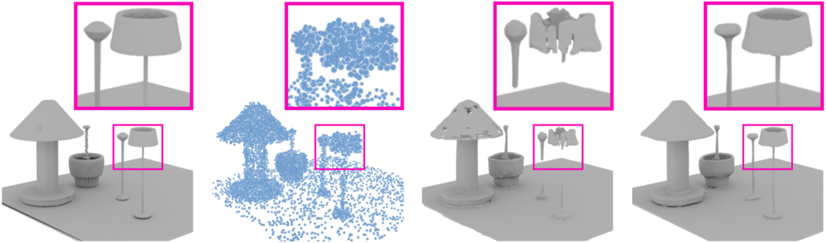
ALTO: Alternating Latent Topologies for Implicit 3D Reconstruction

Zhen Wang[1*] Shijie Zhou[1*] Jeong Joon Park[2] Despoina Paschalidou[2]
Suya You[3] Gordon Wetzstein[2] Leonidas Guibas[2] Achuta Kadambi[1]

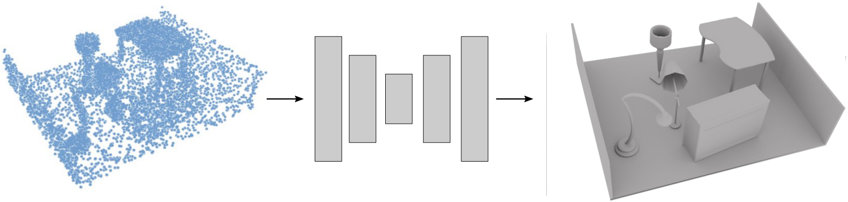
[1] University of California, Los Angeles [2] Stanford University
[3] DEVCOM Army Research Laboratory

https://visual.ee.ucla.edu/alto.htm/



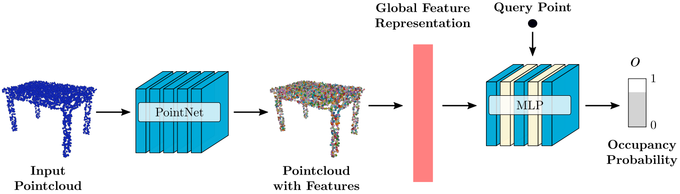
(a) Ground Truth (b) Input (10,000 points) (c) POCO [3] (36.8 sec) (d) **ALTO (10.5 sec)**

Problem Statement

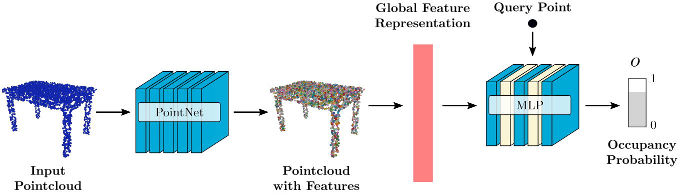


Can we recover 3D geometries of high fidelity given a (noisy) pointcloud as input?

Implicit Neural Representations for 3D Reconstructions

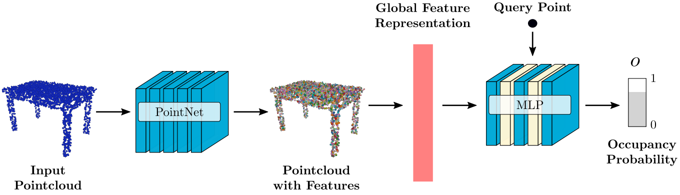


Implicit Neural Representations for 3D Reconstructions



✓ Results in **continuous representations**

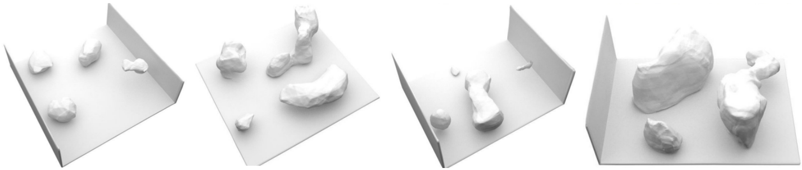
Implicit Neural Representations for 3D Reconstructions



