58/0.6014



RCAN
21.05/0.6660



EDSR
21.23/0.6778



360-SS
18.58/0.5553



LAU-Net
20.89/0.6535



OSRT
21.05/0.6635



BPOSr
21.31/0.6827



SUN360 ($\times 8$): 049



HR
PSNR/SSIM



SRCNN
21.67/0.6549



RCAN
22.48/0.7186



EDSR
22.73/0.7281



360-SS
19.80/0.6088



LAU-Net
22.35/0.7077



OSRT
22.54/0.7167

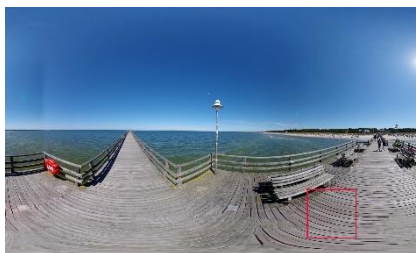
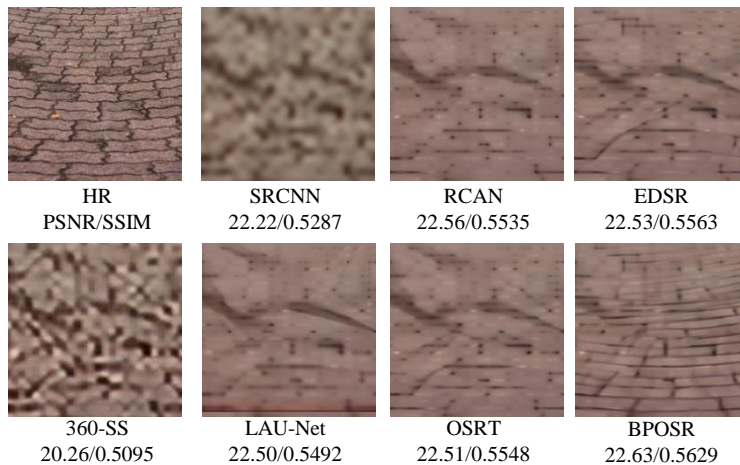


