ECCV 2026 | MALMÖ | SEPT 8–12

The 19th European Conference on Computer Vision, ECCV 2026

EraseLoRA: MLLM-Driven Foreground Exclusion and Background Subtype Aggregation for Dataset-Free Object Removal

Sanghyun Jo[1, 2*†] **Donghwan Lee[2*]** **Eunji Jung[2*]** **Seong Je Oh[2]** **Kyungsu Kim[2†]**

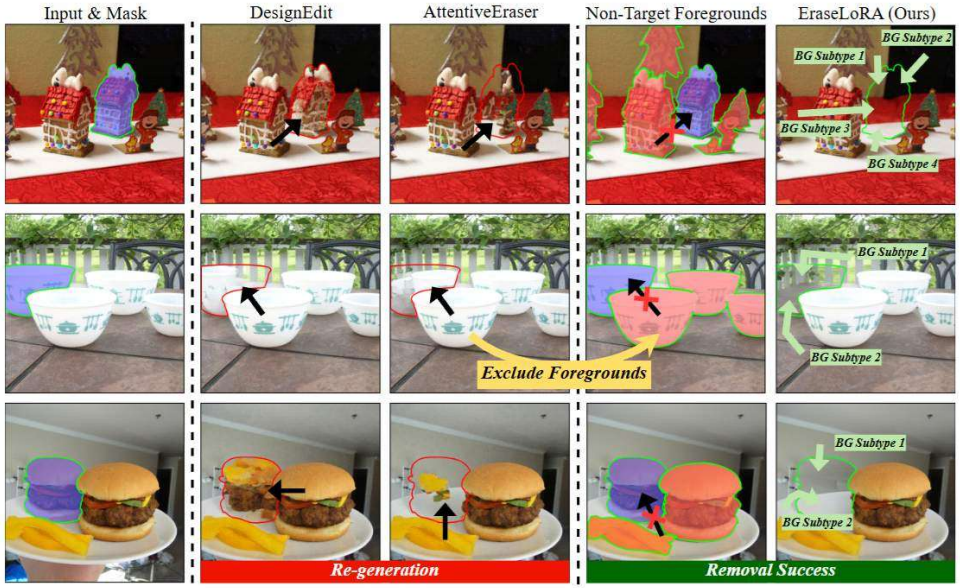
[1]OGQ corp
[2]Seoul National University

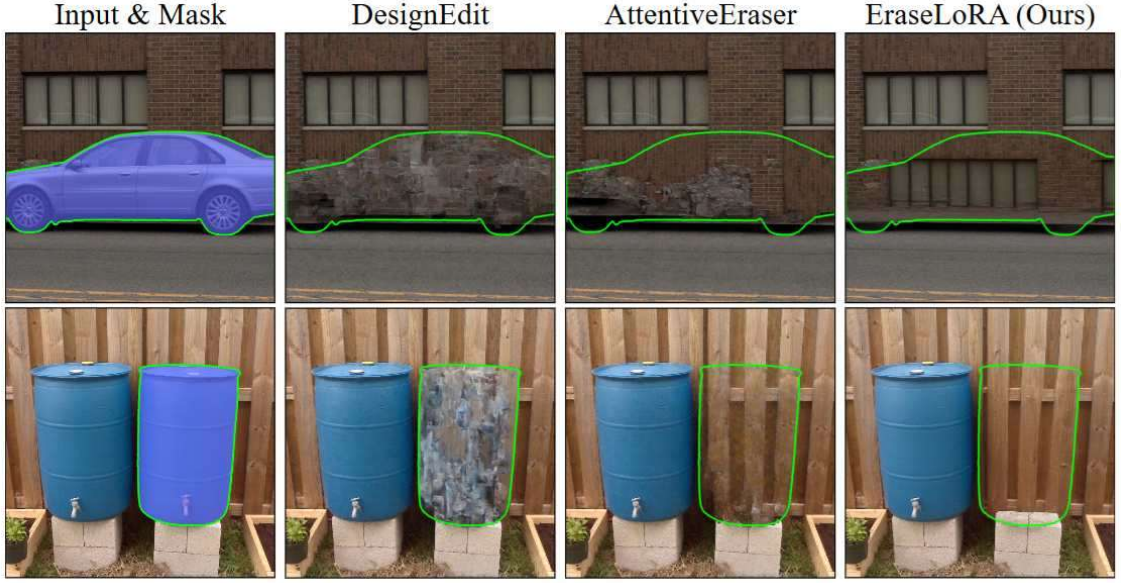
*Equal Contribution, †Corresponding Author