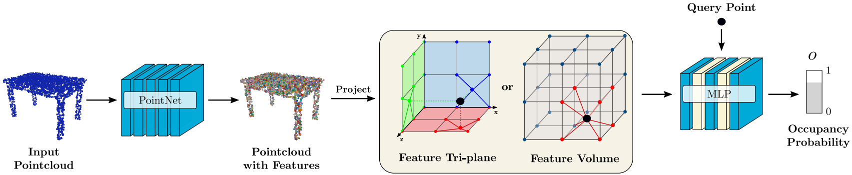
✓ Results in **continuous representations**

✕ The **global latent code** yields **overly smooth geometries**

✕ Fully connected layers are not **translation equivariant**

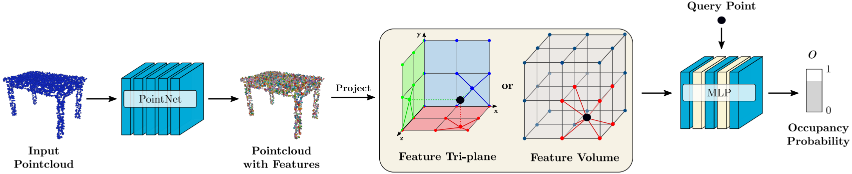


Implicit Neural Representations for 3D Reconstructions



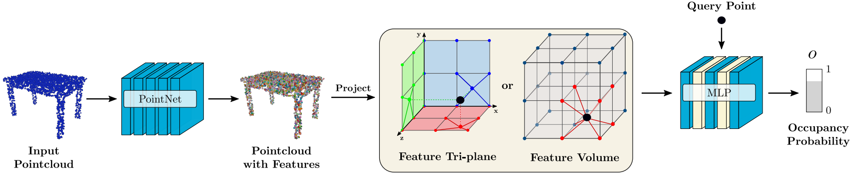
Peng: Convolutional Occupancy Networks. ECCV, 2020.

Implicit Neural Representations for 3D Reconstructions



✓ Utilizing **local features** recovers more detailed geometries

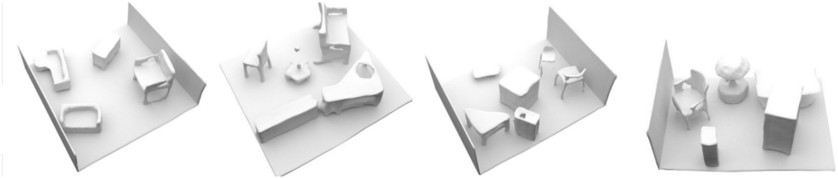
Implicit Neural Representations for 3D Reconstructions



✓ Utilizing **local features** recovers more detailed geometries

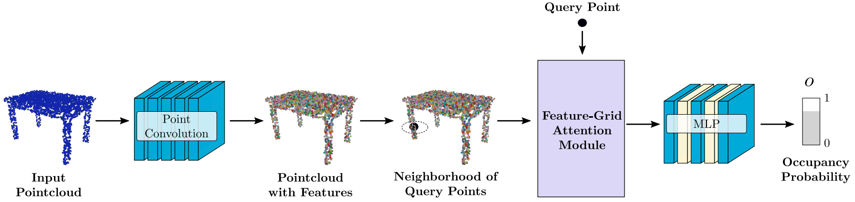
✗ Latent vectors are **uniformly distributed in space**

✗ Struggles to capture **fine-grained geometries around the surface boundaries**

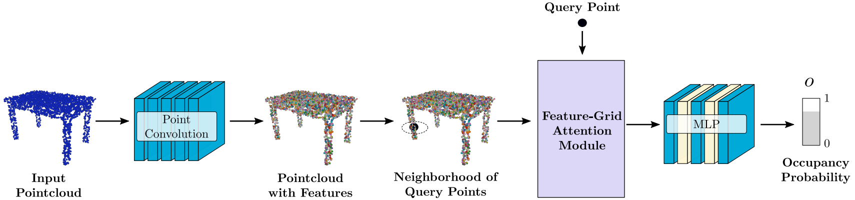


Peng: Convolutional Occupancy Networks. ECCV, 2020.

