OGC: Unsupervised 3D Object Segmentation from Rigid Dynamics of Point Clouds

Ziyang Song, Bo Yang

vLAR Group, The Hong Kong Polytechnic University

Introduction

Our task: 3D object segmentation from point clouds

Fully-supervised methods:

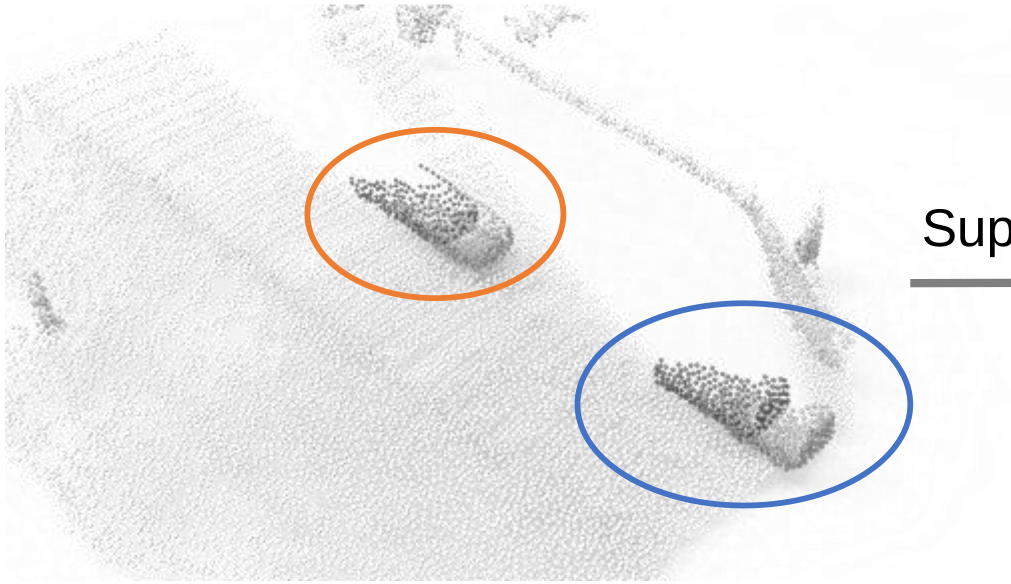
- Costly human annotations
- Limited generalization

→ **Our goal:**
Unsupervised 3D object segmentation

Introduction

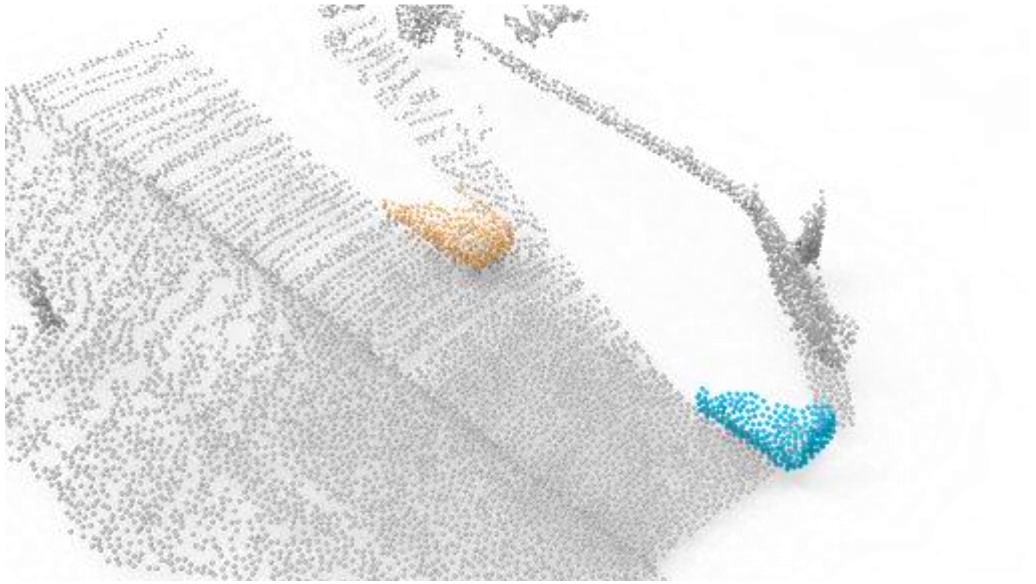
Gestalt theory: *The raw sensory data with similar motion are likely to be organized into a single object* [1]

