amazon science prime video

NEURAL INFORMATION PROCESSING SYSTEMS

MACRO: Training-free Multi-plane Attention for Closeup Render Optimization

Nitzan Hodos 

Roy Amoyal 

Lior Fritz 

Ianir Ideses 

Sagie Benaim 

Netalee Efrat 

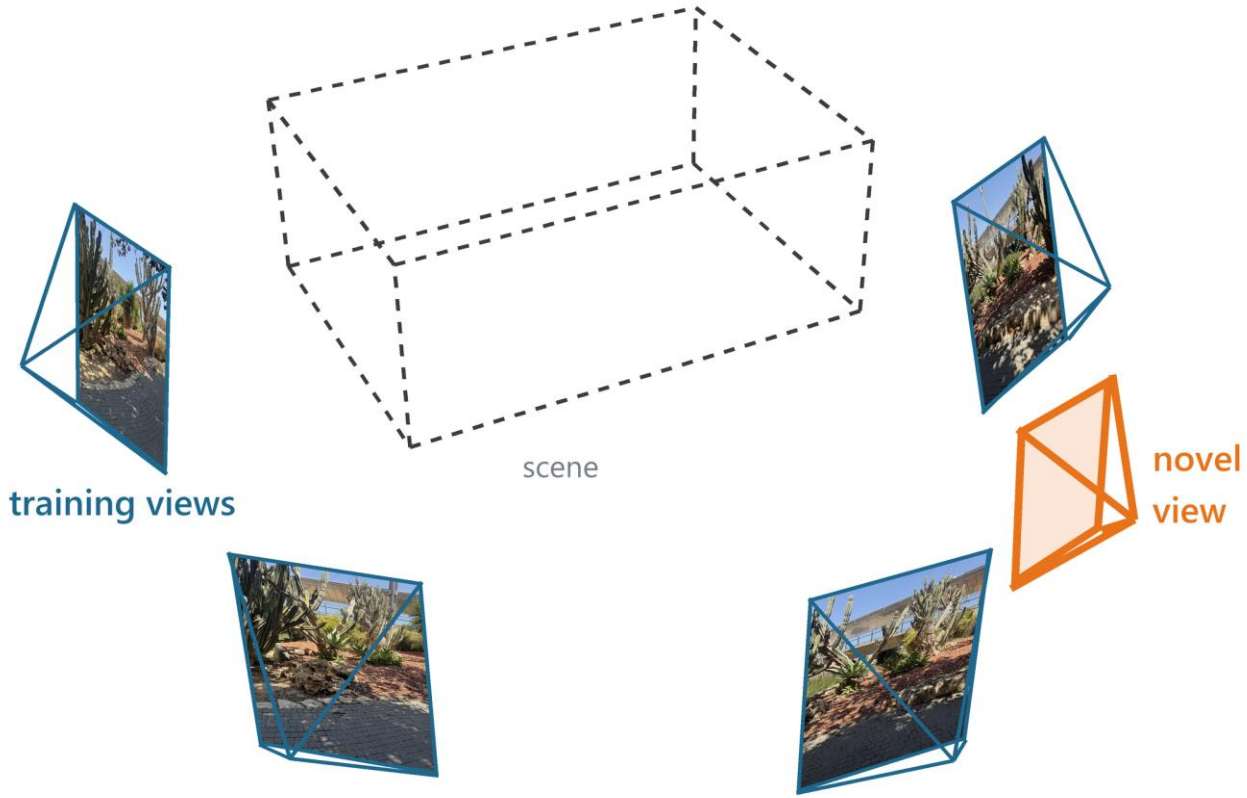
NeurIPS 2026 Spotlight

TECHNION

Ben-Gurion University of the Negev

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

3DGS [1]: real-time novel view synthesis