Implicit Neural Representations for 3D Reconstructions

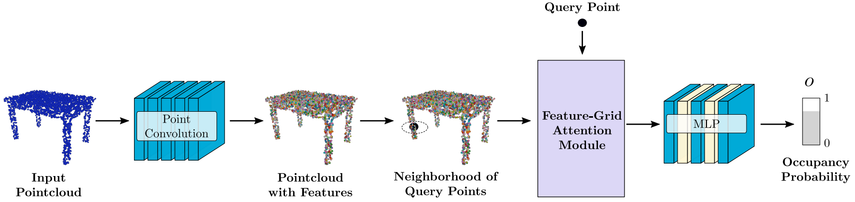


Implicit Neural Representations for 3D Reconstructions

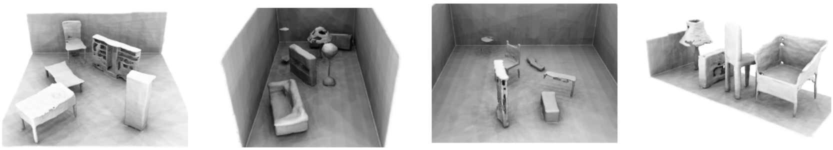


✓ The latent vectors are **concentrated around the surface boundaries**

Implicit Neural Representations for 3D Reconstructions

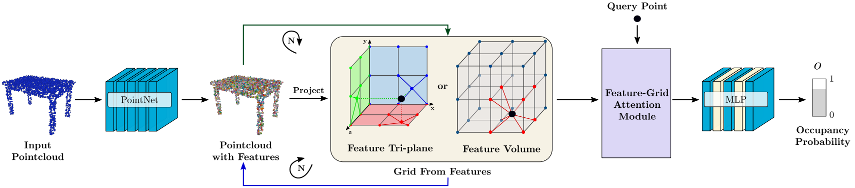


- ✓ The latent vectors are **concentrated around the surface boundaries**
- ✕ Very slow **inference time**
- ✕ Struggles to capture **fine-grained geometries**

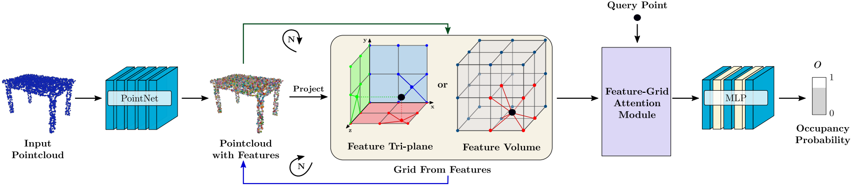


Boulch: POCO: Point Convolution for Surface Reconstruction. CVPR, 2022.

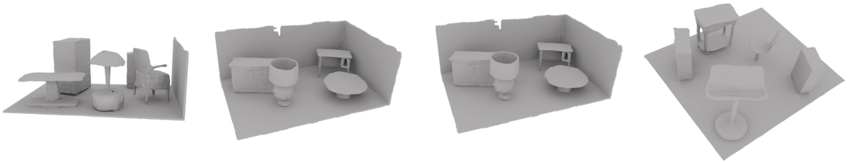
Our Implicit Neural Representation for 3D Reconstructions



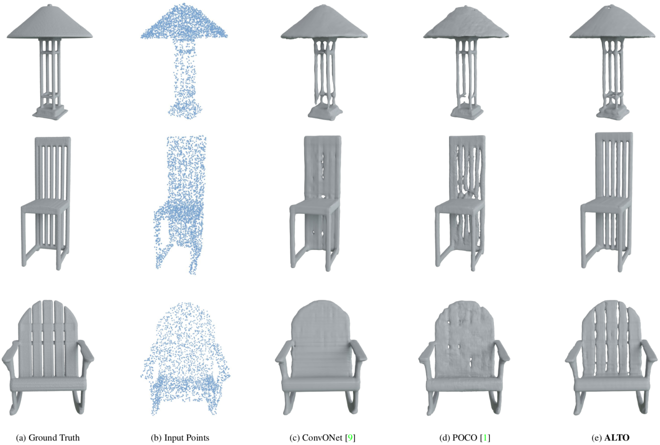
Our Implicit Neural Representation for 3D Reconstructions



