

Sem-NeuS: Semantically-Guided High-Fidelity Neural Surface Reconstruction via Geometry Distillation

Abderraouf Amrani and Hamid Laga

School of Information Technology

Murdoch University - Australia

Ngala kwop biddi.
Building a brighter
future, together.

ICPR 2026 Oral Session O6-E.T2
Computer and Robot Vision · Lyon, France

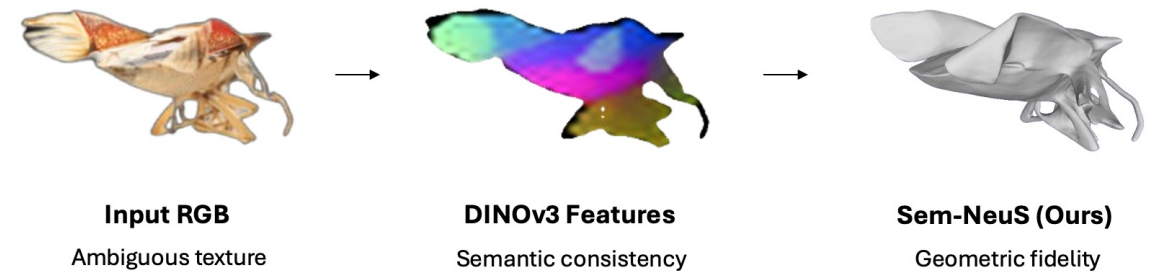


Motivation

Problem: RGB supervision becomes ambiguous

Why photometric training can fail?

- Many nearby surfaces can produce similar RGB observations
- Weak texture makes correspondence unreliable
- Thin geometry can be smoothed, fused, or mistaken for noise

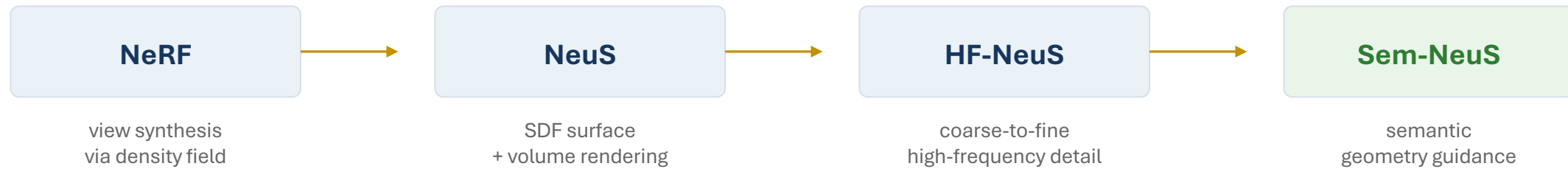


Goal: recover a continuous high-fidelity surface, not just plausible training views.

Paper focus: multi-view RGB $\rightarrow$ SDF surface reconstruction with improved detail in ambiguous regions.

