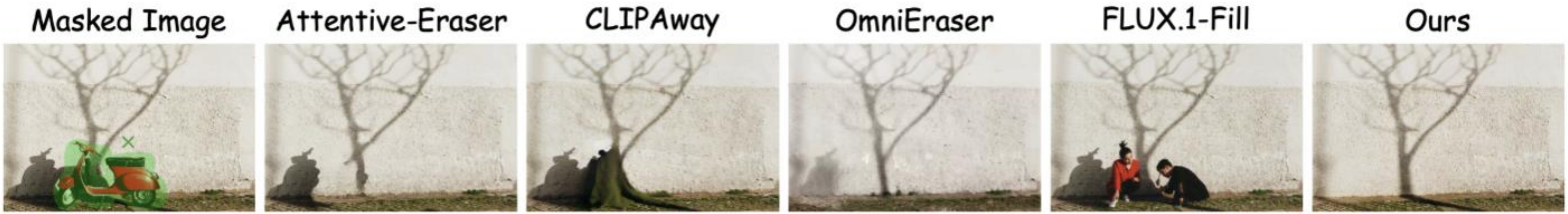
GEOREMOVER: REMOVING OBJECTS AND THEIR CAUSAL VISUAL ARTIFACTS

Zixin Zhu[1,2], Haoxiang Li[2], Xuelu Feng[1], He Wu[2], Chunming Qiao[1], Junsong Yuan[1]

[1]University at Buffalo, [2]Pixocial Technology

Project

Background



Task: Object Removal vs. Inpainting

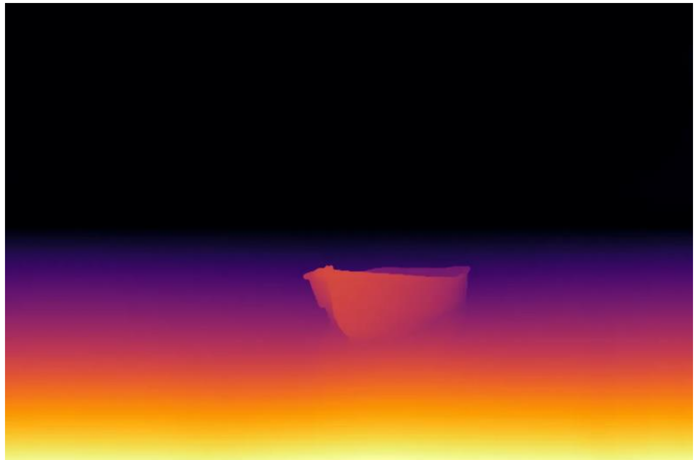
- **Inpainting:** only fills the masked region.
- **Object Removal:** removes a **specific object in the masked region** and also **preserves outside-mask consistency** (no new objects, no lighting or edge mismatches).

Observation

Direct paired supervision blurs what to remove vs. what to preserve.