Artificial Intelligence & Biomedicine Lab
Core AI Foundations · Biomedical Intelligence · Deployable Systems

Problem

Goal: Remove the target and reconstruct the occluded background



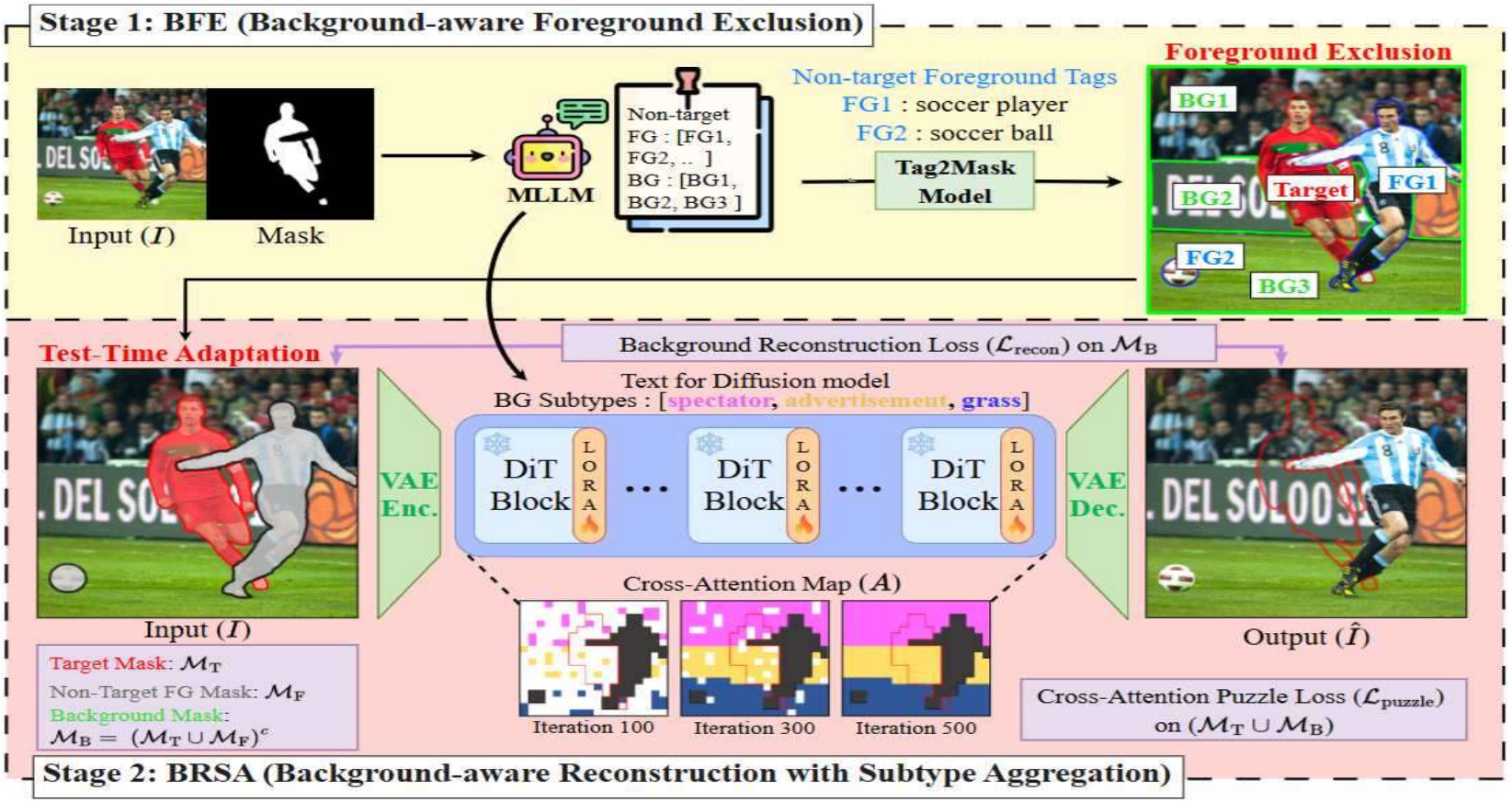


Failure 1:
Non-target foregrounds are mistaken for background
→ Unwanted object re-genration