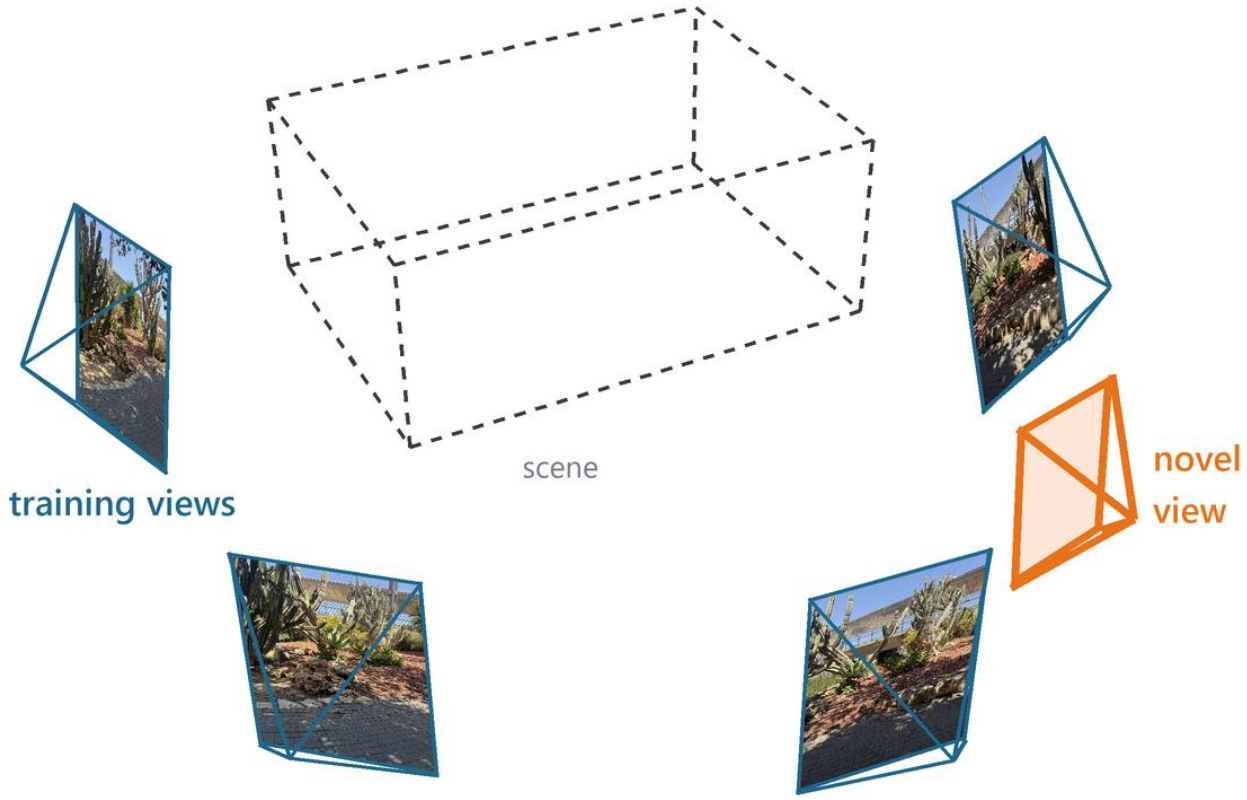
Optimized from **training views**

Renders **novel views** in real time

Best **near the training cameras**

[1] Kerbl, Kopanas, Leimkühler, Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Trans. Graph. 42(4), 2023.

Close-up rendering: 3DGS degrades

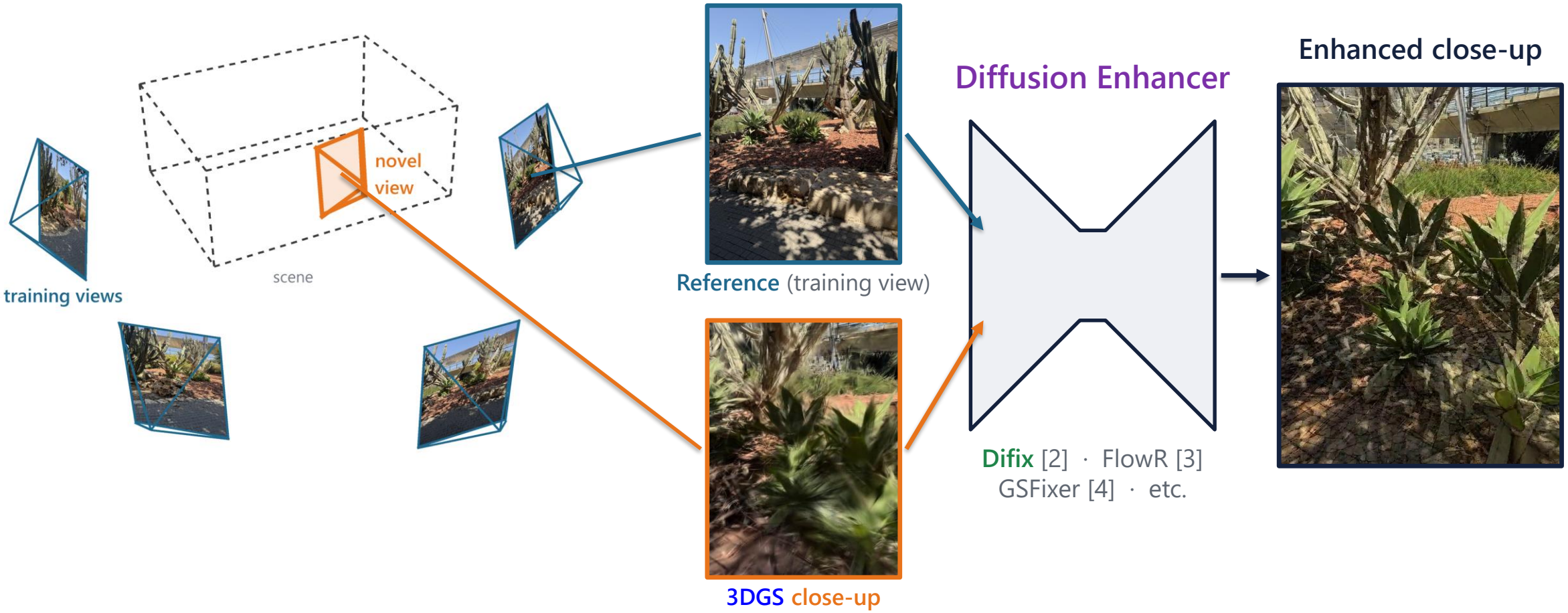


3DGS render from the orange camera

The close-up is **blurry and broken**, due to:

(1) Limited image resolution (2) Limited number of Gaussians (3) Disocclusions

Diffusion enhancers use real references



Conditions on **real reference images** to restore the **degraded rendering**

[2] Wu et al. DIFIX3D+: Improving 3D Reconstructions with Single-Step Diffusion Models. CVPR 2025.
[3] Fischer et al. FlowR: Flowing from Sparse to Dense 3D Reconstructions. ICCV 2025.
[4] Yin et al. GSFixer: Improving 3D Gaussian Splatting with Reference-Guided Video Diffusion Priors. arXiv:2508.09667, 2025.

Existing diffusion enhancers fail on close-ups

3DGS | Difix | MACRO (Ours)

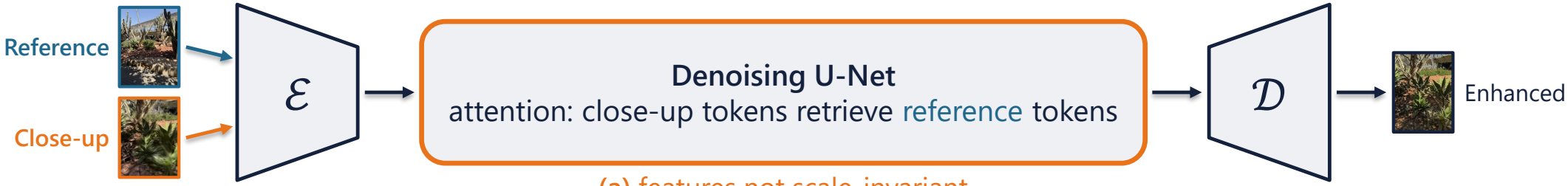


Training view → close-up → back (half speed)

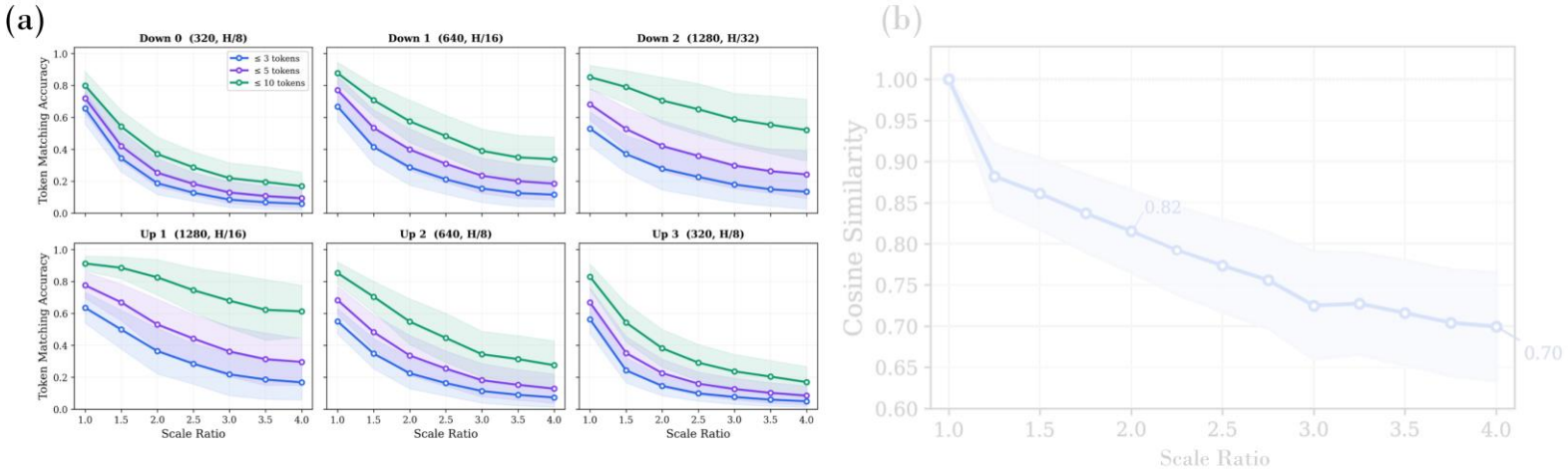
Difix copies by matching textures, but across a scale gap:

(1) **Wrong object**: copies a look-alike texture → broken geometry
(2) **Wrong scale**: right object, but its texture copied too small