Motivation

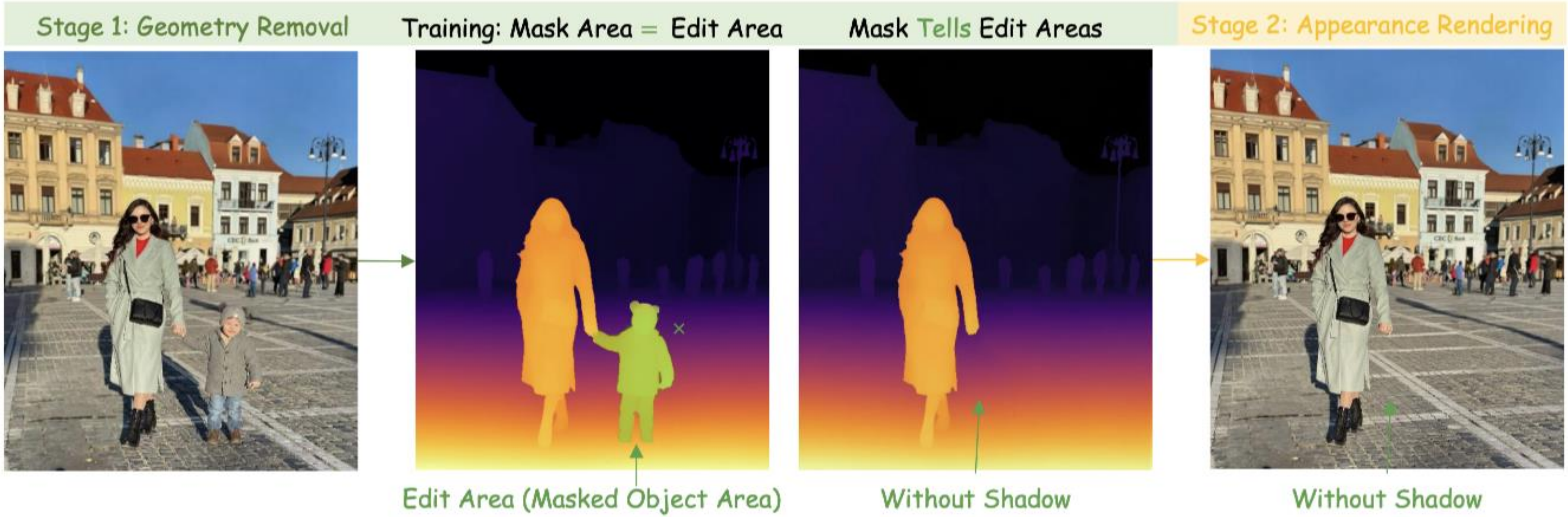
Visual effect (Images)



Geometry (depth maps)

Causal view: object geometry *(cause)* → outside-mask artifact fields *(effect: shadows, reflections, contacts)*.

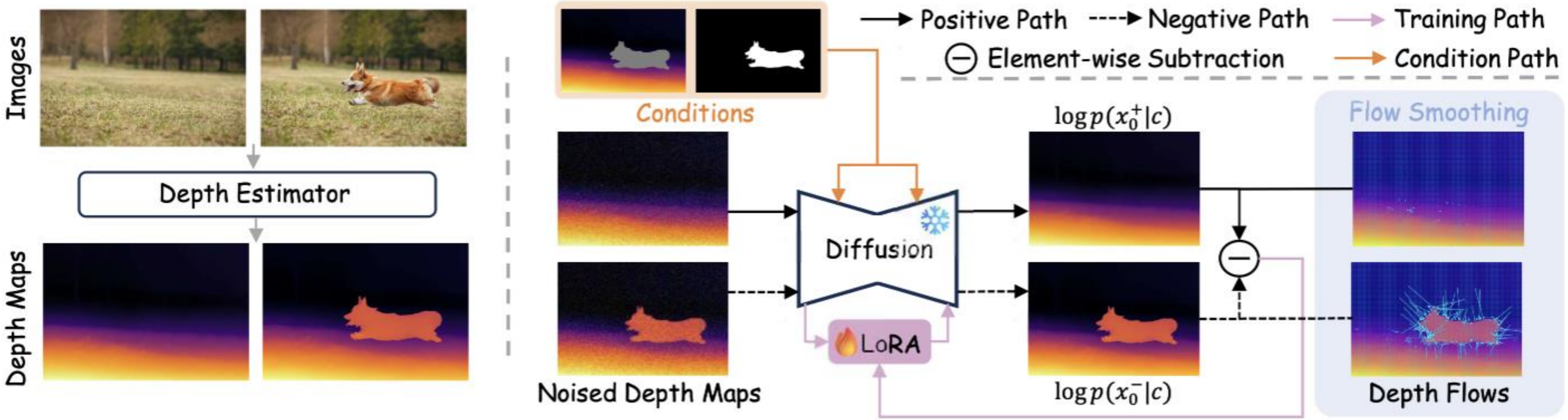
Motivation



Method. We recast object removal as a causal two-stage pipeline: (1) **geometry removal**—modify scene geometry (e.g., depth) to excise the object; (2) **appearance rendering**—re-synthesize the image from the updated geometry so shadows/reflections disappear.