dynamic motions in
point cloud sequences



Supervise **???**

3D object segmentation



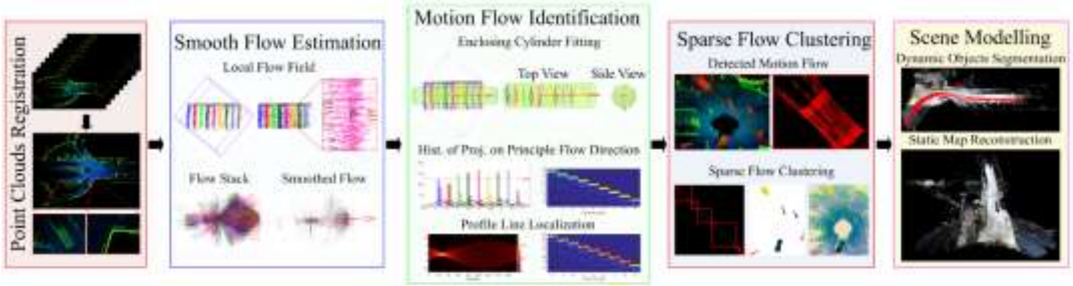
[1] J. Wagemans, J. H. Elder, M. Kubovy, et al. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. Psychological Bulletin, 138(6):1172–1217, 2012.

Introduction

Prior works & Limitations

Subspace Clustering (BMVC'18) [1]

3D-MOD (TITS'21) [2]



SLIM (ICCV'21) [3]

Limitations:

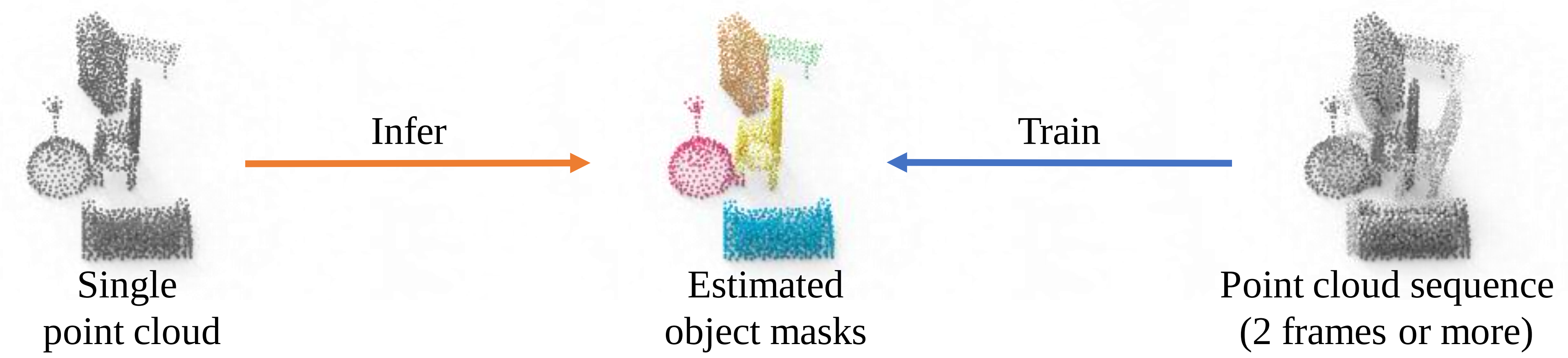
① Only for specific scenarios
② Binary segmentation
③ Needing multiple frames in inference

[1] U. M. Nunes and Y. Demiris. 3D motion segmentation of articulated rigid bodies based on RGB-D data. BMVC, 2018.
[2] C. Jiang, D. P. Paudel, D. Fofi, et al. Moving Object Detection by 3D Flow Field Analysis. TITS, 22(4):1950–1963, 2021.
[3] S. A. Baur, D. J. Emmerichs, F. Moosmann, et al. SLIM: Self-Supervised LiDAR Scene Flow and Motion Segmentation. ICCV, 2021.

