OmniZoom: A Universal Plug-and-Play Paradigm for Cross-Device Smooth Zoom Interpolation

Xiaoan Zhu*
Zhejiang University
xiaoanzhu@zju.edu.cn

Yue Zhao*
Huawei Noah's Ark Lab
zhaoyue53@huawei.com

Tianyang Hu
Zhejiang University
huty@zju.edu.cn

Jiaming Guo
Huawei Noah's Ark Lab
guojiaming5@huawei.com

Yulan Zeng
Sungkyunkwan University
zengyulan@skku.edu

Renjing Pei
Huawei Noah's Ark Lab
peirenjing@huawei.com

Fenglong Song
Huawei Noah's Ark Lab
songfenglong@huawei.com

Huajun Feng†
Zhejiang University
fenghj@zju.edu.cn

Introduction -- Zoom interpolation (ZI)

Main camera image

Wide-angle camera image

Merged camera image

➤**Key Challenges of Zoom Interpolation (ZI) vs. Frame Interpolation (FI)**

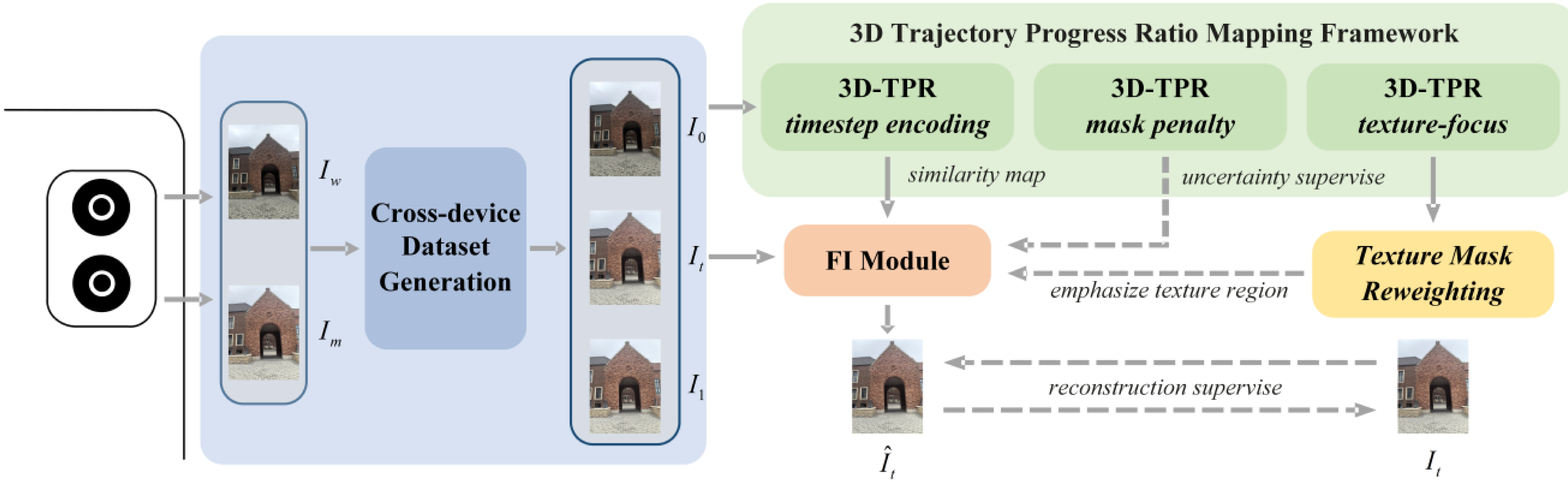
- **Lack of Dense Supervision**

  ☐**Core Issue:** Absence of suitable ground-truth intermediate frames for training.

- **Inherent Motion Ambiguity**

  ☐**Core Issue:** Severe motion ambiguity that violates the linear motion assumption in FI.