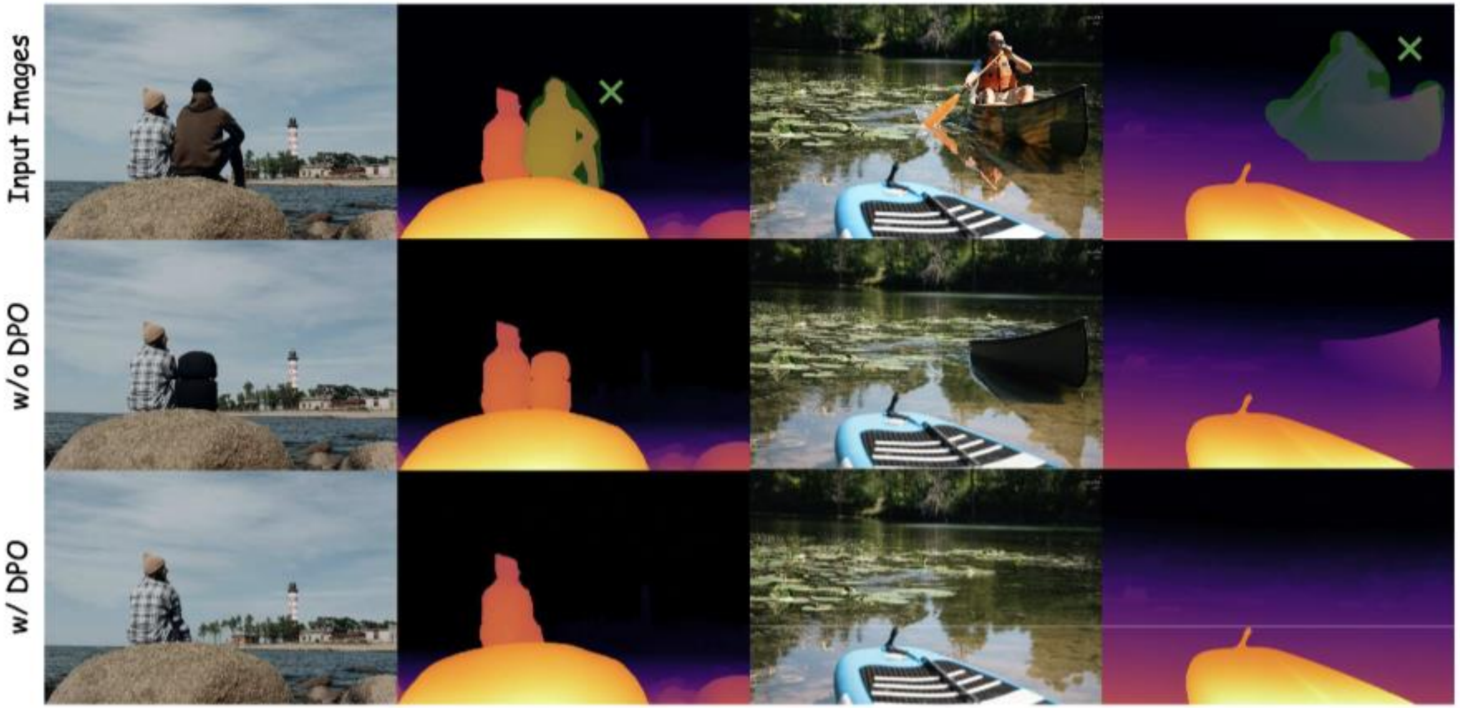
Benefits. The geometry stage supports strictly **mask-aligned, well-posed supervision** (no unintended outside-mask edits), and the rendering stage **naturally removes object-induced artifacts while preserving nearby content**, learned from paired data that links objects to their effects.