Why? Diffusion enhancers are not scale-invariant



(a) features not scale-invariant



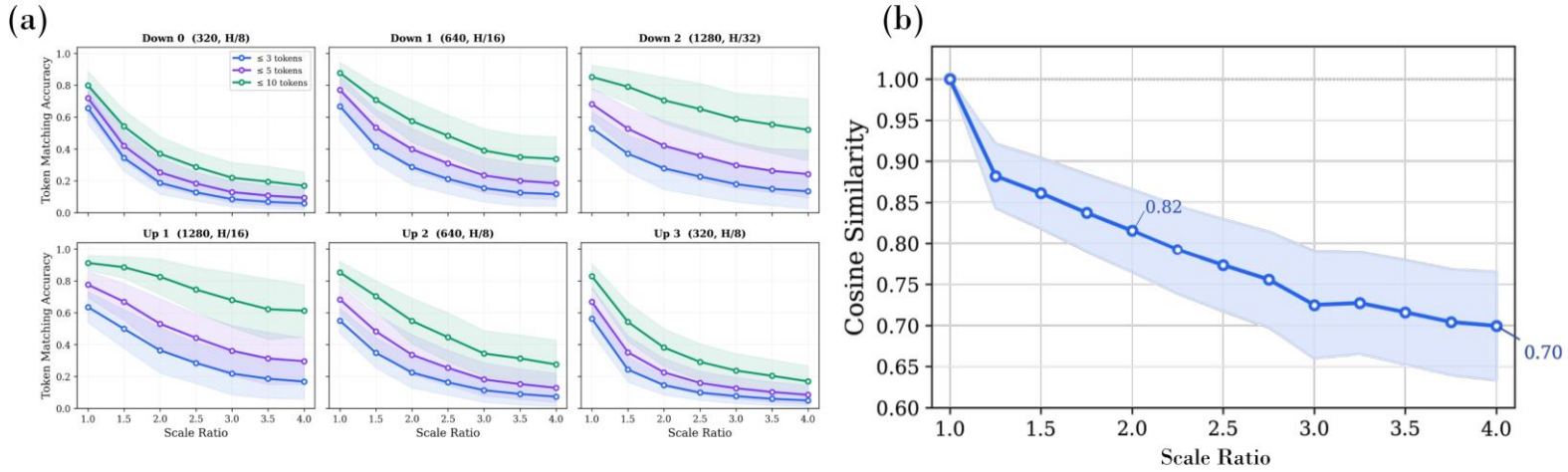
(a) Attention retrieves **incorrect correspondences** as the scale ratio grows

Why? Diffusion enhancers are not scale-invariant



(b) not scale-equivariant

(a) features not scale-invariant



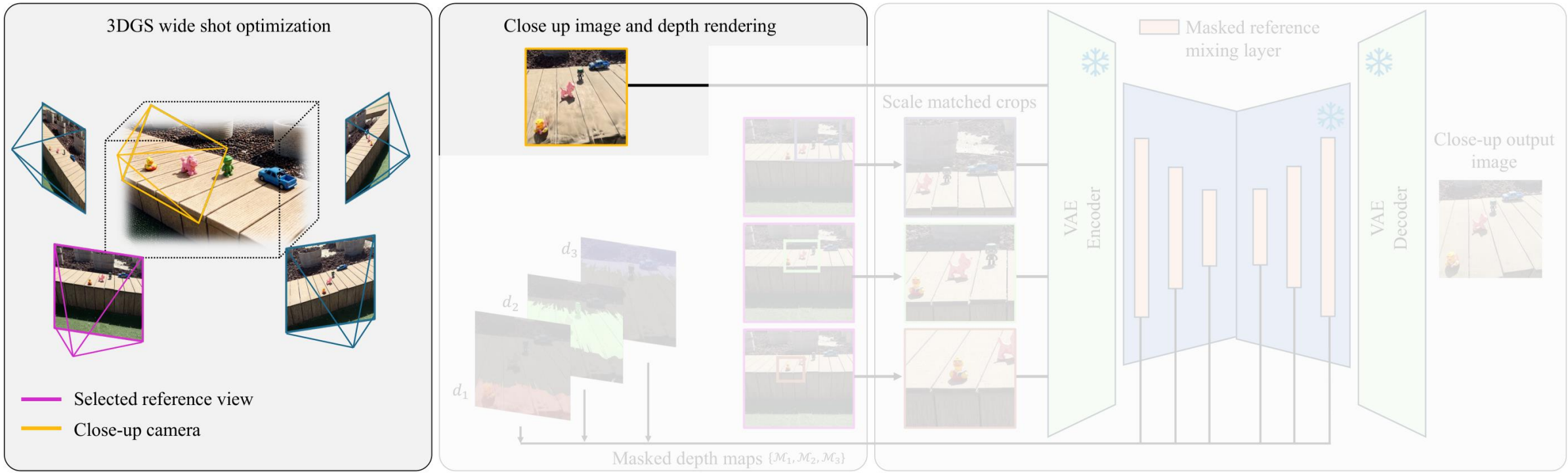
(a) Attention retrieves **incorrect correspondences** as the scale ratio grows

