SGRec3D

Self-Supervised Pre-training for 3D Scene Graph prediction

BOSCH [1] universität ulm [2] TECHNISCHE UNIVERSITÄT WIEN [3]

Sebastian Koch[1,2] Pedro Hermosilla[3] Narunas Vaskevicius[1] Mirco Colosi[1] Timo Ropinski[2]

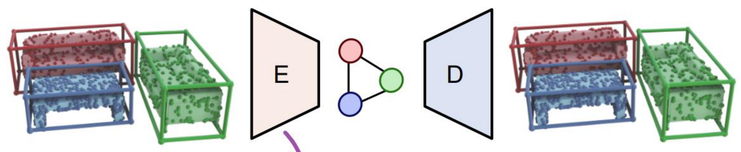
JAN 4-8 WACV 2024 WAIKOLOA HAWAII

1. Introduction

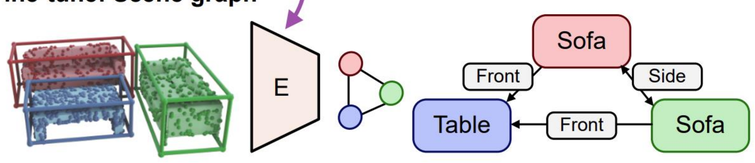
Problem: Large-scale datasets with high-quality relationship labels are scarce for 3D scene graph learning

Key Idea: Increase label efficiency by self-supervised pre-training

Pre-training: Reconstruction



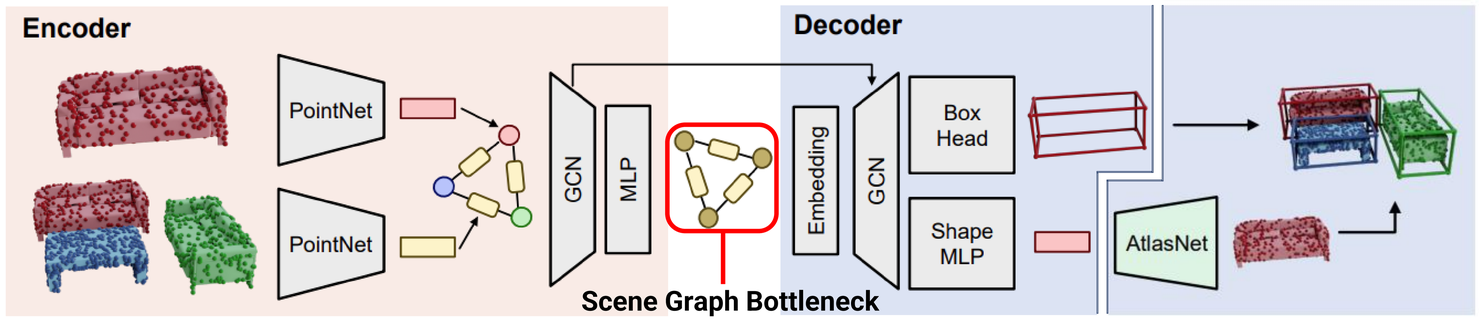
Fine-tune: Scene graph



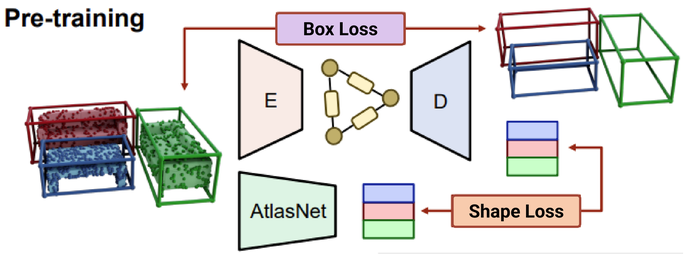
✔ No explicit scene graph labels required for pre-training

✔ Trainable on large-scale 3D datasets such as ScanNet

2 . Method



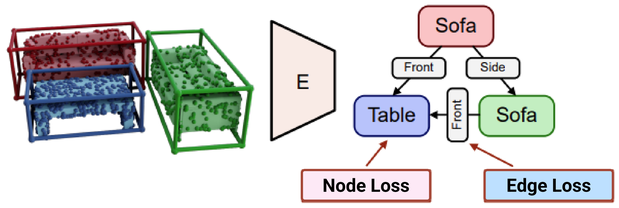
Pre-training



Use of a **bounding box loss** to learn to reconstruct the overall layout of the scene.

Use of a **latent shape loss** to learn the shapes of objects encoded by a pre-trained **AtlasNet**.

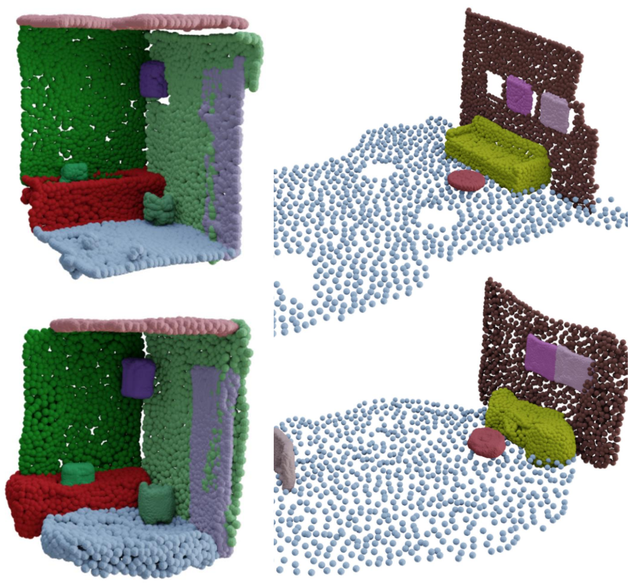
Fine-tuning



The **decoder** is discarded for fine-tuning.