Pipeline -- OmniZoom

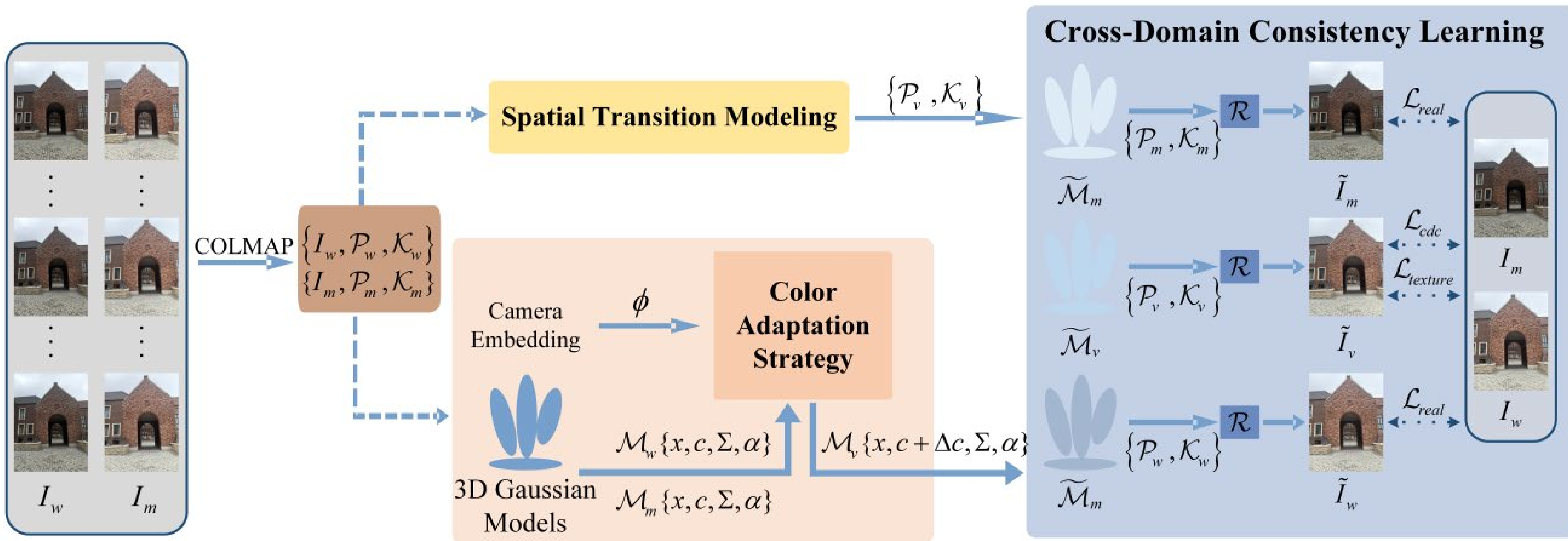


OmniZoom: Plug-and-play integration pipeline for crossdevice ZI.

➢ **Key Contributions:**

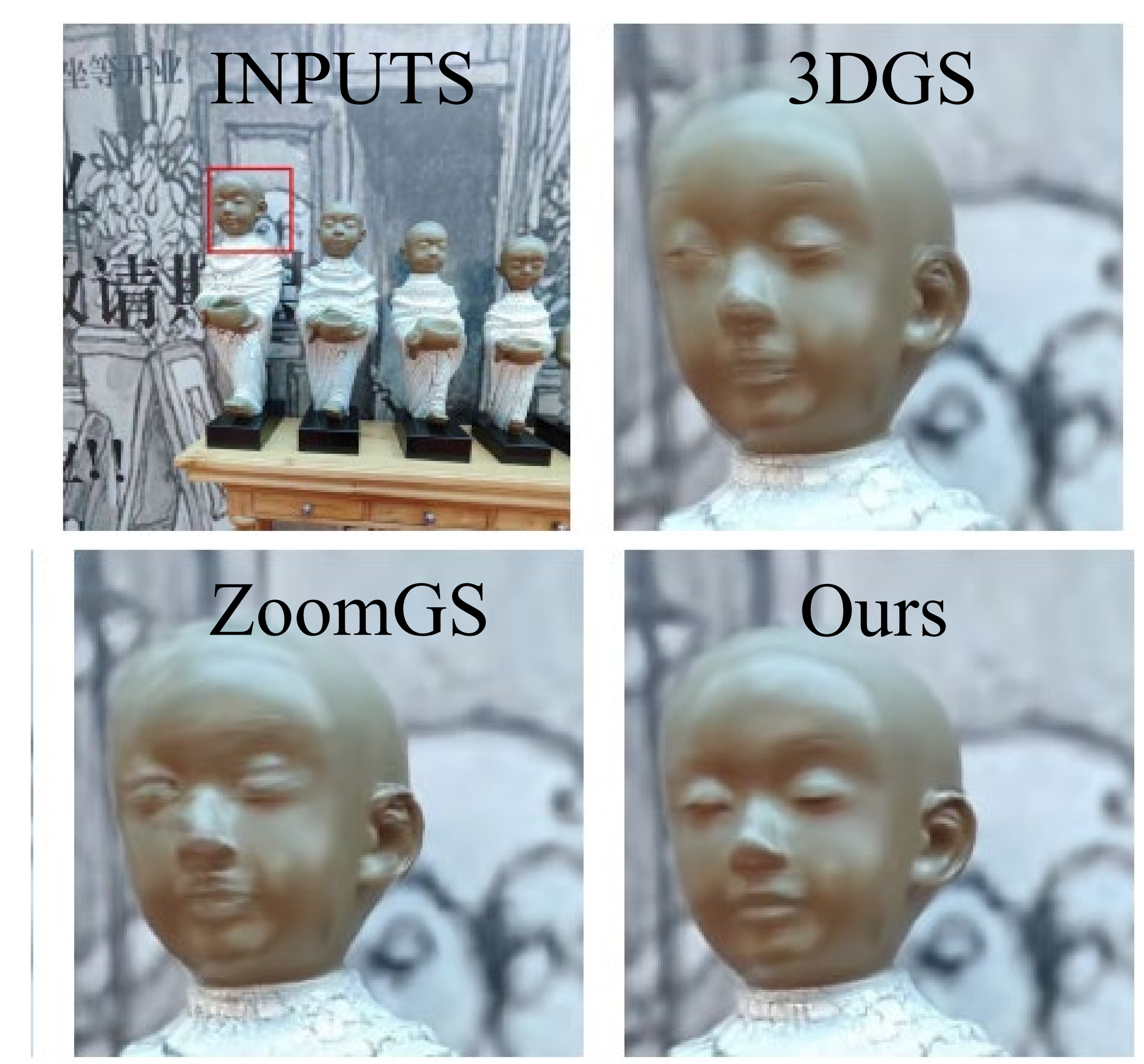
- **A novel ZI dataset pipeline** that models cross-device inconsistencies to enable high-quality, device-agnostic supervision.
- **The 3D-TPR framework** that leverages disparity-aware encoding, texture focusing, and a mask penalty for superior perceptual quality.
- **OmniZoom, a universal paradigm** that seamlessly integrates with FI networks for robust, high-quality zoom interpolation across diverse devices.