Failure 2:
Background subtypes are not distnguished
→ Blurred textures and structural misalignment

Both failures originate from the lack of explicit background-aware reasoning

Method Overview



Stage 1 · BFE:
Separate target FG, non-target FG, and background subtypes

Stage 2 · BRSA:
Reconstruct and aggregate background subtypes with test-time LoRA

Dataset-free

No attention blocking

Model-agnostic

BFE isolates clean background supervision

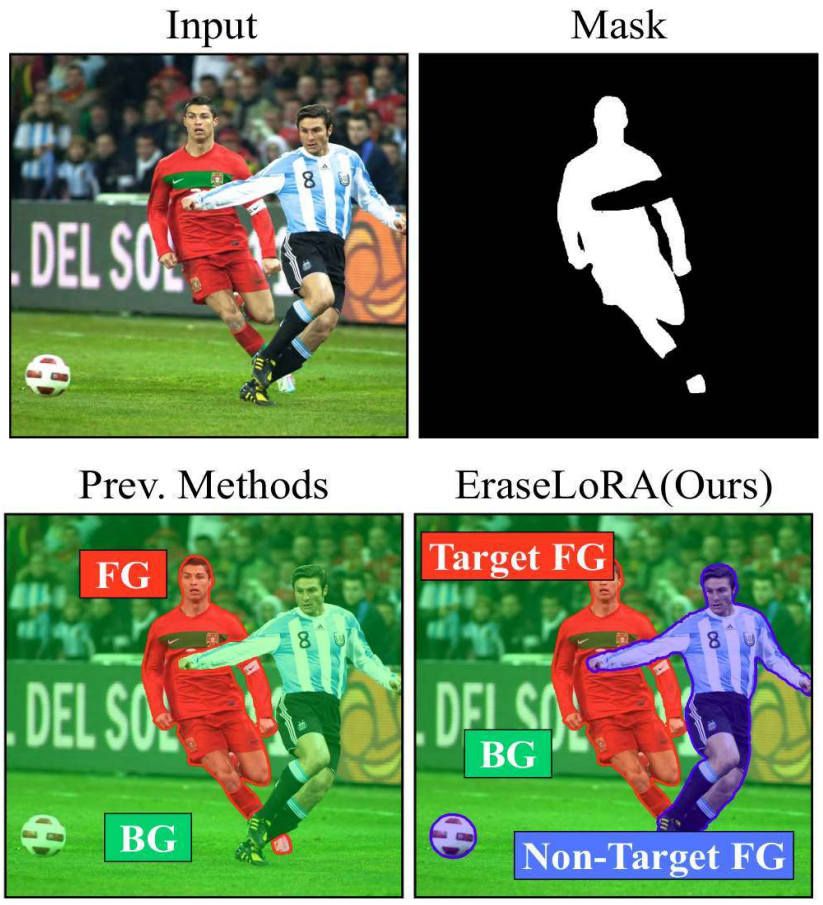
Stage 1 · Background Aware Foreground Exclusion

MLLM → F: Non-target foreground tags
MLLM → B: background subtype tags
Tag2Mask(F) → $\mathbf{M_F}$