Introduction

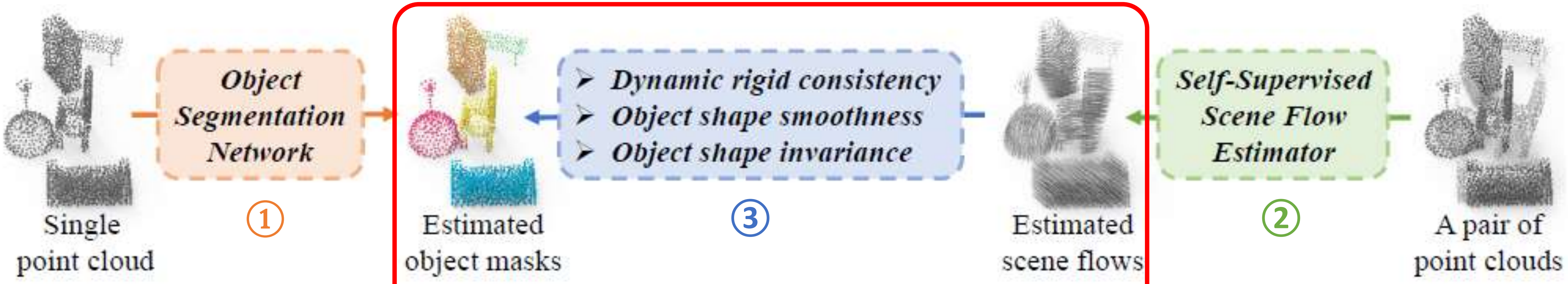
Our goal:

- A **general** framework
- **Multi**-object segmentation
- Learning from **unlabeled** sequences
- Inferring on **single** point clouds

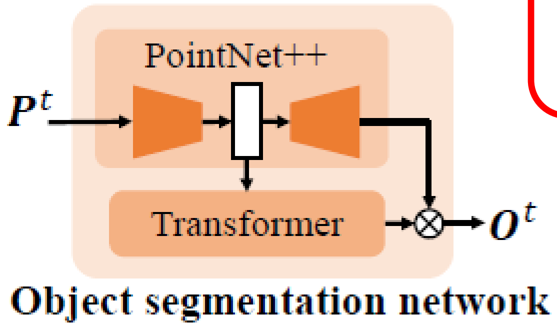


OGC

OGC (**O**bject **G**eometry **C**onsistency)



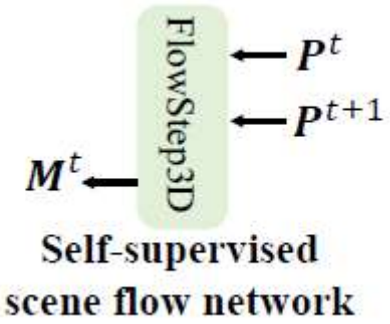
1 3D Extension of MaskFormer [1]



3 ???