(b) $\mathcal{E}$(Resize(x)) ≠ Resize($\mathcal{E}$(x))

⇒ Align the scale in image space, before encoding

MACRO: a scale-matched crop for every depth plane

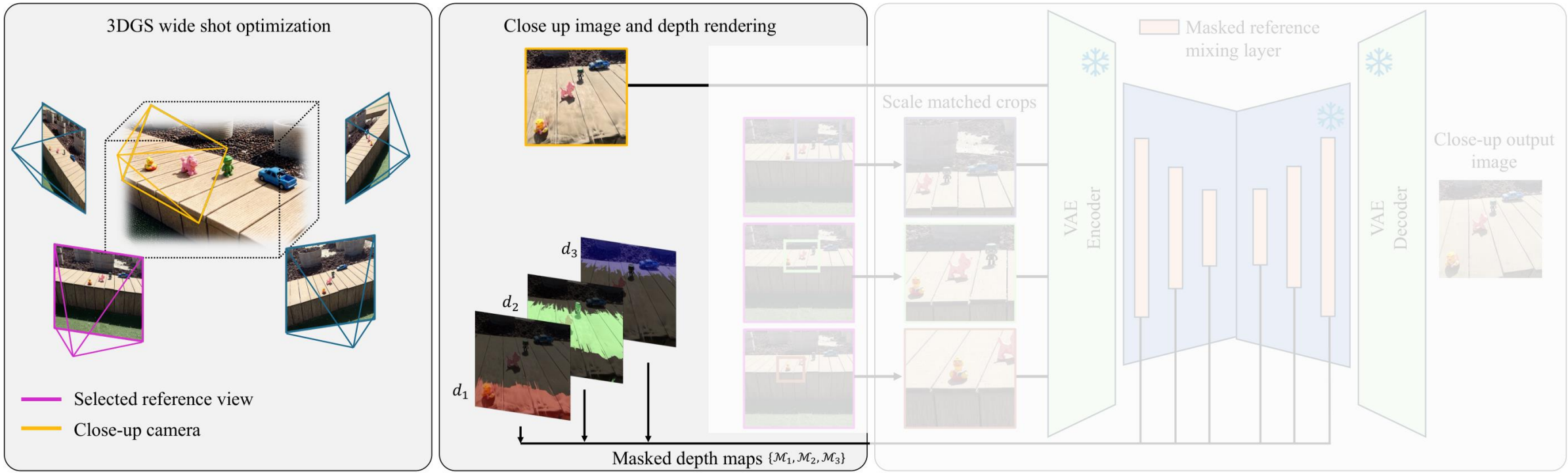
Training-free · on top of a frozen **Difix** · single forward pass



1 **3DGS renders** the close-up (orange) and its depth; **references** (pink) are picked to cover it

MACRO: a scale-matched crop for every depth plane

Training-free · on top of a frozen **Difix** · single forward pass

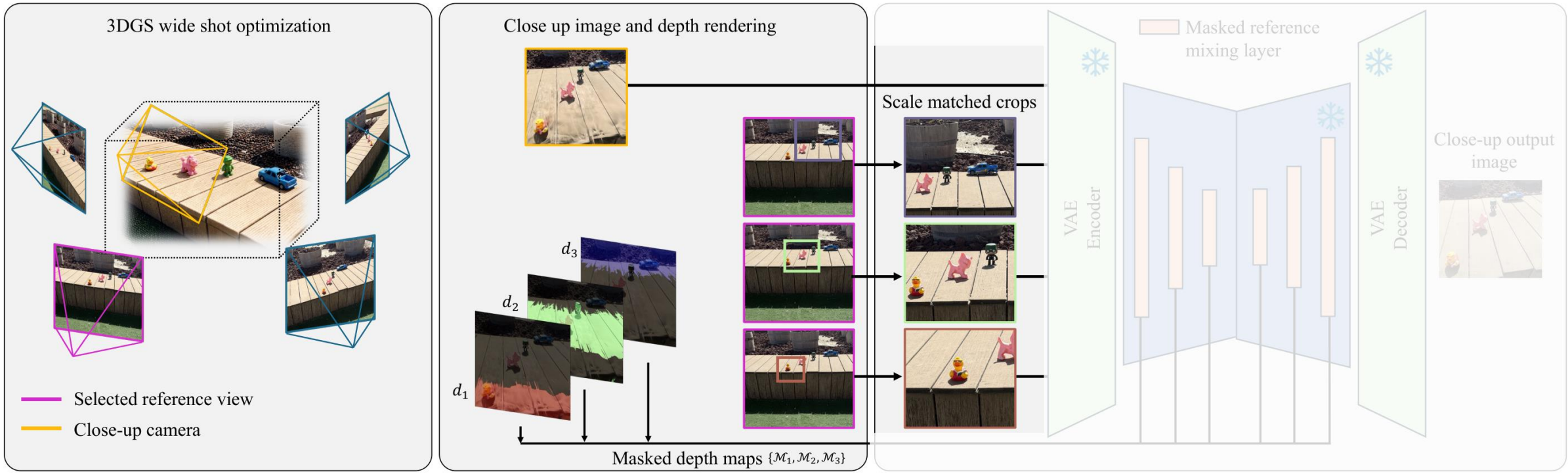


1 **3DGS renders** the close-up (orange) and its depth; **references** (pink) are picked to cover it