Clean Background

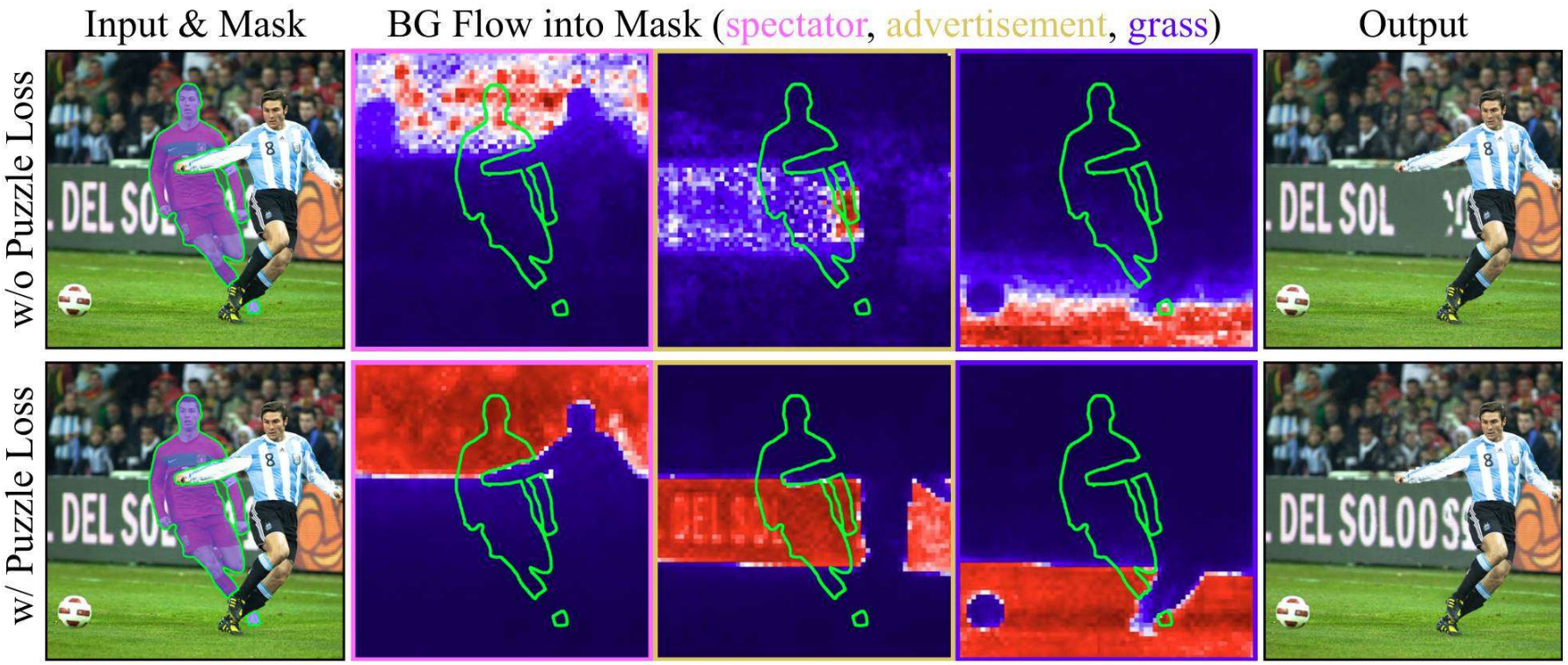
$$\mathbf{M_B} = \Omega - \mathbf{M_T} - \mathbf{M_F}$$

→ BFE prevents foreground distractors from being used as background references



BRSA assembles complementary background subtypes

Stage 2 · Background Aware Reconstruction with Subtype Aggregation



(1) Reconstruction Loss
- Preserve clean-background appearance and structure for input image

$$\mathcal{L}_{\text{recon}} = \frac{1}{M} \sum_{} \left\| \widehat{z[p]} - z[p] \right\|^2$$

(2) Puzzle Loss
- Aggregate background subtypes in valid regions

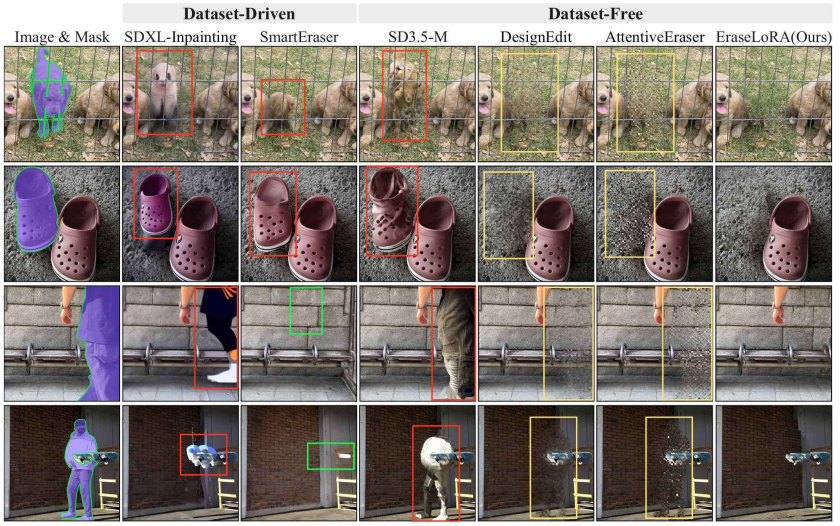
$$\mathcal{L}_{puzzle} = 1 - \boldsymbol{Dice}(A^{dom}, \mathbf{1}_{M_T \cup M_B})$$