- ✓ Utilizing **local features** recovers more detailed geometries
- ✓ Our model can reconstruct a 3D scene **up to** $10\times$ **faster**
- ✓ Can capture **fine-grained geometries**

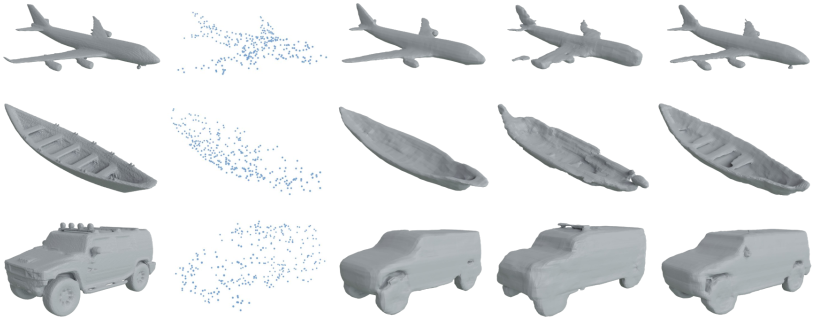


Object-Level Reconstruction on ShapeNet



Object-level reconstructions using 3k points as input

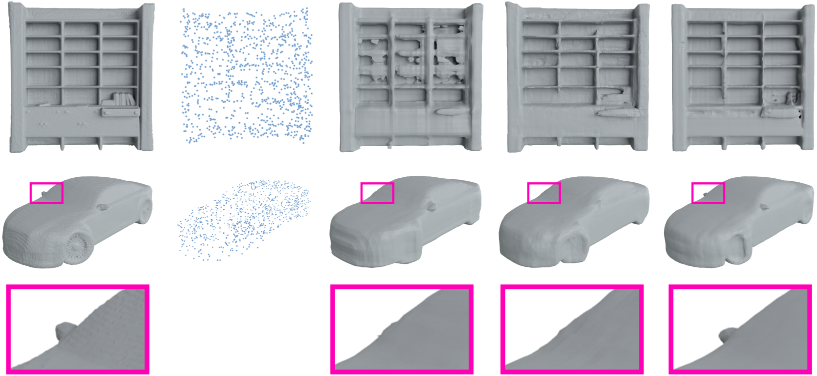
Object-Level Reconstruction on ShapeNet



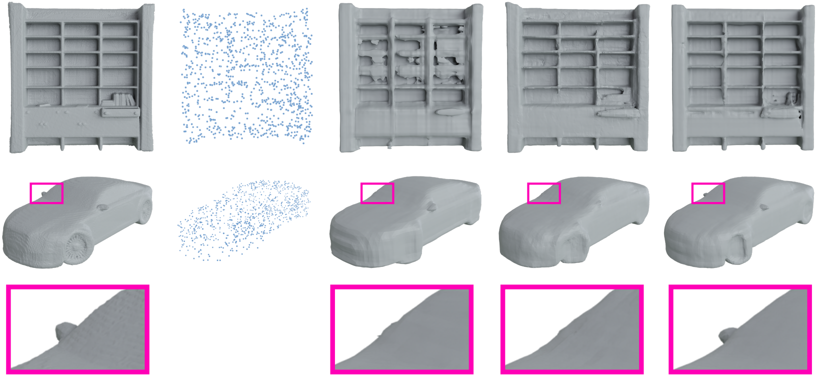
(a) Ground Truth (b) Input Points (c) ConvONet [9] (d) POCO [1] (e) **ALTO**