Method

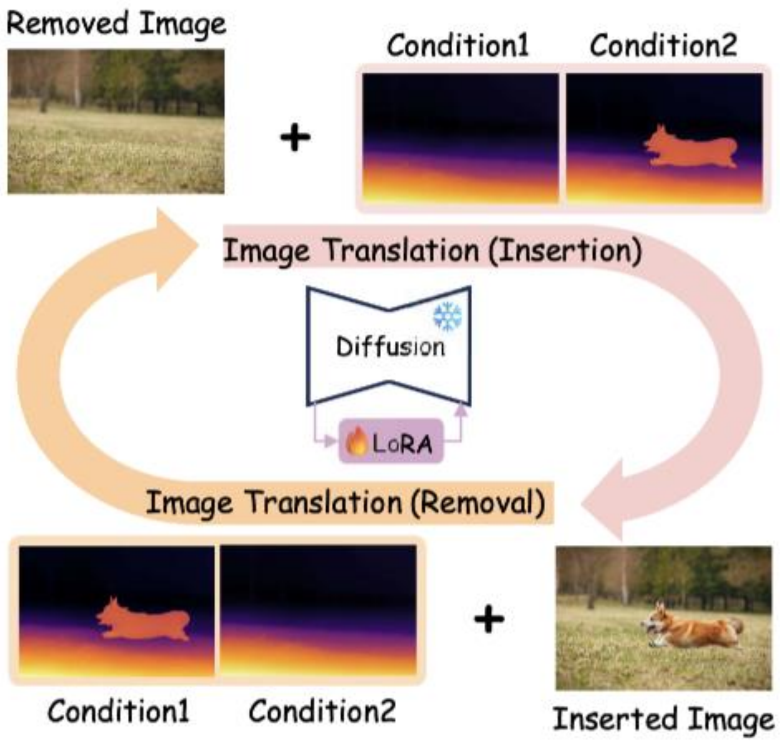


The training framework of Stage 1: Geometry Removal.

Method



Effect of direct preference optimization (DPO) in Stage 1.



Stage 2: Appearance rendering.

Experiments

Table 1: Comparison with state-of-the-art methods on RemovalBench and RORD-Val.

Method	RemovalBench					RORD-Val				
	FID ↓	CMMD ↓	LPIPS ↓	PSNR ↑	AS ↑	FID ↓	CMMD ↓	LPIPS ↓	PSNR ↑	AS ↑
ZITS++ [41]	108.38	0.374	0.158	19.62	4.56	107.44	0.448	0.274	21.17	4.12
MAT [19]	123.78	0.366	0.164	17.88	4.51	136.53	0.455	0.281	19.18	4.38
LaMa [42]	99.88	0.351	0.156	18.72	4.55	100.21	0.294	0.229	20.50	4.23
RePaint [20]	102.65	0.741	0.378	19.86	4.38	114.64	2.345	0.525	17.68	4.71
BLD [43]	128.66	0.553	0.233	17.43	4.39	224.61	0.862	0.273	17.13	4.74
LDM [7]	108.79	0.365	0.157	19.24	4.47	128.19	0.506	0.221	19.02	4.12
SD-Inpaint [7]	119.60	0.419	0.274	17.02	4.48	143.69	0.494	0.308	16.83	4.61
SDXL-Inpaint [7]	104.97	0.398	0.187	17.87	4.63	147.01	0.460	0.210	17.69	4.76
BrushNet [35]	120.97	0.549	0.191	18.68	4.63	234.87	0.745	0.293	16.51	4.41
FLUX.1-Fill [8]	115.79	0.487	0.193	17.12	4.59	141.39	0.450	0.217	18.50	4.55
PowerPaint [44]	114.55	0.392	0.240	18.25	4.56	102.33	0.408	0.241	18.29	4.38
CLIPAway [5]	108.40	0.272	0.254	18.78	4.48	81.28	0.545	0.278	16.36	4.19
Attentive-Eraser [45]	55.49	0.232	0.146	20.60	4.50	96.77	0.233	0.221	20.24	4.77
OmniEraser [9]	39.52	0.208	0.133	21.11	**4.66**	43.71	**0.153**	0.166	22.13	**4.99**
Ours	**29.88**	**0.089**	**0.124**	**25.52**	4.54	**31.15**	0.182	**0.103**	**23.70**	4.69

Table 2: Ablation study on RORD-Val to evaluate the effectiveness of our design components. "Insert." denotes the percentage of cases where a new object is wrongly inserted into the removal region.