Method -- Cross-device Data Generation



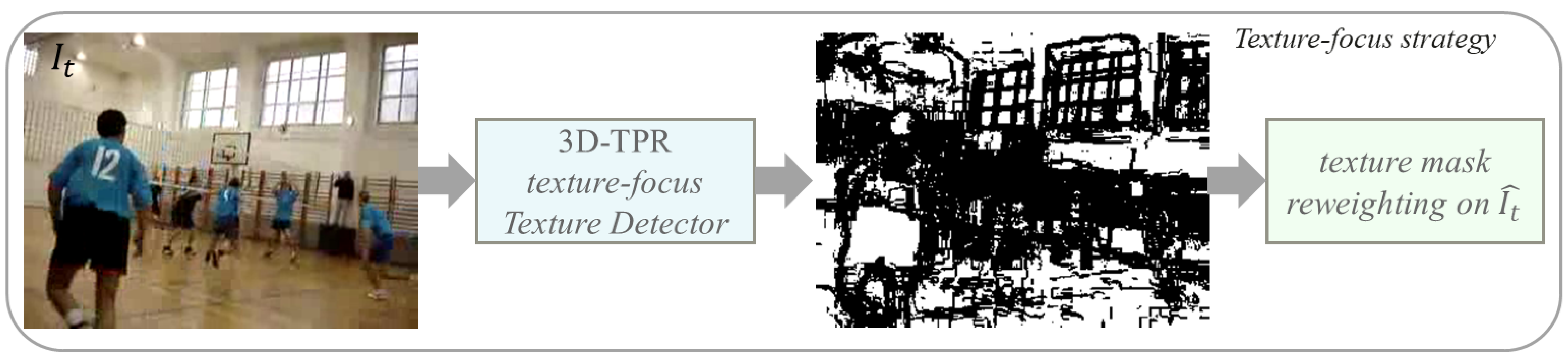
Dual-camera ZI dataset generation via spatial-color calibration and cross-domain optimization.

Sample triplets from our ZI dataset, each showing the wide-angle I_w, the synthetic intermediate frame I_v, and the main camera image I_m, for two devices: Huawei and Redmi.