Object-level reconstructions using 300 points as input

Attention to Detail

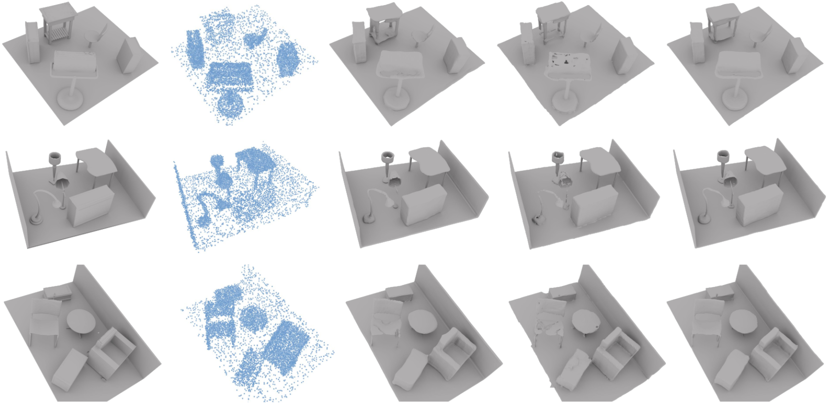


Object-Level Reconstruction on ShapeNet



Method	Input points 3K				Input points 1K				Input points 300			
	IoU ↑	Chamfer-L_1 ↓	NC↑	F-score↑	IoU ↑	Chamfer-L_1 ↓	NC↑	F-score↑	IoU ↑	Chamfer-L_1 ↓	NC↑	F-score↑
ONet [41]	0.761	0.87	0.891	0.785	0.772	0.81	0.894	0.801	0.778	0.80	0.895	0.806
ConvONet [49]	0.884	0.44	0.938	0.942	0.859	0.50	0.929	0.918	0.821	0.59	0.907	0.883
POCO [3]	0.926	**0.30**	0.950	**0.984**	0.884	0.40	0.928	0.950	0.808	0.61	0.892	0.869
ALTO	**0.930**	**0.30**	**0.952**	0.980	**0.905**	**0.35**	**0.940**	**0.964**	**0.863**	**0.47**	**0.922**	**0.924**

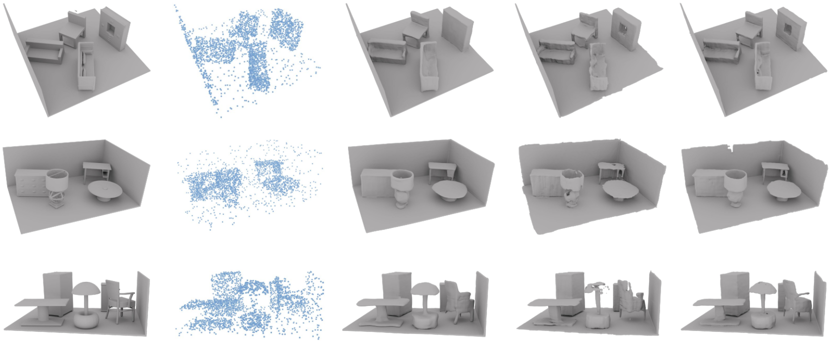
Scene-Level Reconstruction on Synthetic Rooms



(a) Ground Truth (b) Input Points (c) ConvONet [9] (d) POCO [1] (e) **ALTO**

Scene-level reconstructions using 10k points as input

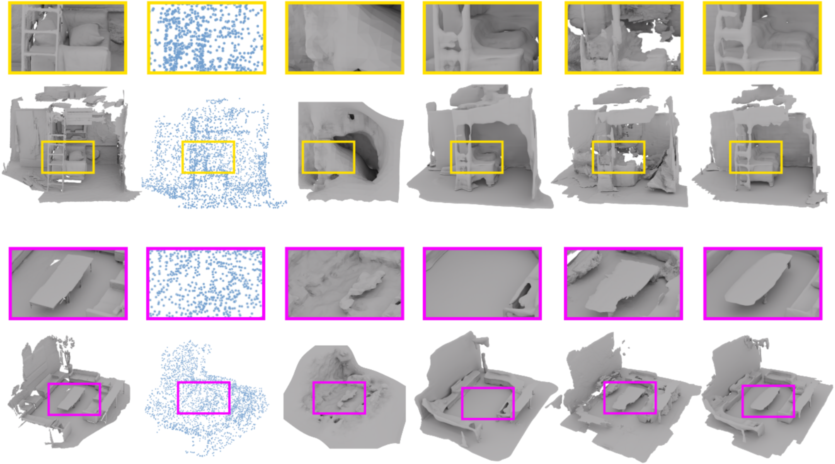
Scene-Level Reconstruction on Synthetic Rooms