The **encoder** is fine-tuned on a smaller scene graph dataset with a supervised loss for **objects** and **predicates**.

3. Scene Generations

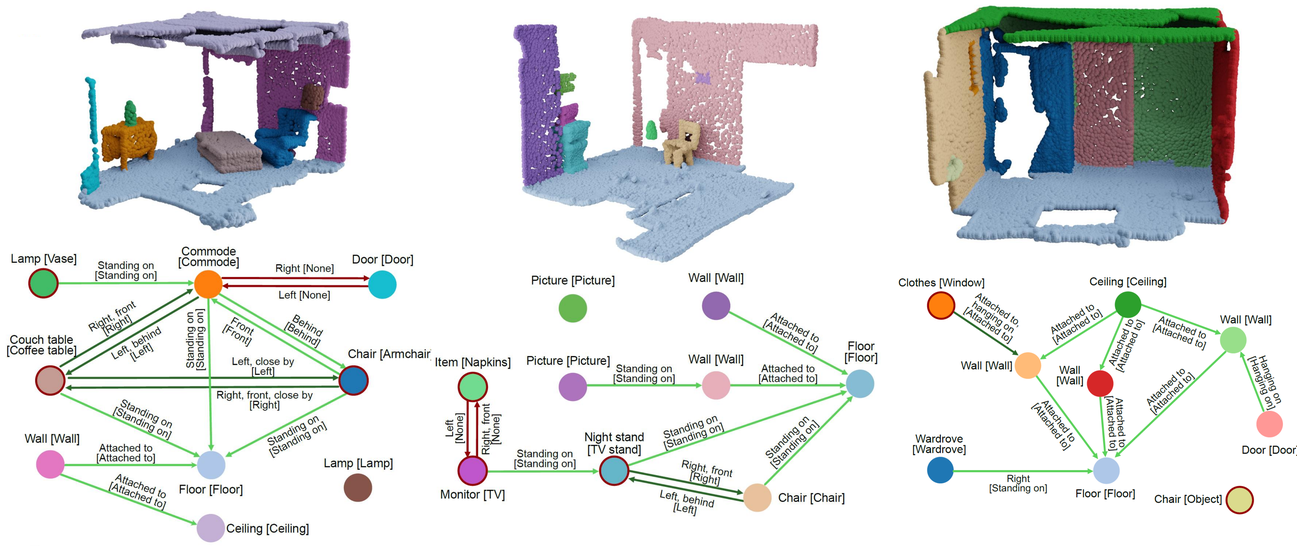


Preserved Relationships

Relationship	Graph-to-3D	Ours
left of	0.85	**0.92**
right of	0.85	**0.92**
front of	0.79	**0.90**
behind of	0.79	**0.90**
higher than	**0.96**	**0.96**
lower than	**0.96**	**0.96**
bigger than	**0.98**	0.96
smaller than	**0.98**	0.96
same as	**1.00**	**1.00**
average	0.90	**0.94**

❗ Preserved relationships are a good indication for learned relationship knowledge

4. Scene Graph Predictions



❗ 3D Scene Graph predictions contain precise object labels and detailed predicate descriptions

5. 3DSSG Evaluation

	Object		Predicate		Relationships	
	R@5	R@10	R@3	R@5	R@50	R@100
3DSSG	0.68	0.78	0.89	0.93	0.40	0.66
SGFN	0.70	0.80	**0.97**	**0.99**	0.85	0.87
SGRec3D	**0.80**	**0.87**	**0.97**	**0.99**	**0.89**	**0.91**

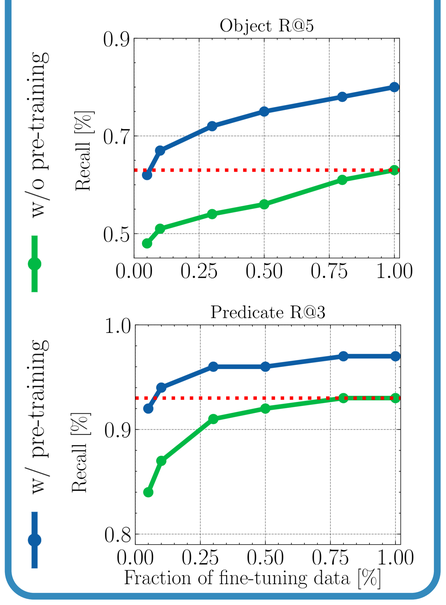
* More baseline results can be found in the paper

		Head	Body	Tail	All
Objects	w/o pre-train	0.88	0.45	0.06	0.30
	w/ pre-train	**0.92**	**0.78**	**0.24**	**0.45**
Predicates	w/o pre-train	0.94	0.83	0.41	0.57
	w/ pre-train	**0.97**	**0.96**	**0.65**	**0.69**

❗ Pre-trained model outperforms baselines by a large margin

❗ Pre-training is especially effective for rare classes

7. Label-efficiency



8. Pre-training strategy

	Pre-train			Object		Predicate	
	GCN	PCL	SG	R@5	mR@5	R@3	mR@3
STRL		✓		0.75	0.35	0.94	0.50
STRL	✓	✓		0.63	0.23	0.92	0.48
DepthContrast		✓		0.77	0.36	0.94	0.51
DepthContrast	✓	✓		0.60	0.22	0.93	0.50
Ours (no pre-train)	✓			0.63	0.30	0.94	0.57
Ours (no GCN)			✓	0.75	0.31	0.94	0.48
Ours	✓		✓	**0.80**	**0.45**	**0.97**	**0.69**

PCL: point cloud pre-training — SG: scene graph pre-training

❗ Point cloud-based pre-trainings are ineffective for 3D scene graphs

❗ Our SGRec3D scene graph pre-training is very effective