Background and gap

From neural rendering to surface-oriented reconstruction

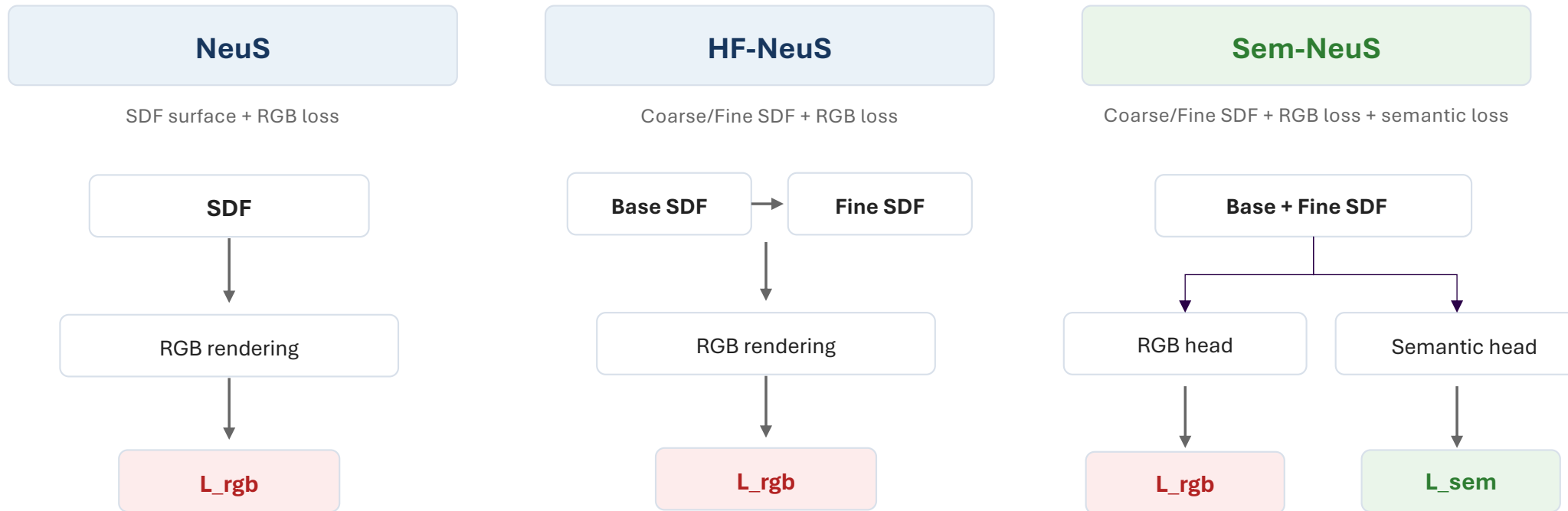


Remaining gap after high-frequency capacity

HF-NeuS can represent fine detail, but supervision is still primarily photometric. When RGB evidence is ambiguous, extra capacity can still be spent on smoothing errors or photometric artifacts.

Background and gap

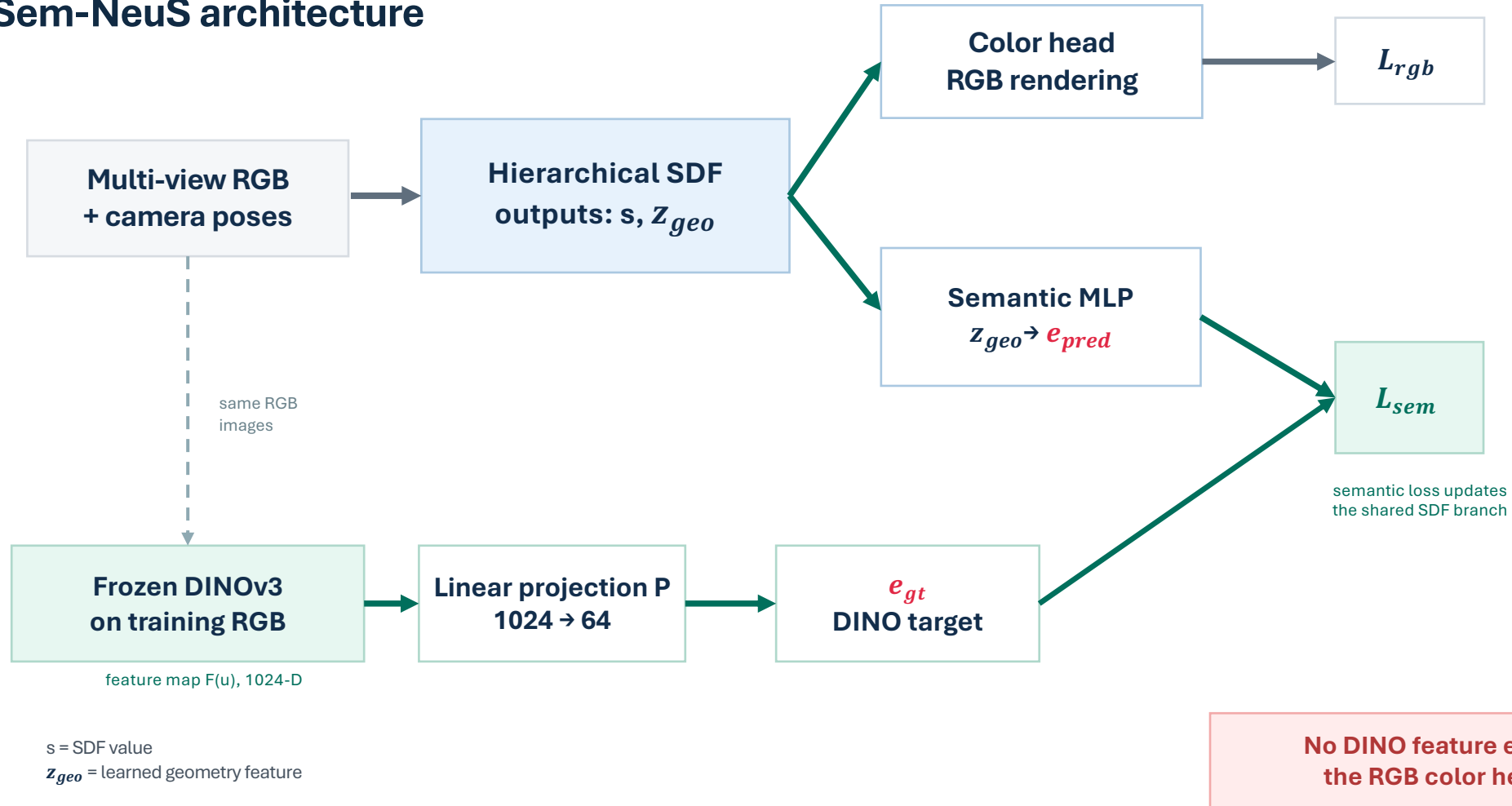
What changes from NeuS / HF-NeuS?



