ICLR

中国科学院大学 University of Chinese Academy of Sciences

中国科学院自动化研究所 INSTITUTE OF AUTOMATION CHINESE ACADEMY OF SCIENCES

NeuralPlane: Structured 3D Reconstruction in Planar Primitives with Neural Fields

*Hanqiao Ye, Yuzhou Liu, Yangdong Liu, Shuhan Shen**

Poster Presentation: *Hall 3 + Hall 2B #78*
Session 1, Thu 24 Apr 10:00 CST — 12:30 CST

Oral Presentation: *Peridot 204-205*
Session 2F, Thu 24 April 15:30 CST — 15:42 CST

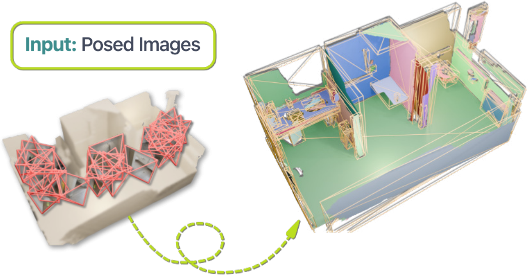
Project Page

Code

Structured Reconstruction In Planes

- Rebuilds structured (man-made) scenes as arrangements of planar primitives.

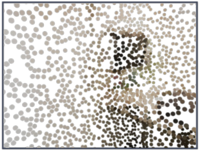
Input: Posed Images

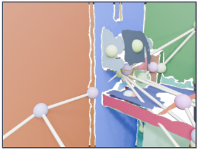


Output a set of 3D planar primitives that briefly describe both geometry and semantics of a man-made environment.

Why Planes?

A form of explicit and compact representation with rich geometric and semantic cues



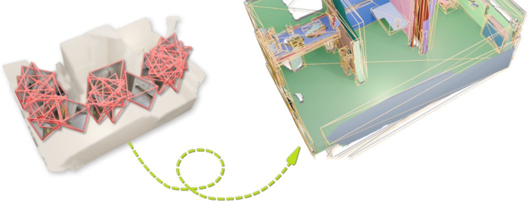


Type (Size): Mesh (~5M) Point Cloud (~0.2M) Plane Map (~0.2M)

Structured Reconstruction In Planes

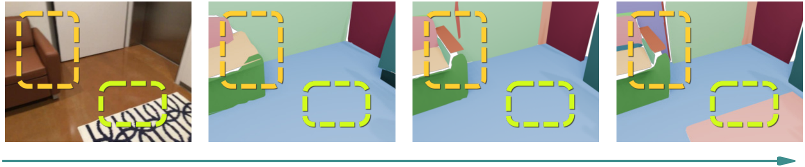
- Rebuilds structured (man-made) scenes as arrangements of planar primitives.

Input: Posed Images



Output a set of 3D planar primitives that briefly describe both geometry and semantics of a man-made environment.

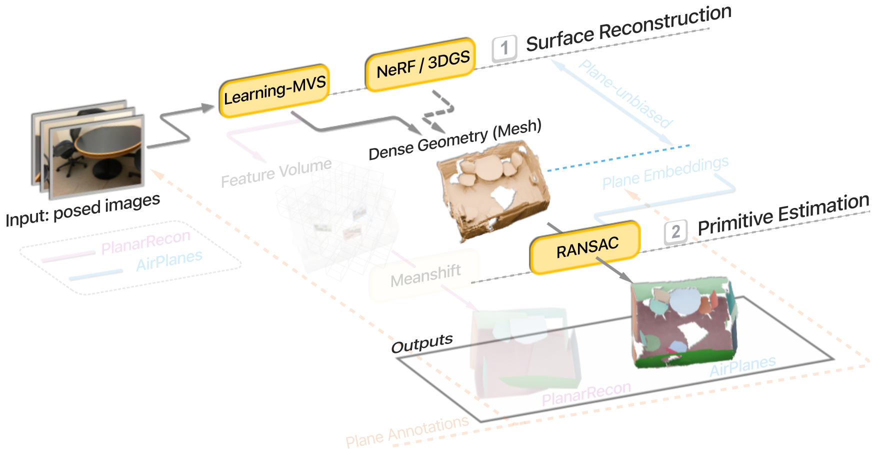
- Challenge: the **abstraction** of GEOMETRY & SEMANTICS