2 Split the close-up into **P depth planes** (k-means on depth)

MACRO: a scale-matched crop for every depth plane

Training-free · on top of a frozen **Difix** · single forward pass



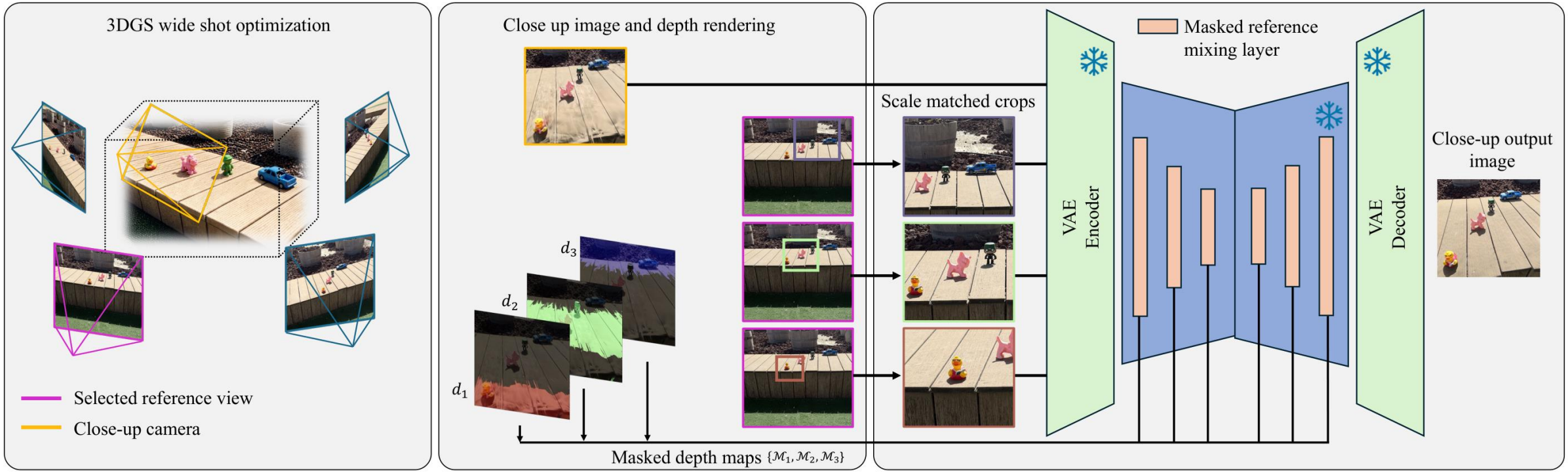
1 **3DGS renders** the close-up (orange) and its depth; **references** (pink) are picked to cover it

2 Split the close-up into **P depth planes** (k-means on depth)

3 **Crop and rescale** the reference for each plane

MACRO: a scale-matched crop for every depth plane

Training-free · on top of a frozen **Difix** · single forward pass



1 **3DGS renders** the close-up (orange) and its depth; **references** (pink) are picked to cover it

2 Split the close-up into **P depth planes** (k-means on depth)

3 **Crop and rescale** the reference for each plane

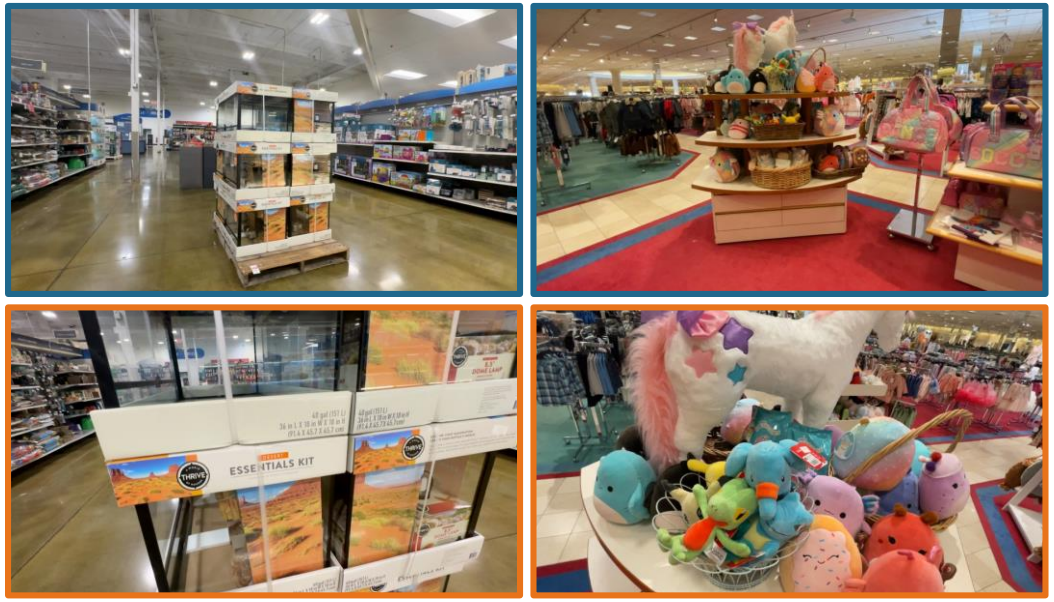
4 Frozen **Difix with a cross-attention mask**: each close-up token attends only to its plane's crops

