FuLLaMa: Training-free Diffusion-based Object Removal with Context Preservation

İlke Demir
Cauth AI

Umur Aybars Ciftci
Binghamton University

WACV 2026
Tucson, Arizona • Mar 6 - 10, 2026

Idea

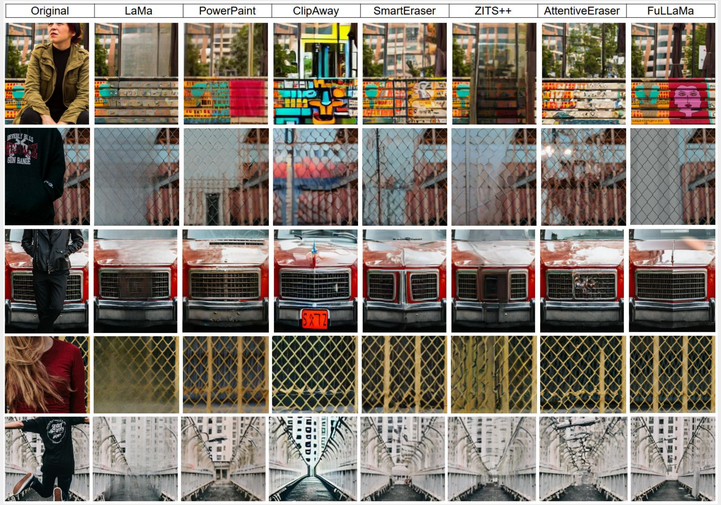
Training-free object removal with LVLM-guided context preservation and adaptive diffusion inference.

Motivation

Existing object removal methods often follow two extremes:

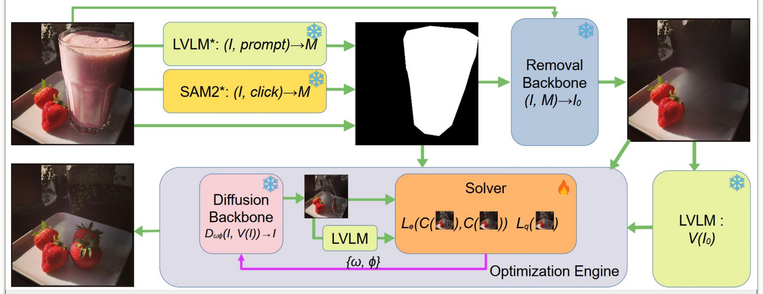
Erasers: remove aggressively, but often cause blur, pixelation, repetitions, broken patterns, and unrealistic fills.

Painters: look realistic, but may fail to remove the target, infill with sundries, or introduce unrelated/nonexisting content.



Goal: remove objects cleanly, preserving scene semantics, do not introduce new things, and keep it as real as possible.

Methodology



FuLLaMa starts from an eraser-like removal, extracts scene context through LVLM→CLIP, and iterates on refining the image with a DM whose inference parameters are optimized by APMN at runtime. All models are used in a frozen state.

Optimization

FuLLaMa initializes removal with a backbone, $R(I, M) = I_0$

extracts semantics via LVLM→CLIP path, $E(I_k) = C(V(I_k))$

refines I with DMs with (ω,ϕ), $D_{\omega,\phi}(I_{i-1}, V(I_{i-1})) = I_k$

while APMN optimizes (ω,ϕ) using $\mathcal{L} = \lambda_q(1 - \mathcal{L}_q) + \lambda_e \mathcal{L}_e$

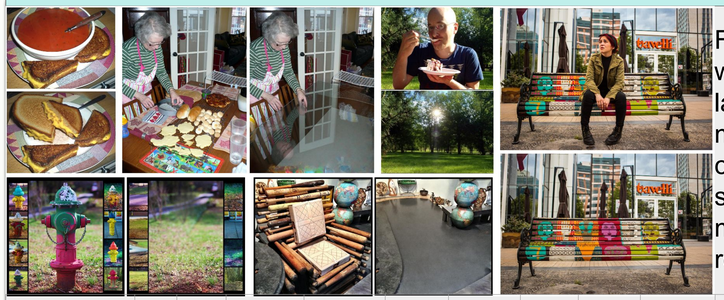
where semantics and quality is balanced with losses as,

$$\mathcal{L}_q = Q(M \odot I_k) \qquad \mathcal{L}_e = \cos(E(I_k), E(I_0))$$

yielding realistic, context-consistent, training-free removal.

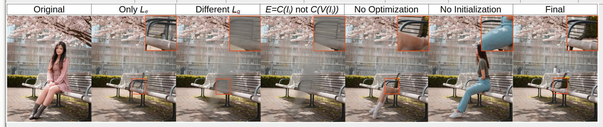
APMN params visually interpret objective values, as context & quality emerges per iter.

Results



FuLLaMa works for large-area, multi-object, complex scene, and more hard removals.

Method	TR	FID	LPIPS	CMMD	CLIP	NsfwR	ReM	MSN	MARS	Maniqa	Claude
ZITS++	✓	121	0.382	1.099	**0.25**	0.055	0.72	1.43	0.593	0.291	4.54%
PowerPaint	✓	114	0.435	1.049	0.32	0.028	0.73	3.09	0.472	0.297	0%
ClipAway	✓	120	0.528	1.073	0.37	0.012	0.84	1.92	0.261	0.305	4.54%
AttentiveEr	✗	168	0.531	1.747	0.35	0.007	**0.87**	**0.73**	**0.121**	0.232	13.63%
SmartEr	✓	**95**	0.358	**0.812**	0.26	0.009	0.73	4.01	0.430	0.337	0%
FuLLaMa	✗	129	**0.550**	1.020	0.32	**0.002**	0.77	1.00	0.209	**0.345**	**72.72%**



Only L_e, only E(I), or different Q: infill is similar to I_o. Without APMN or init, E_0 misses context or sundries.

↳Painters add nudity, erasers replicate body parts. FuLLaMa vs. newest DMs in challenging tasks. →