(a) Ground Truth (b) Input Points (c) ConvONet [9] (d) POCO [1] (e) **ALTO**

Scene-level reconstructions using 3k points as input

Generalization on ScanNet-v2



(a) Ground Truth (b) Input Points (c) SPSR [33] (d) ConvONet [49] (e) POCO [3] (f) **ALTO**

Scene-Level Reconstruction on Synthetic Rooms

Method	IoU ↑	Chamfer-L_1 ↓	NC↑	F-score↑
ONet [41]	0.475	2.03	0.783	0.541
SPSR [33]	-	2.23	0.866	0.810
SPSR trimmed [33]	-	0.69	0.890	0.892
ConvONet [49]	0.849	0.42	0.915	0.964
DP-ConvONet [37]	0.800	0.42	0.912	0.960
POCO [3]	0.884	0.36	0.919	0.980
ALTO	**0.914**	**0.35**	**0.921**	**0.981**

Quantitative Evaluation on Synthetic Room Dataset using 10k points as input

Scene-Level Reconstruction on Synthetic Rooms

Method	IoU ↑	Chamfer-L_1 ↓	NC↑	F-score↑
ONet [41]	0.475	2.03	0.783	0.541
SPSR [33]	-	2.23	0.866	0.810
SPSR trimmed [33]	-	0.69	0.890	0.892
ConvONet [49]	0.849	0.42	0.915	0.964
DP-ConvONet [37]	0.800	0.42	0.912	0.960
POCO [3]	0.884	0.36	0.919	0.980
ALTO	**0.914**	**0.35**	**0.921**	**0.981**

Quantitative Evaluation on Synthetic Room Dataset using 10k points as input

	N_{Train}=10K, N_{Test}=3K		N_{Train}=N_{Test}=3K	
Method	Chamfer-L_1 ↓	F-score↑	Chamfer-L_1 ↓	F-score↑
ConvONet [49]	1.01	0.719	1.16	0.669
POCO [3]	0.93	0.737	1.15	0.667
ALTO	**0.87**	**0.746**	**0.92**	**0.726**

Generalization Capability on ScanNet

Scene-Level Reconstruction on Synthetic Rooms

Method	IoU ↑	Chamfer-L_1 ↓	NC↑	F-score↑
ONet [41]	0.475	2.03	0.783	0.541
SPSR [33]	-	2.23	0.866	0.810
SPSR trimmed [33]	-	0.69	0.890	0.892
ConvONet [49]	0.849	0.42	0.915	0.964
DP-ConvONet [37]	0.800	0.42	0.912	0.960
POCO [3]	0.884	0.36	0.919	0.980
ALTO	**0.914**	**0.35**	**0.921**	**0.981**

Quantitative Evaluation on Synthetic Room Dataset using 10k points as input

	N_{Train}=10K, N_{Test}=3K		N_{Train}=N_{Test}=3K	
Method	Chamfer-L_1 ↓	F-score↑	Chamfer-L_1 ↓	F-score↑
ConvONet [49]	1.01	0.719	1.16	0.669
POCO [3]	0.93	0.737	1.15	0.667
ALTO	**0.87**	**0.746**	**0.92**	**0.726**

Generalization Capability on ScanNet

Method	# Parameters	Inference time (s)
ConvONet [49]	4,166,657	1.6
POCO [3]	12,790,454	36.1
ALTO	4,787,905	3.6

Runtime Comparison

Thank you for your attention!