Our data generation method compared to 3DGS and ZoomGS

Method -- 3D-TPR Framework



- ➢ 3D Similarity for Timestep Input
- ➢ Uncertainty Suppression for Blur Reduction
- ➢ Gradient Reprojection for Edge Enhancement

Results

Quantitative results

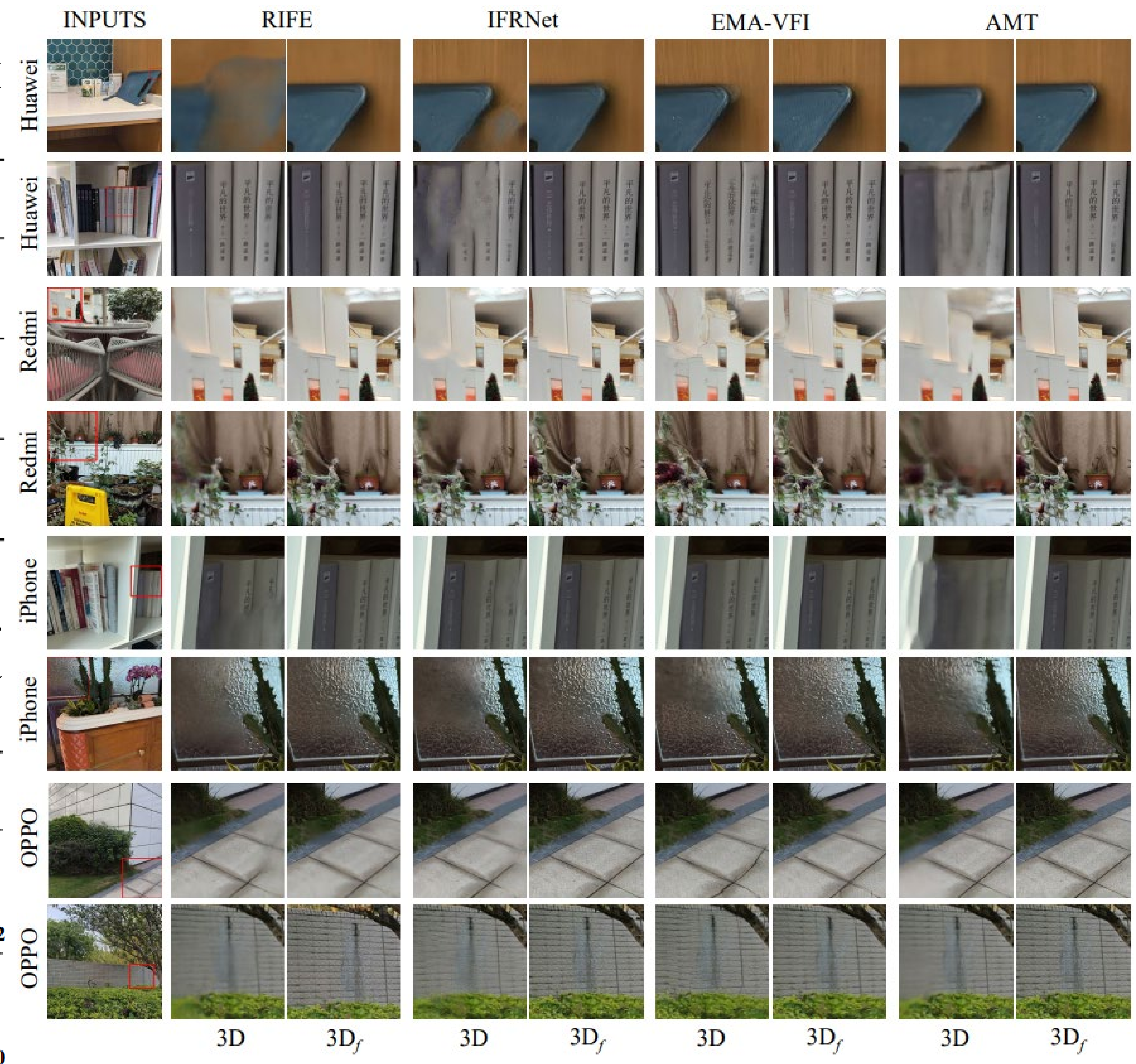
Comparisons on Vimeo90k, Inter4K and UCF101 using 2D and 3D-TPR FI frameworks. **Bold** indicates the best metric.