BPOSr
22.85/0.7313

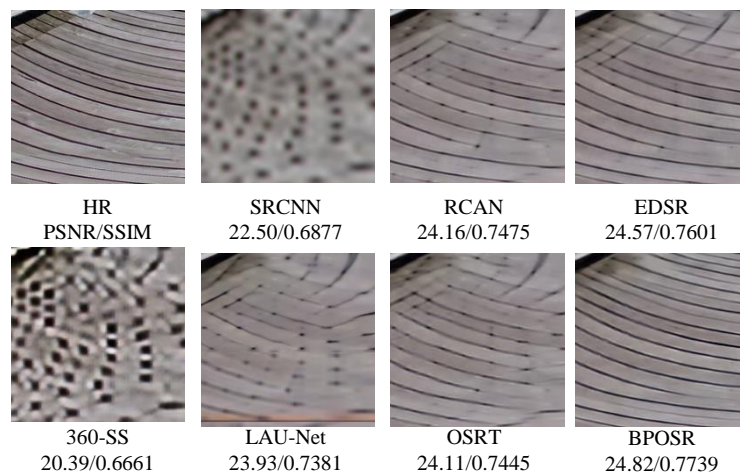
五、实验与结果



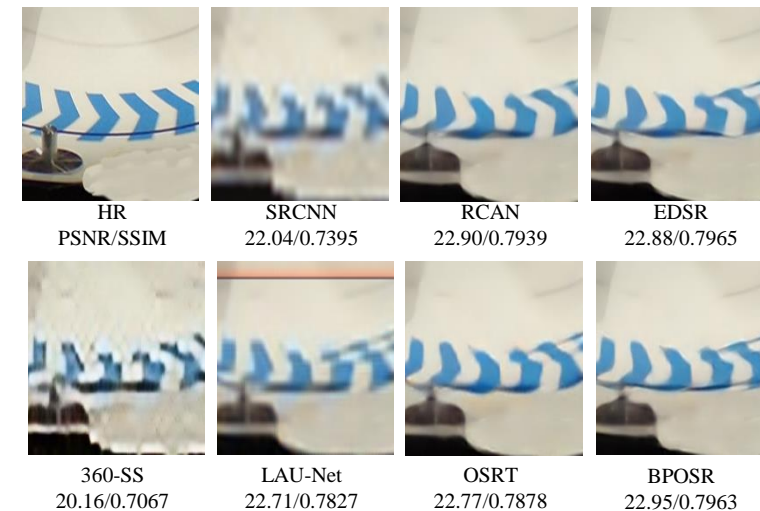
(a) ODI-SR ($\times 8$): 046



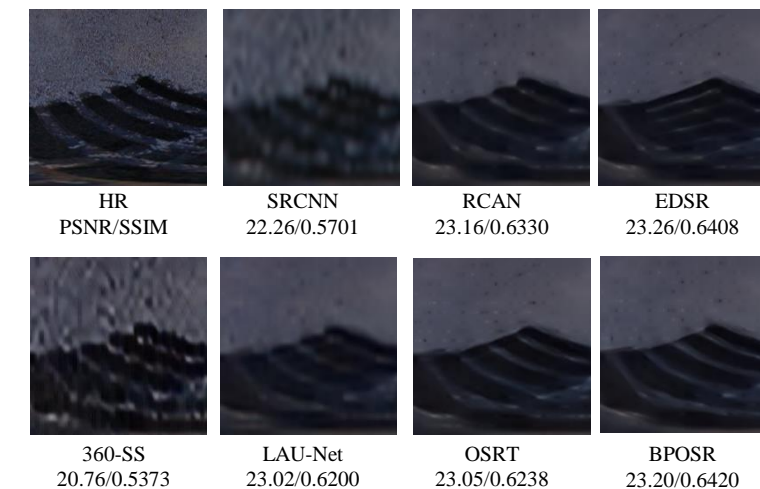
(b) ODI-SR ($\times 8$): 064



(c) ODI-SR ($\times 8$): 005



(d) ODI-SR ($\times 8$): 091



六、消融实验

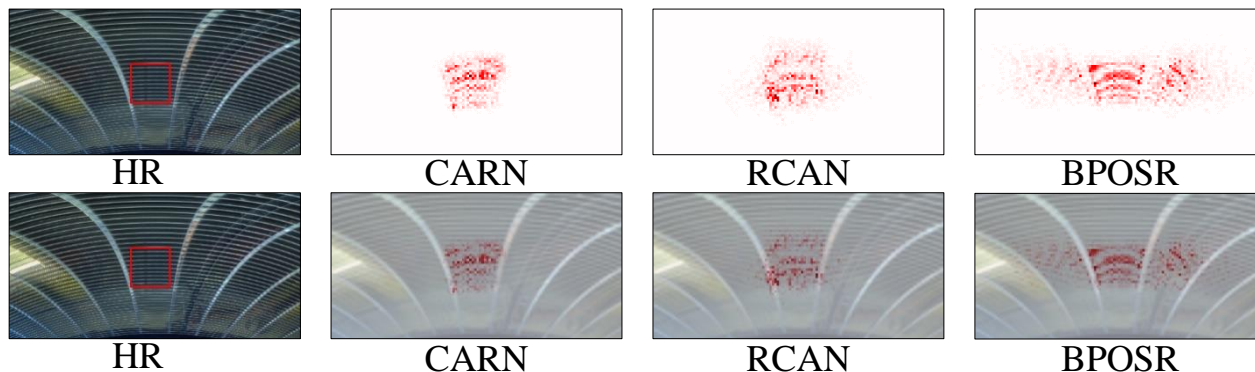


Figure: Local Attribution Maps (LAM) results for different networks. The LAM attribution reflects the importance of each pixel in the input LR image when reconstructing the patch marked with a box.

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

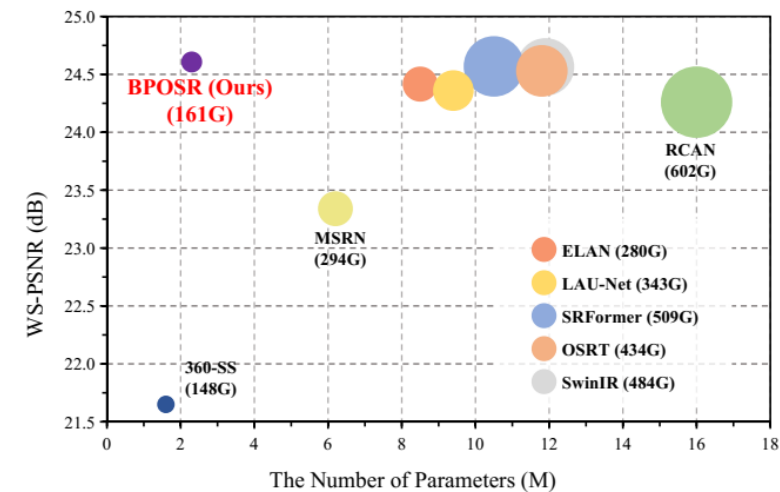


Figure: WS-PSNR vs. the number of parameters. The circle size indicates MACs.

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