Motivation

- An intuitive solution: fitting planes to dense geometry, i.e., two stages of

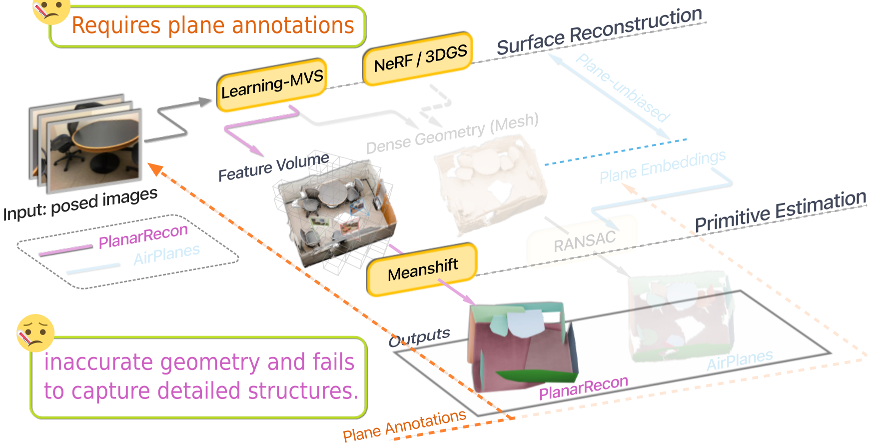
  1 Surface Reconstruction   2 Primitive Estimation



Motivation

- PlanarRecon [Xie et al, CVPR'22]:
  1 predicts plane parameters for each voxel
  2 detects plane instances by Meanshift

Requires plane annotations



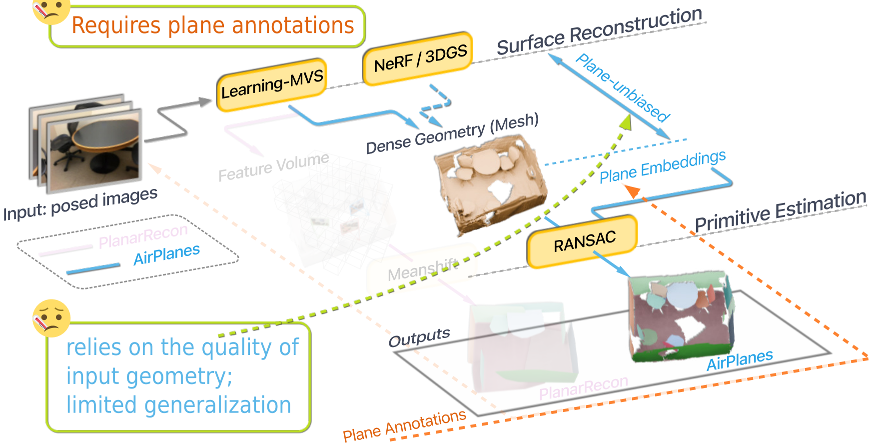
inaccurate geometry and fails to capture detailed structures.

Motivation

- AirPlanes [Watson et al, CVPR'24]: 1 off-the-shelf surface reconstruction

  2 detects plane instances by running RANSAC on geometry PLUS learned plane embeddings

Requires plane annotations

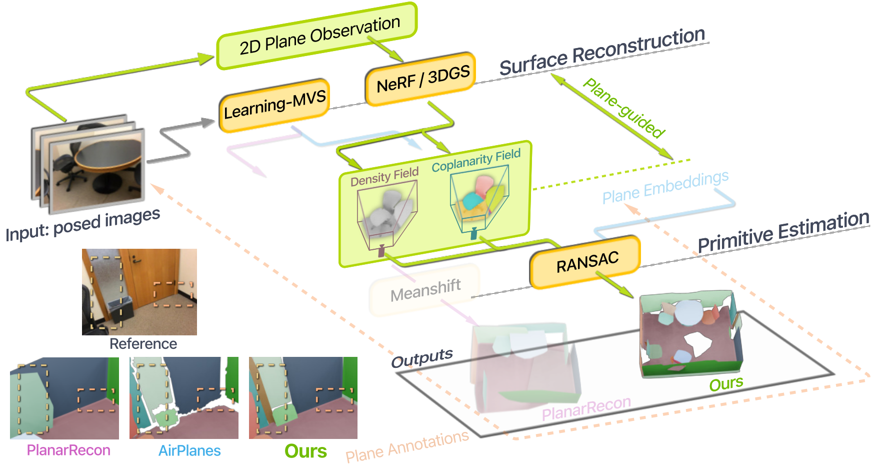


relies on the quality of input geometry; limited generalization

Motivation

Core idea: 1 detecting plane instances on 2D

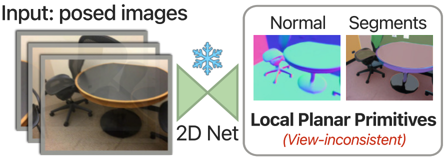
2 distilling inconsistent 2D observations into a unified 3D neural representation, which unlocks the full use of plane attributes



