Multi-Object Occupancy Networks

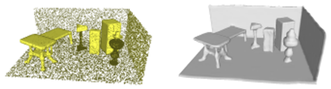
Gianluca Galletti[1] Nina Kirakosyan[1]

[1]Technical University of Munich

Motivation

Goal: Robust 3D reconstruction of multi object scenes from point clouds

Some current methods fail to capture independent object properties.
▹ Not leveraging semantic/instance information.
▹ Missing object-wise mesh reconstruction.



Our aim: directly predict object-wise occupancy.
▹ Scale up to object-wise scene reconstruction.
▹ Allow cheap object transformations in the reconstructed space.

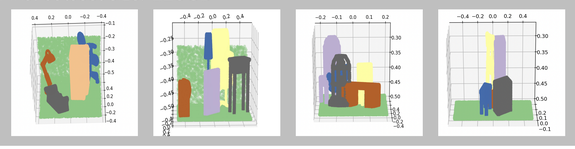
Semantic Synthetic Rooms Dataset

Custom instance-labeled dataset

▹ **Reference point**: Synthetic Rooms Dataset from ONet
▹ **Why?** *Instance* vs *Semantic* segmentation

Dataset Structure:
- Objects of 5 classes from ShapeNet: *chair, table, sofa, lamp, cabinet*.
- Rooms: 4-8 objects per room, 1000 rooms per each object number.
- 10 distinct sampled point clouds for each room.

▹ Total number of training inputs: **50k**
▹ Final dataset size: **223GB**



Components

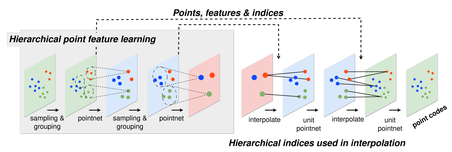
Instance segmentation

SoftGroup 3D Instance Segmentation on Point Clouds
▹ Integrated Synthetic Rooms Dataset into the SoftGroup pipeline.
▹ Data feature and format limitations to be addressed in future work.

Encoders

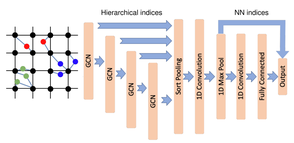
MO-DGCNN: Dynamic Graph CNN extended to *tag-indexed* point clouds.

MO-PointNet++: PointNet++ adapted to *tag-indexed* point clouds.