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda \mathcal{L}_{puzzle}$$

Quantitative & Qualitative comparison on OpenImages V7 and RORD

BG Sim. (↑) : Faithful background reconstruction / FG Sim. (↓) : Less object re-generation / BG Pres. (↑) : Unmasked-region preservation

	OpenImages V7			RORD		
	BG Sim.(↑)	FG Sim.(↓)	BG Pres.(↑)	BG Sim.(↑)	FG Sim.(↓)	BG Pres.(↑)
Dataset-Free Approaches:						
SD3.5-M [8]	0.605	0.286	**0.934**	0.582	0.319	0.907
+ AttentiveEraser [34]	0.559	0.276	0.931	0.541	0.302	0.901
+ DesignEdit [14]	0.600	0.255	0.933	0.597	0.273	**0.908**
+ **EraseLoRA**	**0.743**	**0.151**	0.924	**0.779**	**0.138**	0.901
Dataset-Driven Approaches:						
SDXL-Inpainting [26]	0.677	0.212	0.742	0.645	0.234	0.720
PowerPaint [54]	0.669	0.217	0.719	0.729	0.176	0.687
CLIPAway [7]	0.656	0.223	0.713	0.744	0.156	0.705
SmartEraser [15]	0.709	0.185	0.727	0.768	0.148	0.672
EntityErasure [53]	0.679	0.204	0.728	0.766	0.175	0.716



≥ 23%

Higher BG Similarity
Vs. prior dataset-free methods

≥ 1/2

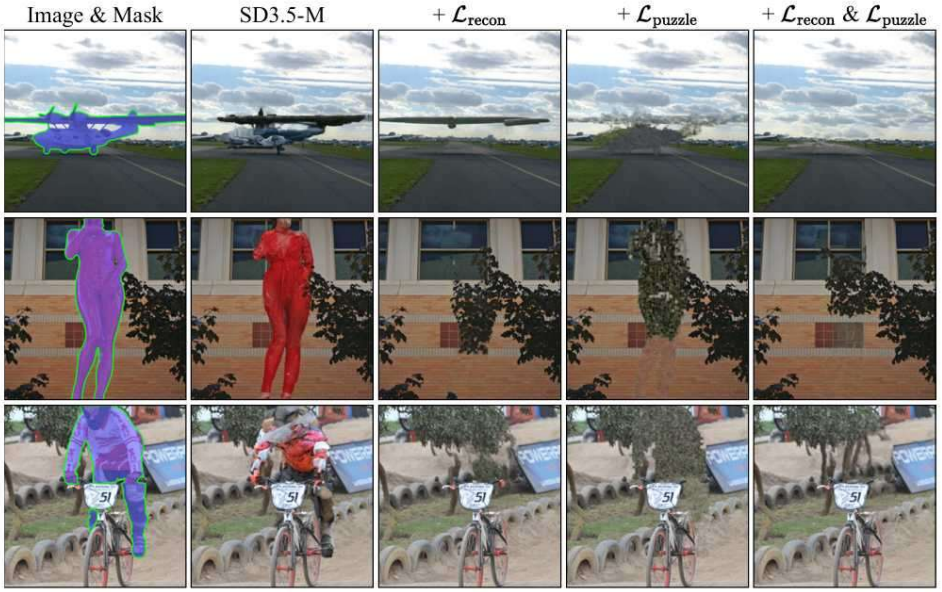
Foreground Similarity
Vs. SD3.5-M

≈ 0.9

Background preservation
Vs. prior dataset-driven methods

Ablation Study

BG Sim. (↑) : Faithful background reconstruction / FG Sim. (↓) : Less object re-generation / BG Pres. (↑) : Unmasked-region preservation