Methodology: Initializing Local Planar Primitives

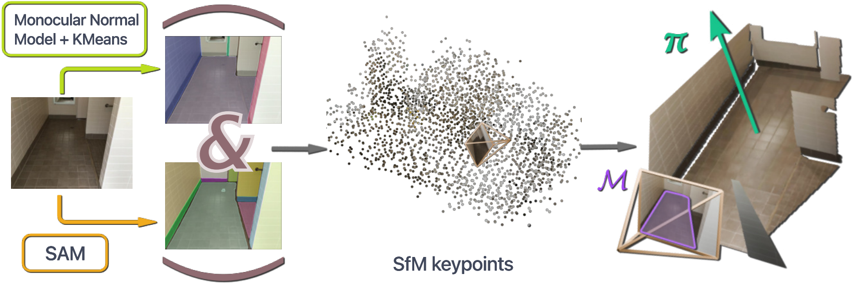
- To detect and recover spatial planar regions from each single view.

Local Planar Primitive: $P = (\mathcal{M} \subseteq I, \pi)$ 2D mask Plane parameters



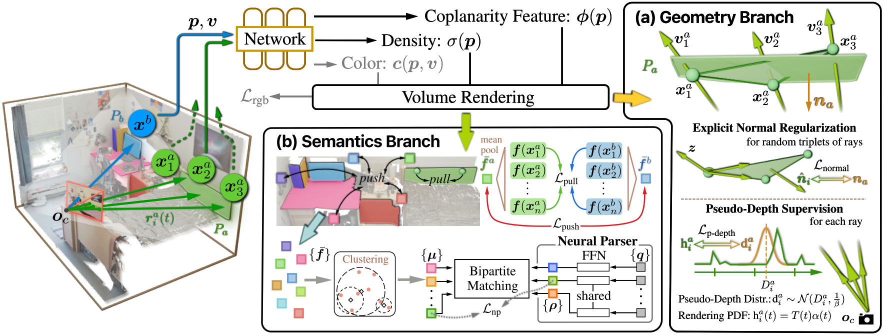
What makes ideal 2D plane observations?

1. Normal Consistency ← Monocular Normal Estimation
2. Geo. & Sem. Continuity ← Over-segmentation



Methodology: Plane-guided Neural Representation

- Local Planar Primitives are generated with little dependence on viewpoint:
  - ✗ **explicitly** establishing and merging correspondences
  - ✓ fusing them **implicitly** in the context of neural fields:



📏 **Geometry Branch** translates spatial constraints into two intra-primitive regularization terms

📚 **Semantics Branch** performs inter-primitive reasoning via contrastive learning

Methodology: Plane-guided Neural Representation

Geometry Branch

Explicit Normal Regularization

$$\mathcal{L}_{\text{normal}}(\sigma; P_a) = \mathbb{E}_{\mathcal{T}_i \sim P_a} \|1 - \hat{\boldsymbol{n}}_i^\top \boldsymbol{n}_a\|_1 .$$

NeRF-derived surface normal $\hat{\boldsymbol{n}}_i$:

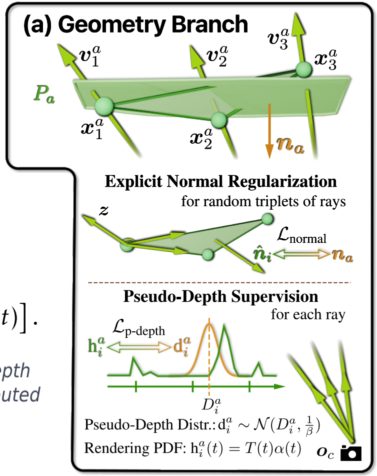
- *sample a triplet of rays from the same 2D segment;*
- *compute the normal of the plane passing through their termination points.*