Method	FID ↓	CMMD ↓	LPIPS ↓	PSNR ↑	AS ↑	Insert. ↓
One-Stage	56.24	0.577	0.315	17.52	4.27	2.81%
Two-Stage w/o DPO	34.24	0.230	0.131	22.81	4.51	5.09%
Two-Stage w/ DPO	**31.15**	**0.182**	**0.103**	**23.70**	**4.69**	**1.48%**

Table 3: Geometry removal accuracy (MAE in masked region) on RORD-Val.

Method	MAE ↓
Input depth	0.0827
Two-Stage w/o DPO	0.0490
Two-Stage w/ DPO	**0.0387**

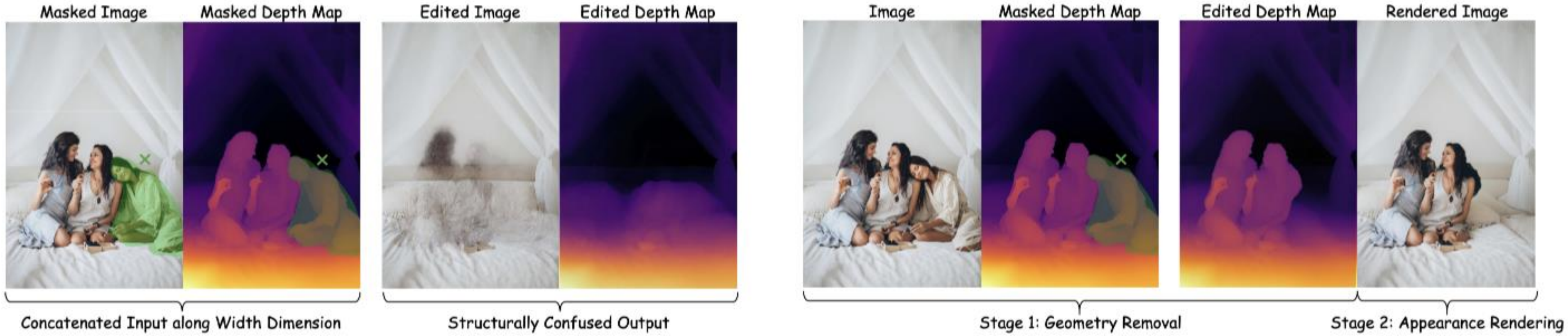
Experiments

Table 4: Removal performance of causal artifacts on CausRem.

Method	IoU% ↑
OmniEraser [9]	68.29
Ours	**73.76**

Table 5: Ablation study on the RORD-Val dataset comparing unidirectional and bidirectional rendering strategies in Stage 2.