Two new close-up benchmarks

Wide-shots
training views

Close-ups
evaluation views

DL3DV-Closeup



44 scenes · 314 close-ups
curated from DL3DV-10K
Zoom ×1.5 – ×5.8

MobileClose-10



10 scenes · 39 close-ups
captured with a smartphone
Zoom ×1.5 – ×5.2

Best perceptual quality on both benchmarks

UP TO

38%

better perceptual metrics than the best enhancement baseline

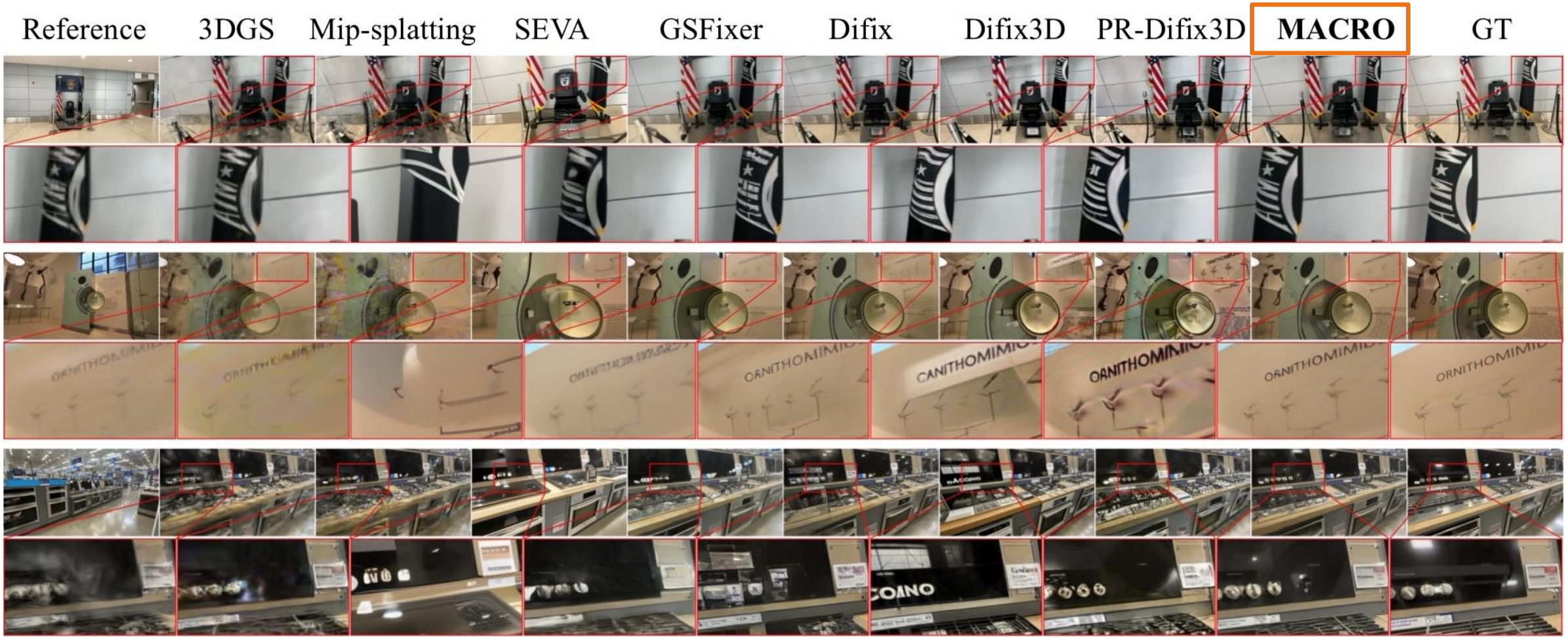
UP TO

67%

better perceptual metrics than unenhanced **3DGS**

Method	*DL3DV-Closeup (44 scenes)*					*MobileClose-10 (10 scenes)*				
	PSNR↑	SSIM↑	LPIPS↓	DSim↓	DINO↓	PSNR↑	SSIM↑	LPIPS↓	DSim↓	DINO↓
3DGS [8]	16.00	0.499	0.457	0.261	0.462	14.58	**0.441**	0.578	0.293	0.512
Mip-Splatting [28]	**16.26**	**0.514**	0.478	0.257	0.440	14.41	0.428	0.667	0.335	0.521
SEVA [32]	12.35	0.335	0.544	0.156	0.263	13.91	0.401	0.647	0.270	0.417
GSFixer [26]	15.91	0.494	0.378	0.121	0.250	14.51	0.401	0.614	0.250	0.408