Method	Loss Components		Metrics	
	$\mathcal{L}_{\text{recon}}$	$\mathcal{L}_{\text{puzzle}}$	BG Sim.(↑)	FG Sim.(↓)
SD3.5-M [8]	✗	✗	0.605	0.286
+ EraseLoRA	✓	✗	0.736	0.158
+ EraseLoRA	✗	✓	0.561	0.278
+ EraseLoRA	✓	✓	**0.743**	**0.151**

(1) The reconstruction Loss anchors the original background structure for input image
(2) The Puzzle Loss integrates background subtypes and suppresses foreground cues
(3) Both Loss coherent and detail-preserving reconstruction

Conclusion

Our project page available here! (Paper & Code)

Project

EraseLoRA introduces explicit background reasoning into dataset-free object removal

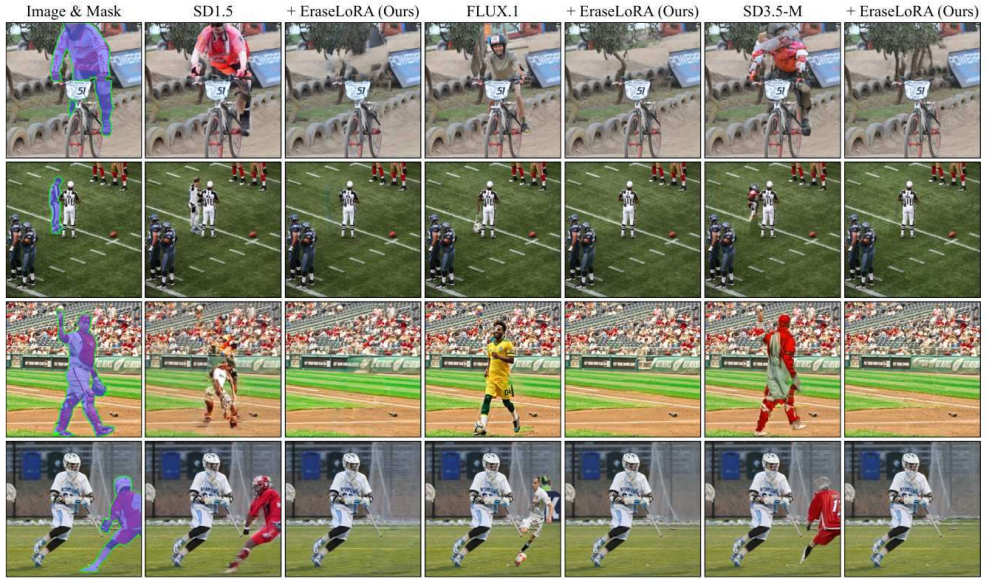


Table 7: EraseLoRA on different diffusion backbones with TTA cost.

Method	Metrics		TTA Cost		
	BG Sim.(↑)	FG Sim.(↓)	Param.	VRAM	Opt. Time
SD1.5 [30]	0.596	0.271	858 M	3.39 GB	-
+ EraseLoRA	**0.702**	**0.176**	+6.38 M	+0.68 GB	117 s
SDXL [26]	0.608	0.297	2,573 M	11.5 GB	-
+ EraseLoRA	**0.730**	**0.186**	+46.5 M	+5.86 GB	250 s
SD3.5-M [8]	0.605	0.286	2,243 M	21.9 GB	-
+ EraseLoRA	**0.743**	**0.151**	+23.9 M	+9.0 GB	191 s
FLUX.1 [3]	0.588	0.205	11,902 M	38.5 GB	-
+ EraseLoRA	**0.760**	**0.146**	+52.3 M	+20.9 GB	495 s

(1) Background-aware removal : EraseLoRA excludes non-target foregrounds and coherently aggregates multiple background subtypes.
(2) Model-agnostic generalization : It consistently improves object removal across diffusion backbones without requiring paired training data.
(3) TTA cost : 500 iterations take ~3 min/image; early stopping averages 141 iterations, achieving >3.5× speedup while retaining >96.5% of BG similarity.