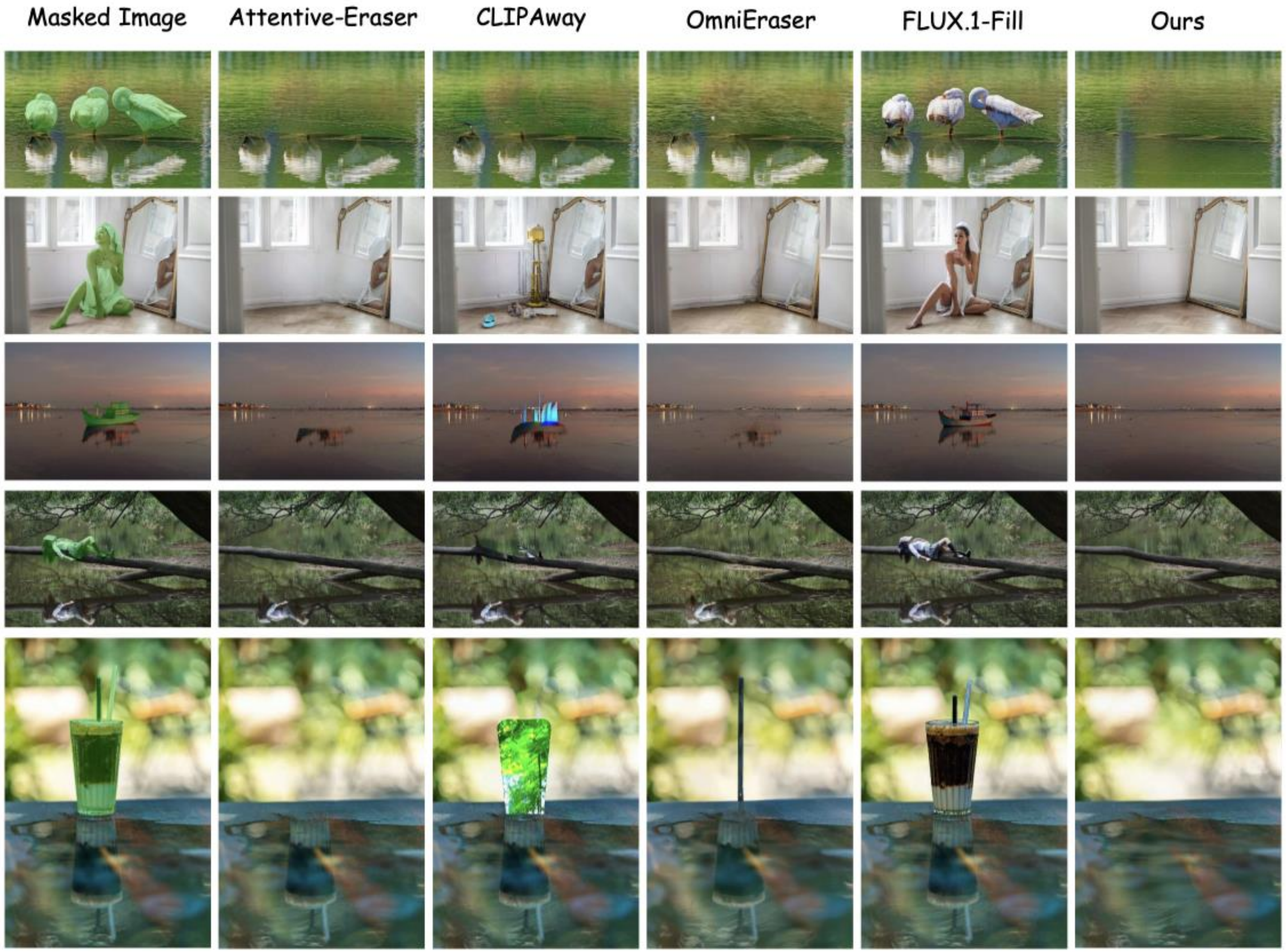
Method	FID ↓	CMMD ↓	LPIPS ↓	PSNR ↑	AS ↑
Unidirectional rendering	38.43	0.215	0.136	23.58	4.19
Bidirectional rendering	**31.15**	**0.182**	**0.103**	**23.70**	**4.69**