New in Sem-NeuS: DINO-derived supervision is attached to the shared geometry representation.

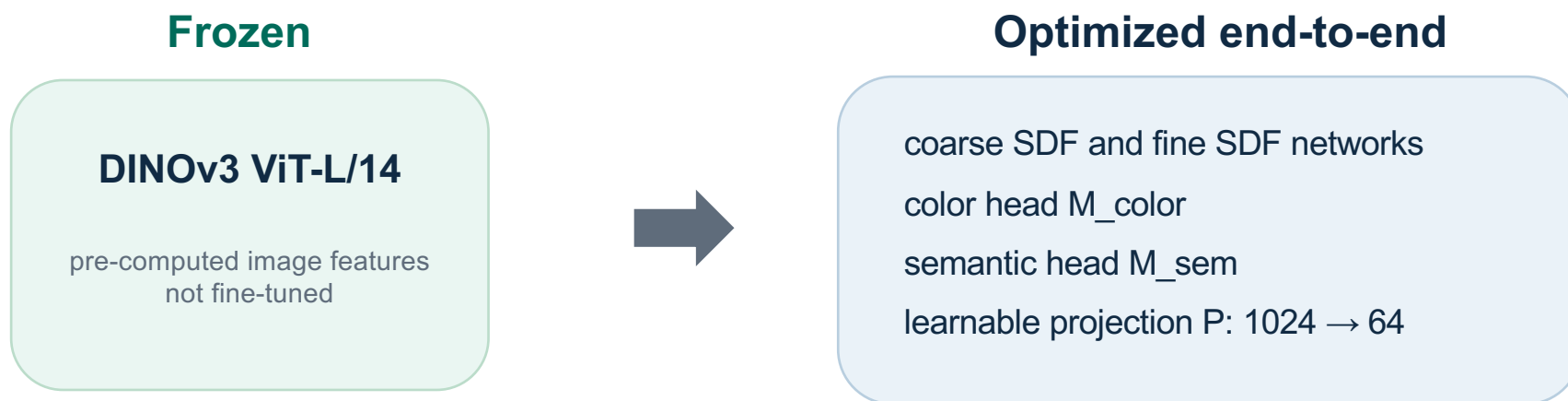
Method Overview

Sem-NeuS architecture



Method overview

Training: frozen teacher, learned geometry



Objective standard SDF rendering losses + $\lambda_{\text{sem}} \cdot L_{\text{sem}}$

$\lambda_{\text{sem}} = 0.05$ · offline DINO extraction: 3–7 min/scene · +8% per-iteration cost

DINO provides a frozen supervision signal; the reconstructed SDF geometry is learned end-to-end.

Method overview

SDF backbone: surface first, details through hierarchy

Surface representation

$$S = \{ x \in \mathbb{R}^3 \mid f(x) = 0 \}$$

NeuS-style rendering converts the SDF into density and ray weights.