Difix [19]	15.81	0.477	0.343	0.128	0.205	14.55	0.407	0.557	0.253	0.352
Difix3D [19]	14.69	0.427	0.394	0.149	0.248	13.76	0.418	0.582	0.258	0.406
Difix3D+ [19]	14.33	0.404	0.393	0.155	0.243	13.36	0.385	0.579	0.270	0.394
PR-Difix3D	15.15	0.433	0.399	0.153	0.262	13.68	0.393	0.563	0.292	0.428
PR-Difix3D+	14.42	0.379	0.430	0.178	0.271	13.00	0.329	0.596	0.320	0.437
MACRO (ours)	16.17	0.505	**0.330**	**0.105**	**0.152**	**14.60**	0.433	**0.551**	**0.168**	**0.217**

Qualitative Results



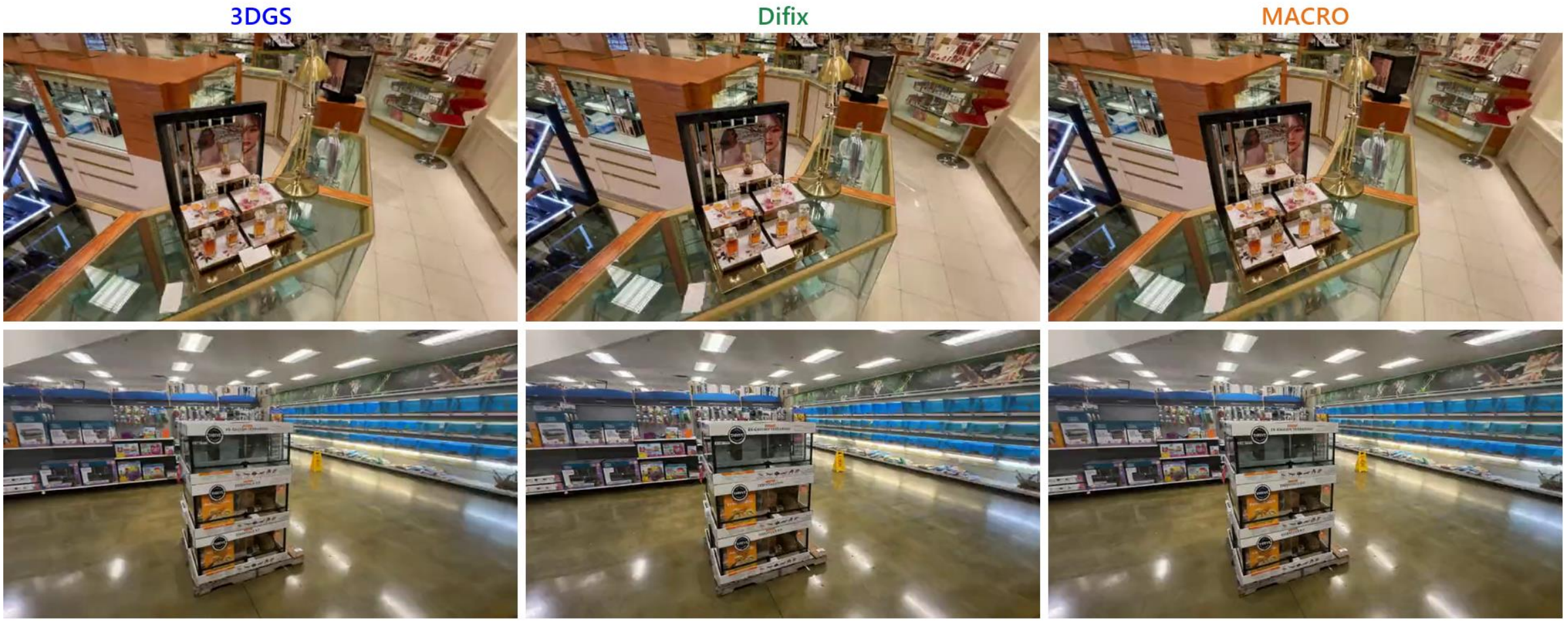
DL3DV-Closeup

Qualitative Results



MobileClose-10

Trajectory Videos



DL3DV-Closeup

Takeaways

1 **Diffusion enhancers** fail at close-ups: their features are **not scale-invariant**

2 **MACRO** matches the scale **per depth plane, in image space**, training-free

3 Two close-up benchmarks: **DL3DV-Closeup** and **MobileClose-10**

Code · Data · Videos
nitzanhod.github.io/MACRO

Poster: Hall 1-4 Wed 10:00 AEDT in Sydney Poster Session 3