(a) Results from our one-stage model. (b) Results from our two-stage model

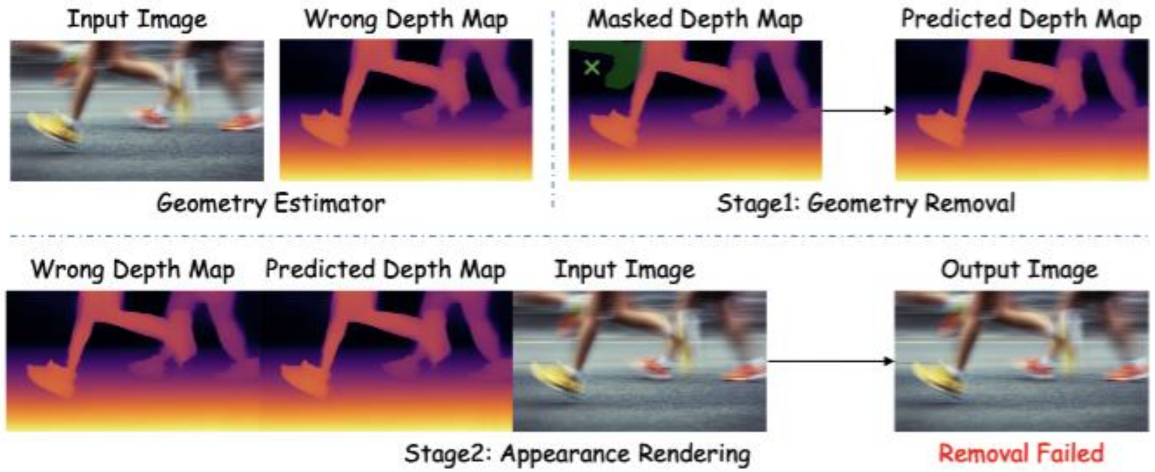
Figure 4: Comparison between our one-stage and two-stage object removal strategies. Two-stage design improves edit quality by separating geometry reasoning from appearance generation.

Experiments

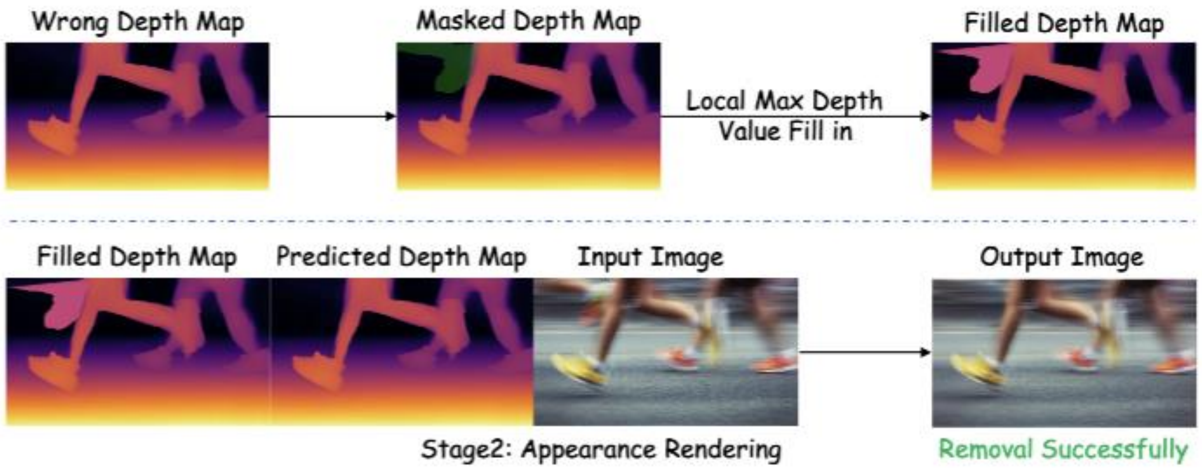


Experiments

Failure Cases



(a) Failure cases under motion blur conditions.



(b) Improved results after applying Fill-in strategy.