Hierarchical representation



What hierarchy gives

capacity for complex surface geometry

What it does not solve alone

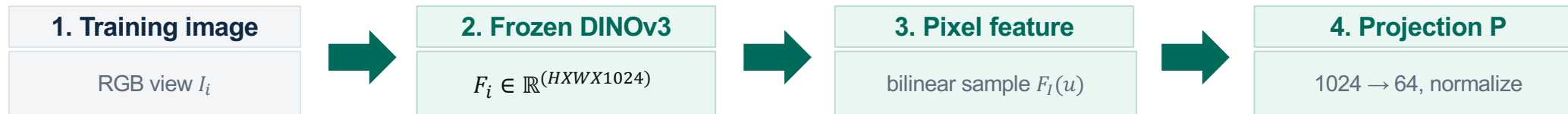
RGB ambiguity in weak texture or thin regions

What Sem-NeuS adds

DINO-based geometry supervision

Method overview

How DINOv3 is used?



DINO target

e_{gt}

Prediction from geometry

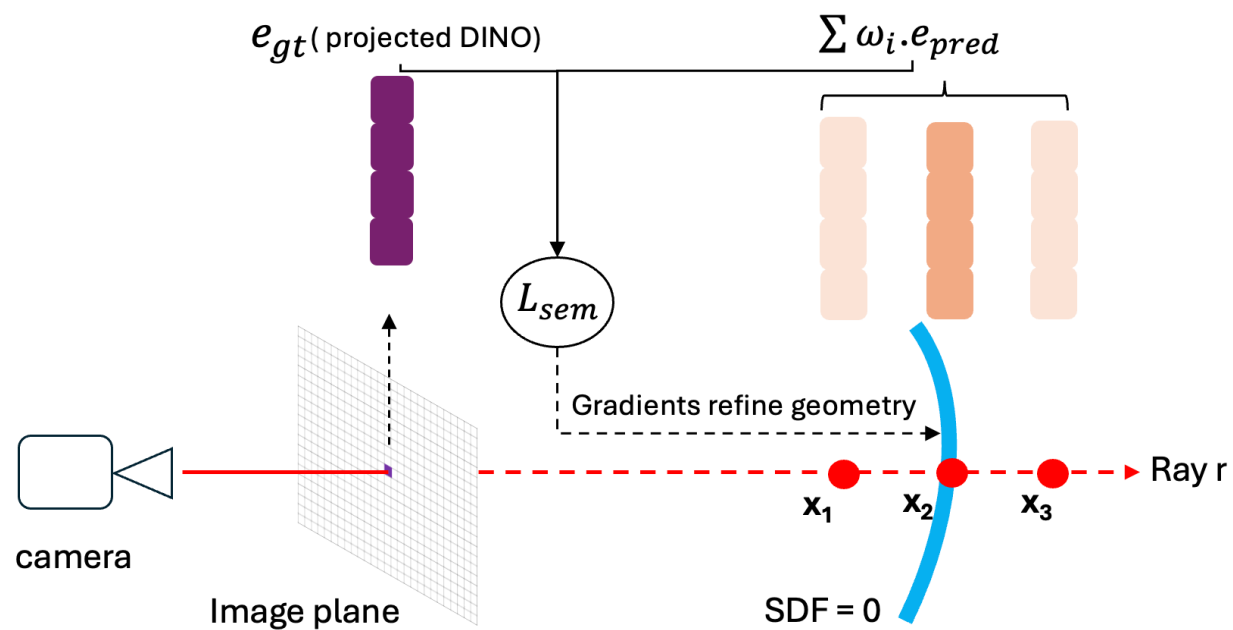
e_{pred}

L_{sem} aligns prediction with target

DINO is a supervision target, not a runtime geometry input.

Method overview

Semantic distillation along a camera ray



Volume-weighted cosine loss

$$L_{sem}(r) = \frac{\sum_{i=0}^N w_i (1 - (e_{pred}(r(t_i)) e_{gt}(r)))}{\sum_{i=0}^N w_i + \epsilon}$$

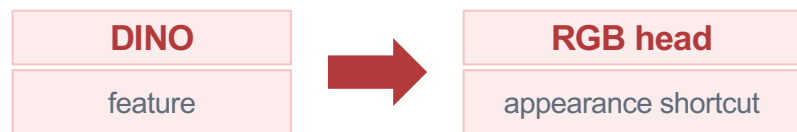
e_{pred}	e_{gt}	w_i
from z_{geo}	image target	SDF weight

weights focus supervision near the visible surface
the semantic head is tied to the shared geometry latent
multi-view consistency is implicit through one shared 3D field