How to supervise with motions

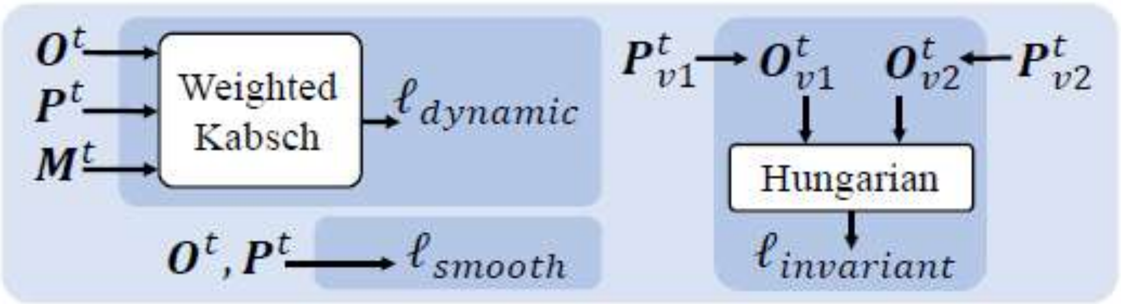
2 FlowStep3D [2]



[1] B. Cheng, A. G. Schwing, and A. Kirillov. Per-Pixel Classification is Not All You Need for Semantic Segmentation. NeurIPS, 2021.
[2] Y. Kittenplon, Y. C. Eldar, and D. Raviv. FlowStep3D: Model Unrolling for Self-Supervised Scene Flow Estimation. CVPR, 2021.

OGC

OGC (**O**bject **G**eometry **C**onsistency) **Losses**



Object geometry consistency losses

① Dynamic rigid consistency:
Motions within each object are rigid

$$\ell_{dynamic} = \frac{1}{N}\Big\|\Big(\sum_{k=1}^{K} \boldsymbol{o}_k^t * (\boldsymbol{T}_k \circ \boldsymbol{p}^t)\Big) - (\boldsymbol{p}^t + \boldsymbol{m}^t)\Big\|_2$$

② Smoothness regularization:
Objects are spatially continual

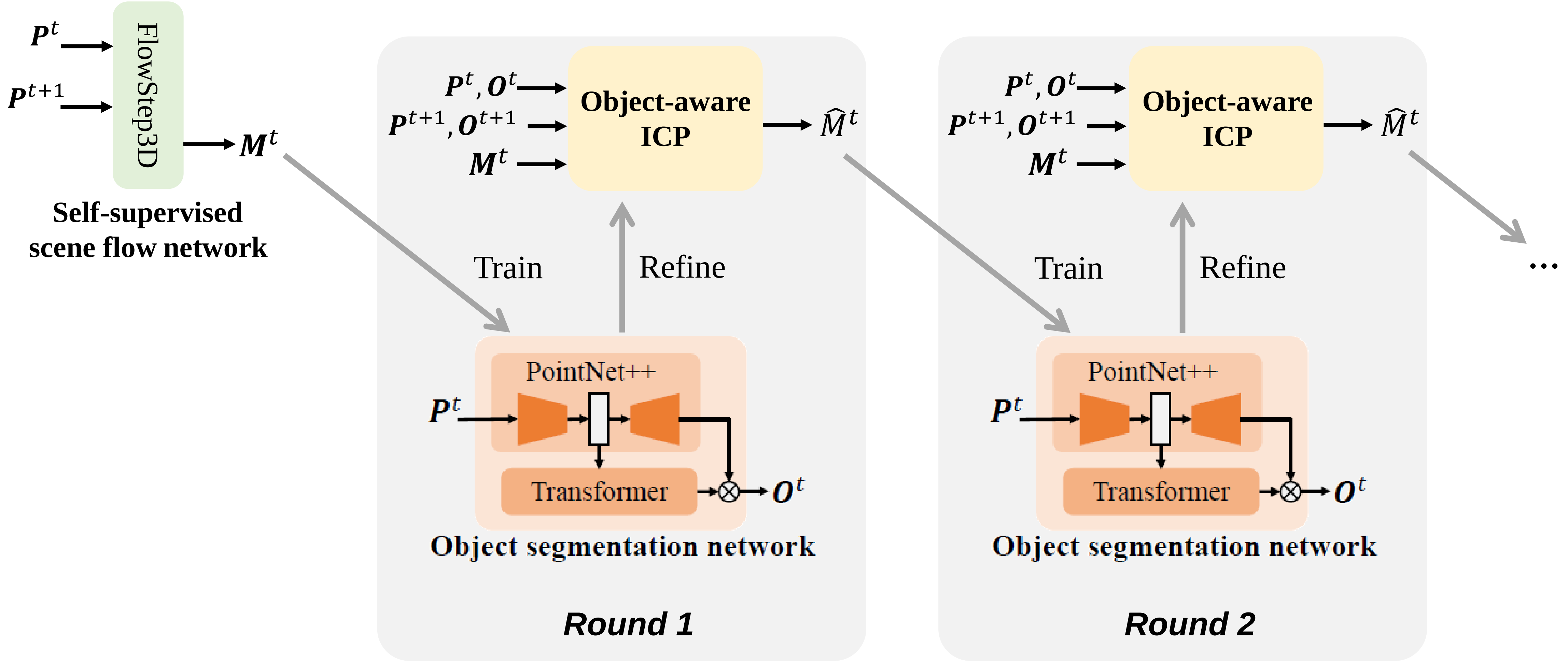
$$\ell_{smooth} = \frac{1}{N}\sum_{n=1}^{N}\Big(\frac{1}{H}\sum_{h=1}^{H} d(\boldsymbol{o}_{p_n}, \boldsymbol{o}_{p_n^h})\Big)$$