Pseudo-Depth Supervision

$$\mathcal{L}_{\text{p-depth}}(\sigma, \boldsymbol{\pi}_a; P_a) = \mathbb{E}_{\boldsymbol{r}_i \sim P_a} D_{\text{KL}} \left[\mathrm{d}_i^a \| \mathrm{h}_i^a(t) \right] .$$

- *DS-NeRF[Deng et al, CVPR'22]: model the depth label d as a random variable normally distributed around the plane-derived pseudo-depth D:*

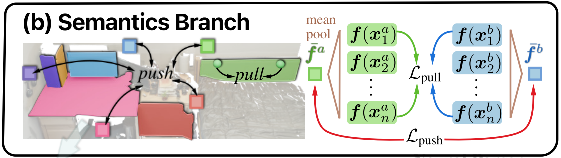
$$\mathrm{d} \sim \mathcal{N}(t; D, \beta^{-1})$$



Methodology: Plane-guided Neural Representation

- Semantics Branch

  Semantics matter: crucial to discerning distinct planes.



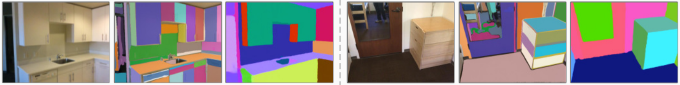
Neural Coplanarity Field (NCF)

- Decompose scenes into groups: GARField [Kim et al, CVPR'24]
- Supervise a feature field with a margin-based contrastive objective

$$\mathcal{L}_{\text{push}}(\phi; P_a, P_b) = \mathbf{1}_{[\|o_a - o_b\| > t_o \text{ or } \|\boldsymbol{n}_a \cdot \boldsymbol{n}_b\| < t_n]} \cdot \text{ReLU}(m - \|\bar{f}_a - \bar{f}_b\|_2)$$

Mitigate the issue of over-segmentation

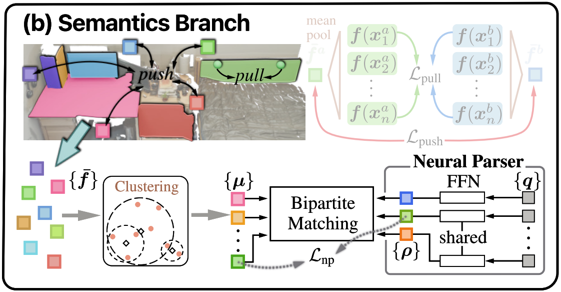
- Analysis of NCF:



Methodology: Plane-guided Neural Representation

❑ Semantics Branch

Semantics matter: crucial to discerning distinct planes.



Neural Parser (NF)

Grouping the learned NCF into instances: ContrastiveLift [Bhalgat et al, NeurIPS'23]

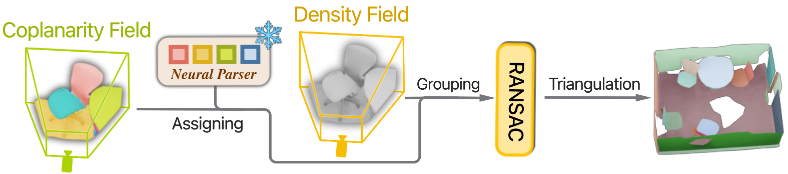
- No need to presume the exact number of instances;
- Just ensure to discern adjacent planes.

Bi. Matching in every single batch

- Assign closely located but distinct features into distinct prototypes;
- The others can be easily handled by **RANSAC**.

Results: Explicification and Quantative Evaluation

- Export Global 3D Parametric Plane Instances (See more in the Appendix)

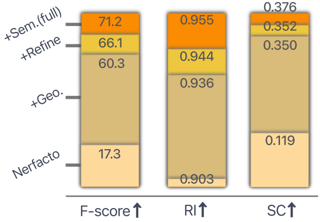


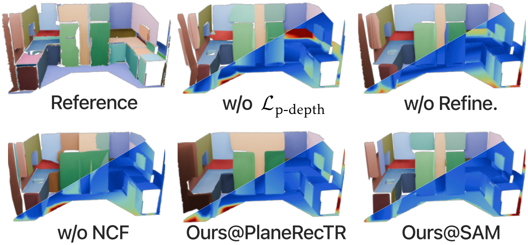
- Table: Comparison on ScanNetv2 (left) & ScanNet++ (right)