Method overview

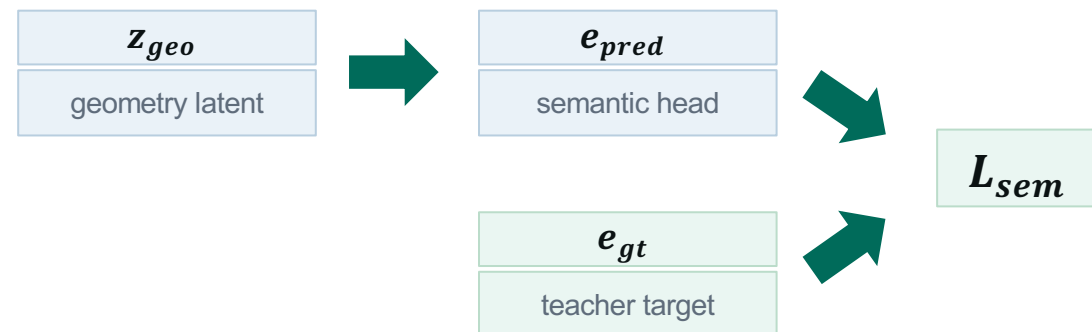
Why distillation rather than DINO injection?

Naïve injection



**risk: fits RGB
without valid surface**

Sem-NeuS distillation



Ablation on Chinese Statue

Variant	CD ↓	NC ↑
Full Sem-NeuS	0.003992	0.945
w/o semantic	0.006252	0.890
Direct injection	0.012500	0.760

The ablation supports the mechanism: DINO should supervise geometry, not enter the renderer.