Benchmarks	Metrics	RIFE		IFRNet		EMA-VFI		AMT	
		2D	3D-TPR	2D	3D-TPR	2D	3D-TPR	2D	3D-TPR
Vimeo90k	PSNR↑	27.40	**27.51**	27.13	**27.21**	24.73	**24.86**	27.17	**27.22**
	SSIM↑	0.901	**0.902**	0.899	**0.901**	0.851	**0.853**	0.902	**0.902**
	LPIPS↓	0.086	**0.081**	0.078	**0.074**	0.081	**0.080**	0.081	0.084
inter4K	PSNR↑	33.92	**34.06**	**33.73**	33.72	30.06	**30.29**	33.57	**33.80**
	SSIM↑	0.951	**0.952**	0.952	**0.952**	0.903	**0.904**	0.953	**0.955**
	LPIPS↓	0.048	**0.046**	0.046	**0.045**	0.044	**0.044**	0.048	**0.047**
UCF101	PSNR↑	35.54	**35.85**	35.42	**35.43**	**36.74**	36.65	**35.37**	35.33
	SSIM↑	0.928	**0.983**	0.984	**0.986**	0.984	**0.985**	0.984	**0.984**
	LPIPS↓	**0.017**	0.018	0.017	**0.017**	0.012	**0.012**	0.018	**0.017**

Comparisons on real-world data across four FI models. Subscript $_f$ denotes models finetuned on our ZI dataset. **Bold** indicates the best performance, and underline is the second best.

Device	Metrics	RIFE			IFRNet			EMA-VFI			AMT		
		2D	3D	$3D_f$	2D	3D	$3D_f$	2D	3D	$3D_f$	2D	3D	$3D_f$
Huawei	NIQE↓	3.8464	3.8651	**3.7422**	3.6801	**3.4852**	3.5035	3.6363	3.8470	**3.8341**	3.5665	3.6459	**3.4547**
	PI↓	4.2505	4.1537	**3.9360**	3.9570	3.3150	**3.2520**	**3.7240**	3.8566	3.8467	3.5147	3.6132	**3.3121**
	CLIP-IQA↑	0.3691	0.4939	**0.5233**	0.5422	0.5784	**0.5909**	0.5612	0.5621	**0.5684**	0.5428	0.5498	**0.6018**
	MUSIQ↑	44.8268	58.8362	**60.8585**	57.4632	73.0233	**74.1220**	61.7263	62.7403	**63.5048**	71.4369	71.7016	**73.7092**
iPhone	NIQE↓	4.2031	4.1027	**3.8601**	3.8786	**3.6923**	3.6953	**3.8155**	4.0555	4.0934	3.5932	3.7053	**3.5087**
	PI↓	4.5340	4.3503	**4.0657**	4.2090	3.3942	**3.3165**	**3.9994**	4.1304	4.1123	3.4293	3.5367	**3.2661**
	CLIP-IQA↑	0.4821	0.5366	**0.5492**	0.5240	0.5829	**0.5949**	0.5352	0.5387	**0.5503**	0.5931	0.5930	**0.6211**
	MUSIQ↑	55.0577	59.6800	**60.7988**	57.4088	73.3569	**73.8292**	58.1422	58.6204	**60.1002**	71.9805	72.1243	**73.6150**
OPPO	NIQE↓	4.6364	4.6100	**4.3968**	**4.5820**	5.2039	5.1133	4.5568	4.5166	**4.5071**	4.9042	5.3017	**4.6699**
	PI↓	5.0848	5.0104	**4.6815**	4.9749	4.9603	**4.7144**	4.9569	4.9022	**4.8752**	4.9676	5.3808	**4.5702**
	CLIP-IQA↑	0.4989	0.5363	**0.5830**	**0.5525**	0.5212	0.5461	**0.5622**	0.5590	0.5591	0.4798	0.4518	**0.5595**
	MUSIQ↑	63.2645	65.7912	**67.1494**	67.2635	75.2718	**75.4545**	67.4327	67.5921	**67.6444**	73.0842	72.1563	**75.2187**
Redmi	NIQE↓	5.0138	4.8764	**4.4720**	5.0098	4.2837	**4.0695**	4.4088	4.3424	**3.8852**	5.1925	4.7426	**3.9835**
	PI↓	5.3615	5.0247	**4.5195**	5.1108	3.5993	**3.3880**	4.6708	4.5974	**4.5519**	5.4335	4.0235	**3.3548**
	CLIP-IQA↑	0.4077	0.4664	**0.4930**	0.4219	0.4913	**0.5152**	0.4599	0.4807	**0.4947**	0.4336	0.4754	**0.5383**
	MUSIQ↑	56.3870	61.9235	**63.6990**	57.5453	73.7573	**74.3871**	57.0371	59.6610	**60.7579**	57.6671	71.9982	**74.3325**

Qualitative results



- **Conclusion:** We propose OmniZoom, a universal ZI solution that bridges pre-trained models and real-world devices via a novel dataset and the 3D-TPR framework.
- **Limitation:** Its plug-and-play deployment may require lightweight, model-aware training tweaks for optimal performance.