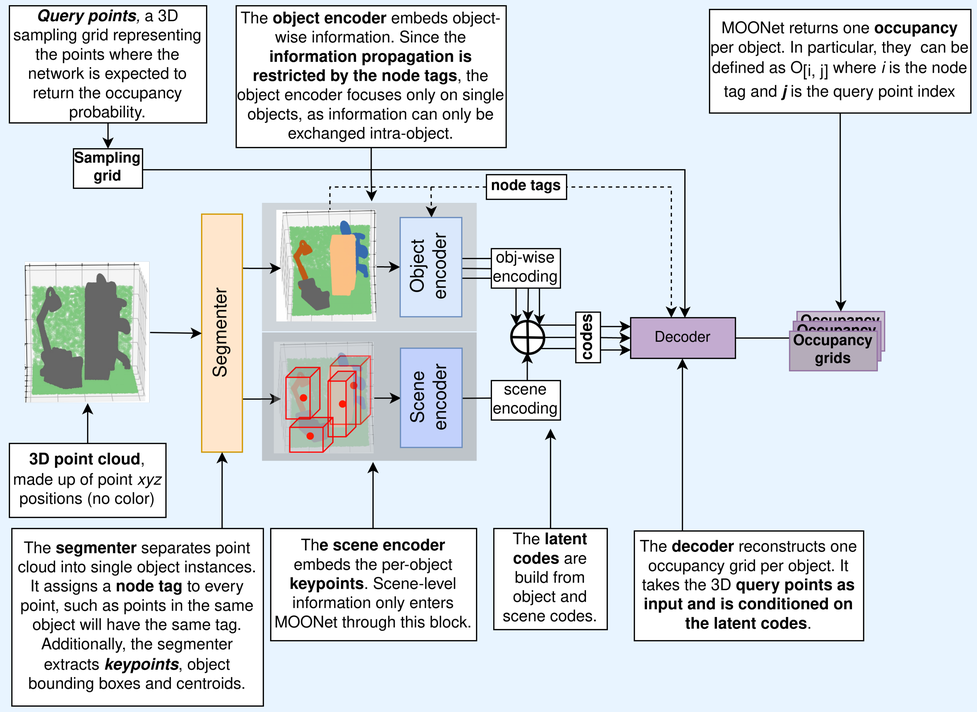


Decoder

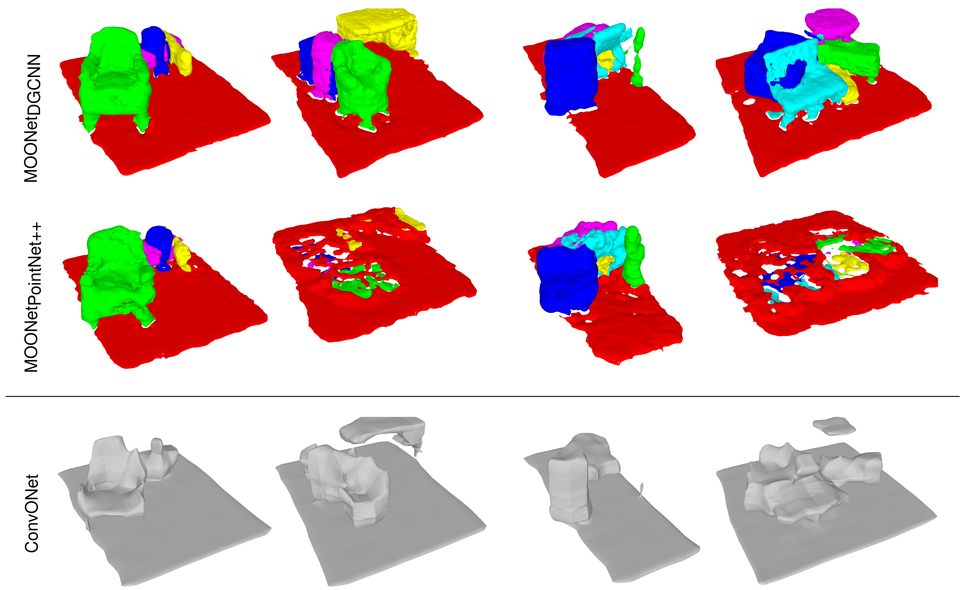
MO-grid-DGCNN: both point cloud and sample grid; multi-object output.



Overview



Scene reconstruction



Results

IoU against baselines

Model	Inference [ms]	Iou@50	IoU@90	IoU@95
Onet	45	0.36	0.30	0.27
ConvONet	**27**	**0.61**	0.44	0.40
MOONet$_{(DGCNN)}$ (ours)	680	0.35	**0.49**	**0.54**
MOONet$_{(PointNet++)}$ (ours)	482	0.33	0.38	0.38
Fake MOONet$_{(DGCNN)}$ (ours)	-	0.32	0.44	0.48

Semantic transformations

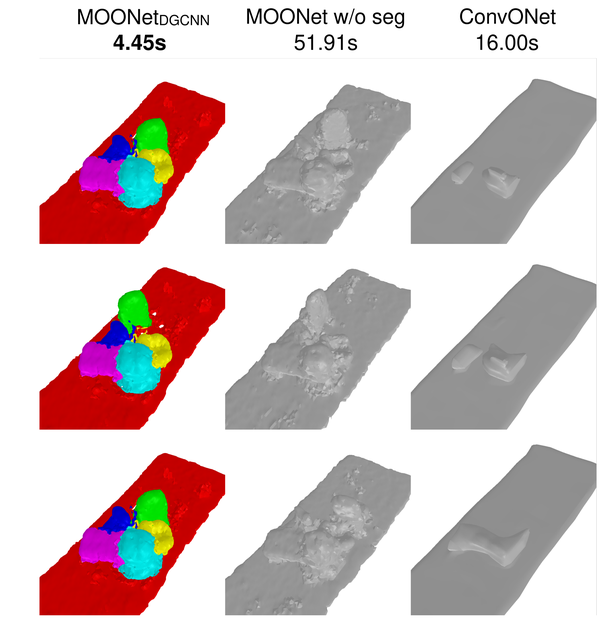


Figure 1. Scene reconstructions with object transformations (translations). The times are reported for 50 reconstructions with different object positions. *"MOONet w/o seg"* treats the entire scene as a single object.

Conclusions

We present MOONets, a novel set of architectures for direct multi-object 3D reconstruction from point clouds, which naturally produce a segmented, object-wise mesh. Abeit not outperforming the previous state of the art, we show two advantages of our architecture, namely: 1) the core advantage of a direct multi-object mesh is the possibility to apply geometric transformations to the object cheaply, since only one forward pass per scene is needed; 2) we observe that introducing semantic information, such as the instance segmentation, can improve performance with little added cost. We suggest that future works in this direction could explore a) dedicated multi-object encoders and decoders; b) better segmentation networks; c) even faster multi-mesh generation.

github.com/gerkone/multiobject-onet

g.galletti@tum.de, nina.kirakosyan@tum.de