Experimental setup

Datasets

OmniObject3D benchmark

Statue, Bread, Lego · ground-truth meshes available

Custom RGB sequence

211 iPhone 13 Pro images · COLMAP poses · no GT mesh

Baselines

- NeuS
- HF-NeuS
- AlphaSurf
- 3D Gaussian Splatting

Metrics

Geometry

CD, F1, Hausdorff

Rendering

PSNR, SSIM, LPIPS

Ablation

component contribution

Custom RGB acquisition: challenging surface geometry



Input data

211 RGB images

captured with iPhone 13 Pro

Camera poses

estimated using COLMAP

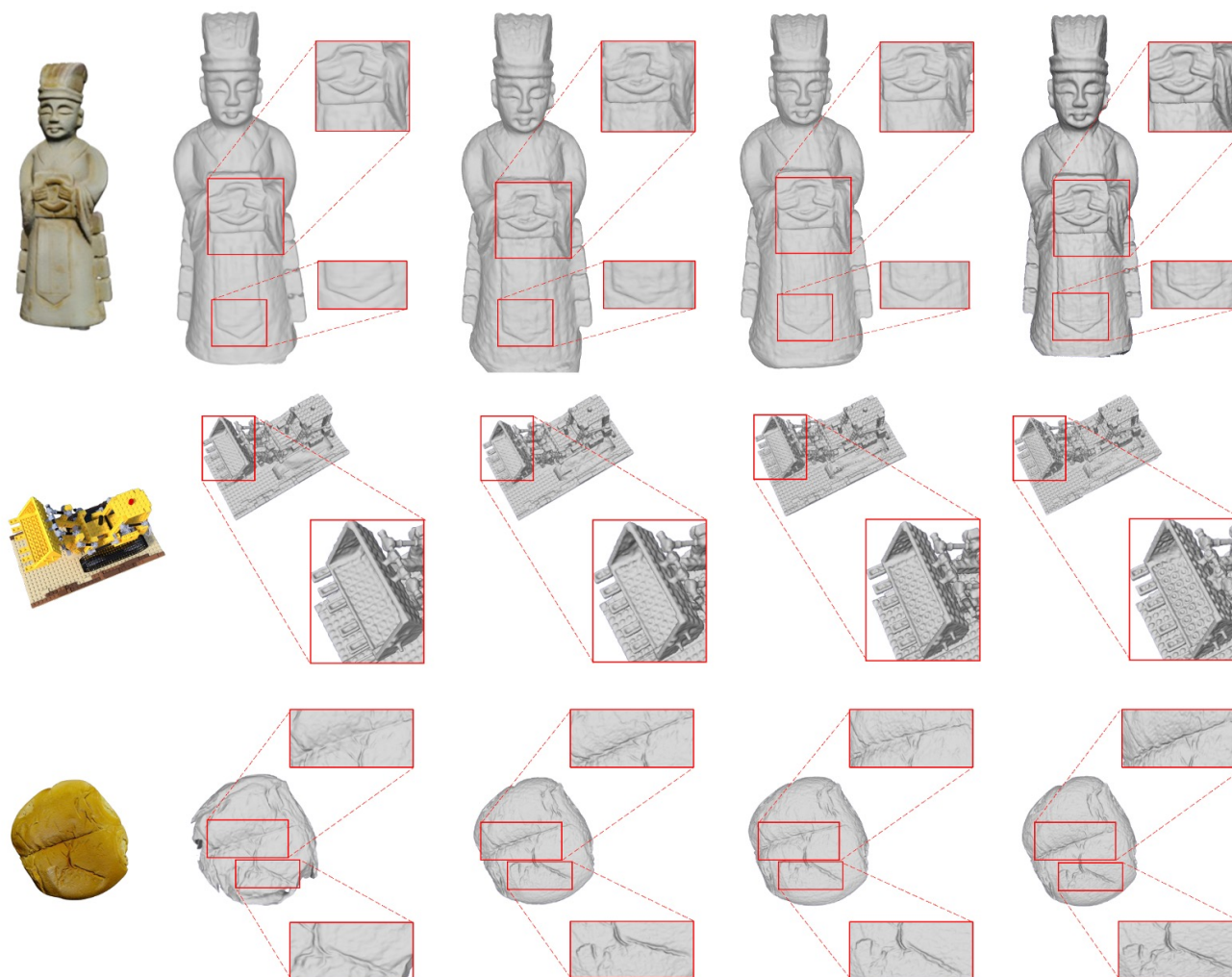
No independent GT mesh

evaluated by held-out views and qualitative fidelity

Why it is challenging

- weak texture
- thin structures
- complex self-occlusion
- small-scale surface features

Qualitative surface reconstruction on benchmark objects



RGB

NeuS

AlphaSurf

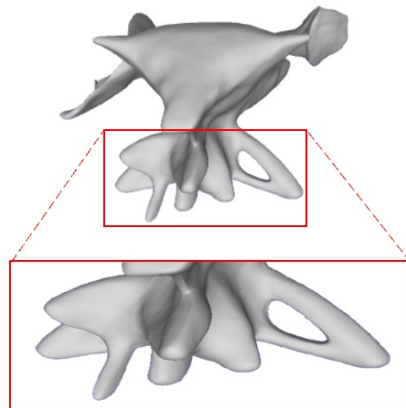
HF-NeuS

Sem-NeuS (Ours)

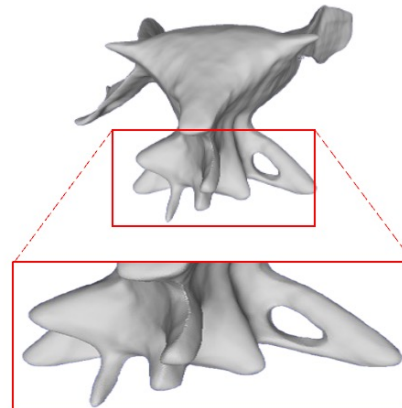
Qualitative stress test on thin structures



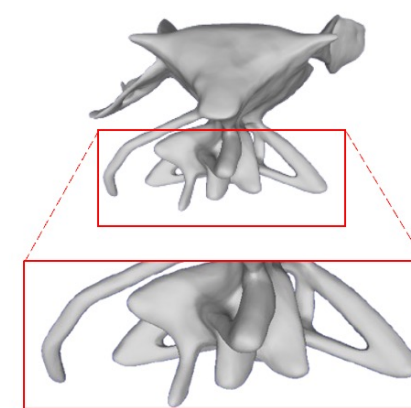
RGB



AlphaSurf



HF-NeuS



Sem-NeuS (Ours)

Interpretation

Baselines

merge or blob thin structures in this example

Sem-NeuS

recovers more distinct continuous surface parts

Scientific restraint

qualitative fidelity, not metric GT geometry

Quantitative geometry: Chamfer Distance

Lower is better · exact CD values from the paper Table 2

Method	Statue ↓	Bread ↓	Lego ↓
NeuS	0.005957	0.006554	0.011715
HF-NeuS	0.006252	0.003583	0.005287
AlphaSurf	0.004311	0.003267	0.005150
Sem-NeuS	0.003992	0.003137	0.005148

F1 and Hausdorff are reported in the paper; CD is shown here for oral clarity.