③ Invariance to spatial transformations:
Generalize to static objects

$$\ell_{invariant} = \frac{1}{N}\sum_{n=1}^{N} \hat{d}(\hat{\boldsymbol{o}}_{v1}^n, \hat{\boldsymbol{o}}_{v2}^n)$$

OGC

Iterative optimization of object segmentation and motion estimation



Experiments

Part instance segmentation on (articulated) objects: SAPIEN dataset

		AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
Supervised Methods	PointNet++ [57]	-	-	-	-	-	51.2	65.0
	MeteorNet [43]	-	-	-	-	-	45.7	60.0
	DeepPart [80]	-	-	-	-	-	53.0	67.0
	MBS [27]	-	-	-	-	-	67.3	77.0
	OGC_{sup}	66.1	48.7	62.0	54.6	71.7	66.8	77.1
Unsupervised Motion Segmentation	TrajAffn [52]	6.2	14.7	22.0	16.3	34.0	45.7	60.1
	SSC [51]	9.5	20.4	28.2	20.9	43.5	50.6	65.9
Unsupervised Methods	WardLinkage [30]	17.4	26.8	40.1	36.9	43.9	49.4	62.2
	DBSCAN [17]	6.3	13.4	20.4	13.9	37.9	34.2	51.4
	OGC(Ours)	**55.6**	**50.6**	**65.1**	**65.0**	**65.2**	**60.9**	**73.4**

Experiments

Object segmentation in indoor scenes: OGC-DR / OGC-DRSV datasets

		AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
Supervised Method	OGC_{sup}	90.7 / 86.3	82.6 / 78.8	87.6 / 85.0	83.7 / 82.2	92.0 / 88.0	89.2 / 83.9	97.7 / 97.1
Unsupervised Motion Segmentation	TrajAffn [52]	42.6 / 39.3	46.7 / 43.8	57.8 / 54.8	69.6 / 63.0	49.4 / 48.4	46.8 / 45.9	80.1 / 77.7
	SSC [51]	74.5 / 70.3	79.2 / 75.4	84.2 / 81.5	92.5 / 89.6	77.3 / 74.7	74.6 / 70.8	91.5 / 91.3
Unsupervised Methods	WardLinkage [30]	72.3 / 69.8	74.0 / 71.6	82.5 / 80.5	**93.9 / 91.8**	73.6 / 71.7	69.9 / 67.2	94.3 / 93.3
	DBSCAN [17]	73.9 / 71.9	76.0 / 76.3	81.6 / 81.8	85.8 / 79.1	77.8 / 84.8	74.7 / 80.1	91.5 / 93.5
	OGC(Ours)	**92.3 / 86.8**	**85.1 / 77.0**	**89.4 / 83.9**	85.6 / 77.7	**93.6 / 91.2**	**90.8 / 84.8**	**97.8 / 95.4**