Method	Chamfer↓	F-score↑	RI↑	VOI↓	SC↑	Chamfer↓	F-score↑	RI↑	VOI↓	SC↑
PlanarRecon	9.80	49.0	0.909	3.27	0.265	14.29	43.8	0.900	3.49	0.231
AirPlanes	6.01	55.1	0.944	2.51	3.41	8.91	44.1	0.931	2.90	0.219
FineRecon	5.16	70.6	-	-	-	5.52	74.0	-	-	-
+Seq.RANSAC	5.43	66.7	0.941	2.56	0.276	5.36	75.3	0.929	2.79	0.252
+AirPlanes	5.44	66.2	0.947	2.43	0.310	5.37	75.5	0.941	2.66	0.277
NeuRIS	7.96	63.2	-	-	-	4.83	81.2	-	-	-
+Seq.RANSAC	8.11	59.3	0.945	2.57	0.293	4.84	80.9	0.941	2.46	0.315
+AirPlanes	6.17	61.0	0.943	2.55	0.291	4.69	79.9	0.943	2.53	0.287
MonoSDF	5.18	69.7	-	-	-	4.85	77.4	-	-	-
+Seq.RANSAC	5.67	65.9	0.945	2.38	0.333	5.09	77.7	0.939	2.47	0.288
+AirPlanes	5.45	66.6	0.948	2.38	0.346	5.29	74.2	0.935	2.69	0.264
Ours@PlaneRecTR	5.02	68.7	0.949	2.37	0.364	6.17	70.0	0.939	2.72	0.301
Ours@SAM	4.59	71.2	0.955	2.25	0.376	4.60	79.7	0.950	2.38	0.356

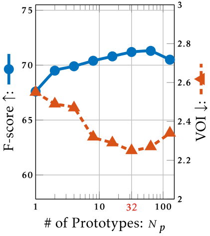
Results: Ablation Study

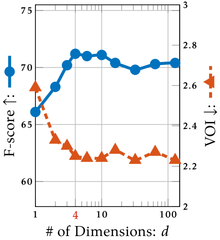
❑ Ablating Components





❑ Ablating Key Hyperparameters

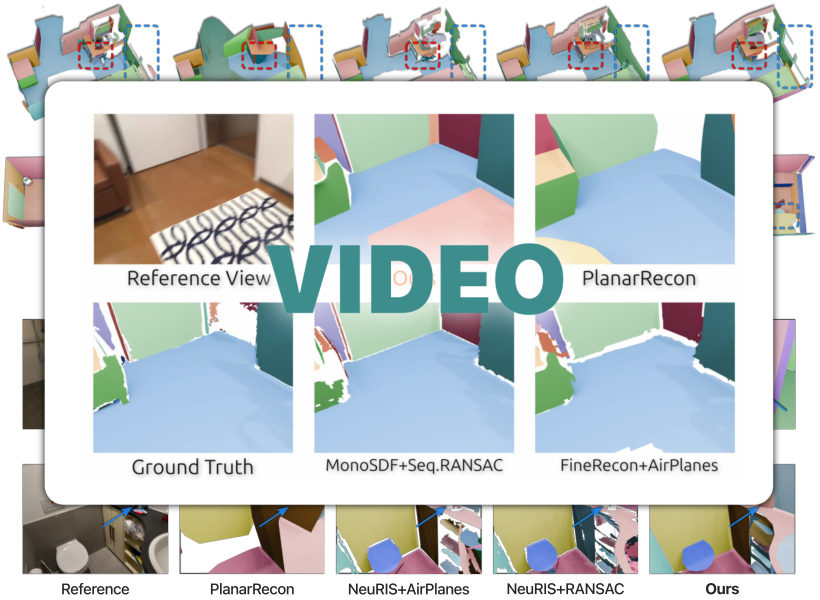




Observations

- ❑ Strikes a good **balance** between Precision and Recall. Even beats SDF-based methods.
- ❑ Geometry and semantics are tightly entangled in the task, and the combination in our method appears notable **synergy**.
- ❑ Inherently ambiguous in defining planar structures. Adjust the granularity.

Results: Visualization



NeuralPlane: Structured 3D Reconstruction in Planar Primitives with Neural Fields

Speaker: Hanqiao Ye

Code: https://github.com/3dv-casia/NeuralPlane

Thank you!

Hanqiao's Homepage

Wechat