Interpretation

Clear CD improvement
Statue and Bread
Approximately matched
Lego vs AlphaSurf
Main evidence
CD results + ablation

Ablation: the integration mechanism matters

Variant	CD ↓	NC ↑
w/o Hierarchical	0.005150	0.915
w/o Semantic Distillation	0.006252	0.890
Direct DINO Injection	0.012500	0.760
Full Sem-NeuS	0.003992	0.945

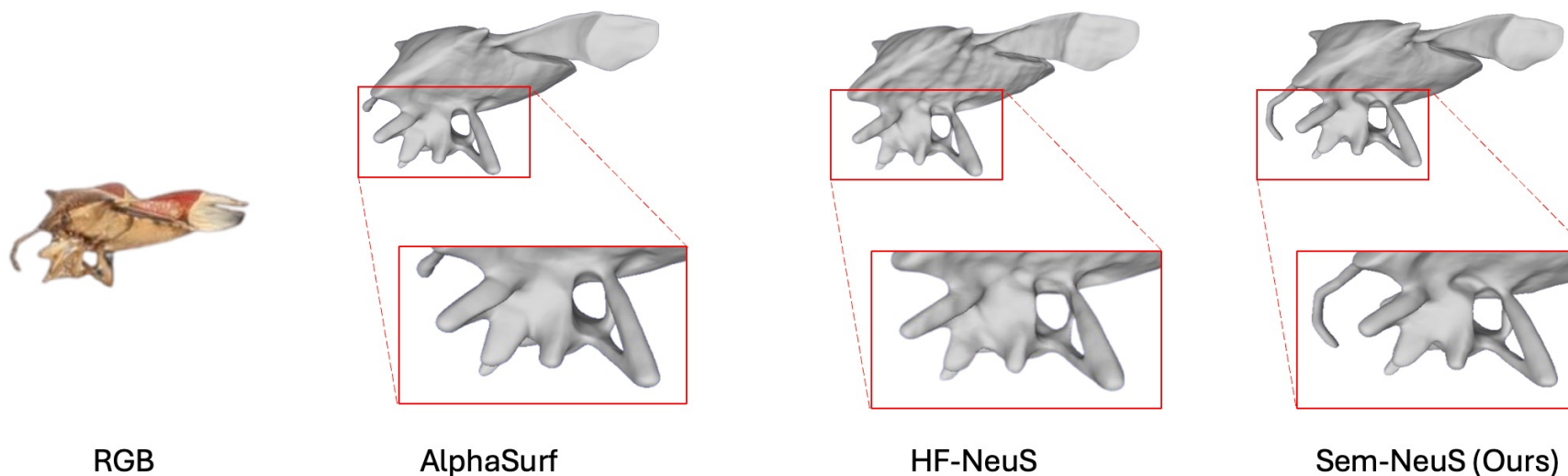
Hierarchy matters
single SDF worsens geometry
Semantics help geometry
removing L_sem increases CD to 0.006252
Direct injection hurts
CD rises to 0.012500

Rendering nuance on the custom sequence

Method	PSNR ↑	LPIPS ↓
AlphaSurf	40.20	0.0110
3DGS	41.10	0.0110
Sem-NeuS	40.50	0.0090

3DGS remains a strong rendering reference; Sem-NeuS keeps a continuous SDF surface.

Limitation: semantic consistency is not topology enforcement



Observed failure mode

Crossing thin structures

some fusion can remain at intersections

Why

2D semantic cues help localization but do not explicitly enforce 3D topology

Future work

topology-aware constraints and dynamic scenes

Conclusion and contributions

Sem-NeuS is a neural surface reconstruction method that uses semantic feature consistency to better constrain SDF geometry when photometric supervision is ambiguous.

1. Geometry distillation

DINOv3 features supervise geometry rather than entering the RGB renderer.

2. Hierarchical SDF integration

Coarse-to-fine SDF provides surface representation capacity.

Thank you



Code and dataset

Ngala kwop biddi.
Building a brighter
future, together.