Datasets

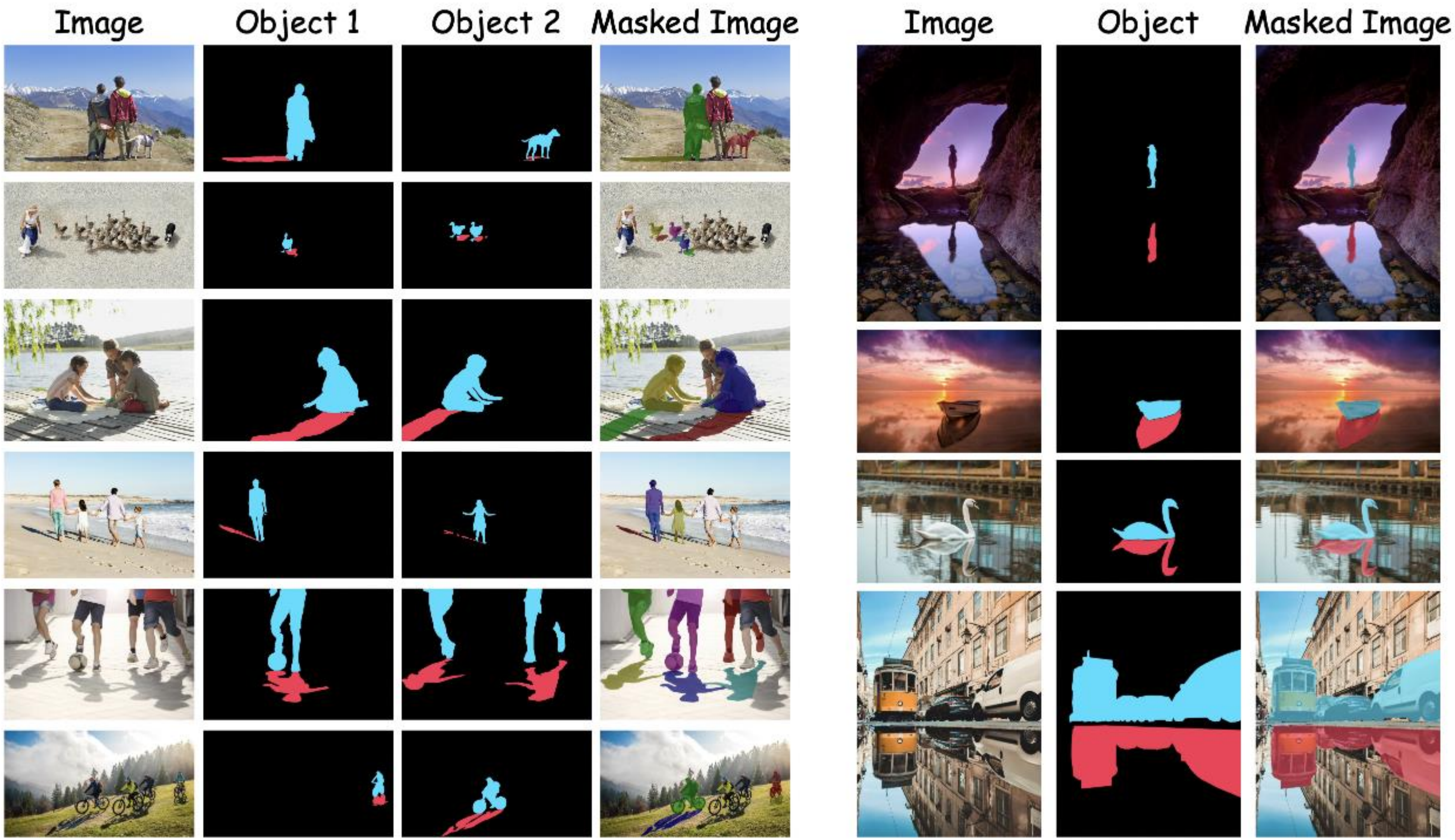


Image
Object 1
Object 2
Masked Image
Image
Object
Masked Image

Conclusion

We propose a **geometry-aware, two-stage framework** that rethinks object removal as a causal process: first **remove the object in geometry space** (depth) under strictly mask-aligned supervision; then **render the appearance from the updated geometry**, so shadows/reflections are naturally cleared while preserving outside-mask content. A DPO-style preference loss stabilizes depth editing and suppresses spurious structure insertions.

Our contributions are three-fold:

- **Decoupled pipeline:** a scalable **two-stage** design—**geometry removal → appearance rendering**—that separates structure reasoning from pixel synthesis.

- **Controllable geometry editing: strictly mask-aligned** training with a **preference-guided (DPO-style) loss** to avoid unwanted insertions and ensure well-posed supervision.

- **Causal artifact removal & SOTA results:** geometry-conditioned rendering (with **bidirectional** training) that removes **outside-mask causal artifacts** and achieves state-of-the-art performance on RemovalBench/RORD-Val and higher IoU on CausRem.

Thank You !