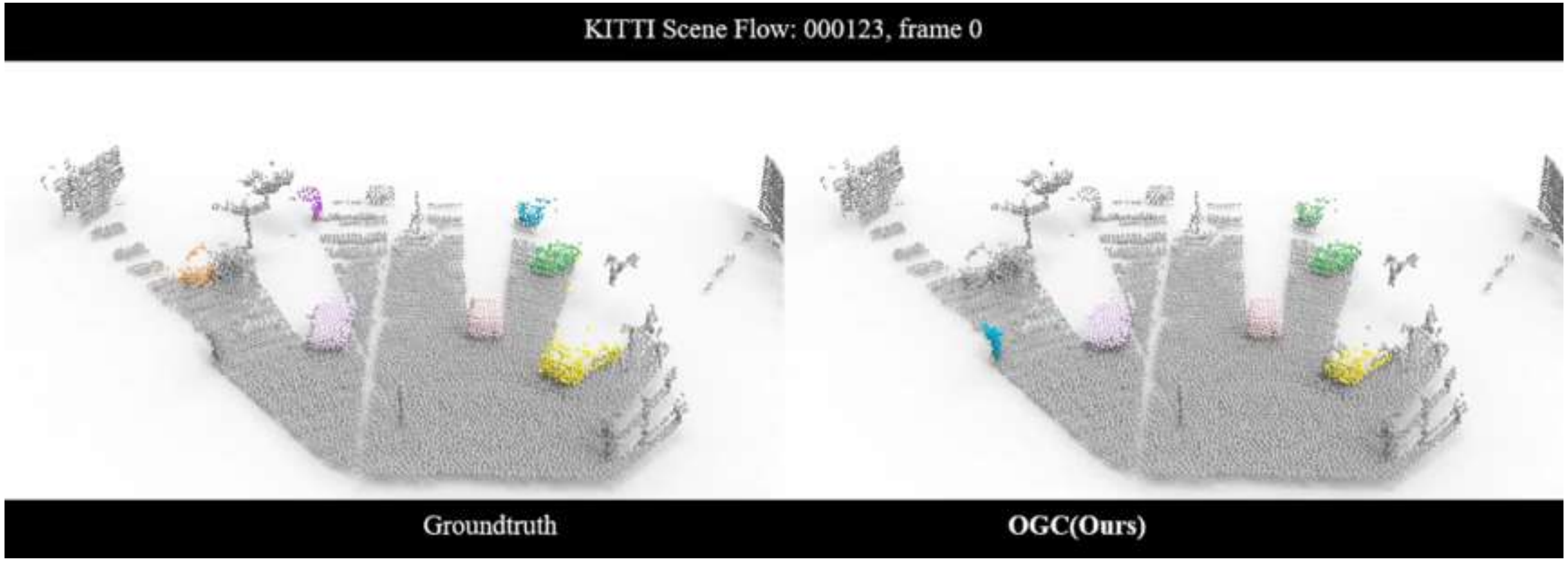
Dynamic Room (OGC-DR)
Complete point clouds

Single-View Dynamic Room (OGC-DRSV)
Incomplete point clouds from
single-view depth scans

Experiments

Object segmentation in outdoor scenes: KITTI-SF dataset

		AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
Supervised Method	OGC_{sup}	62.4	52.7	65.1	63.4	67.0	67.3	95.0
Unsupervised Motion Segmentation	TrajAffn [52]	24.0	30.2	43.2	37.6	50.8	48.1	58.5
	SSC [51]	12.5	20.4	28.4	22.8	37.6	41.5	48.9
Unsupervised Methods	WardLinkage [30]	25.0	16.3	22.9	13.7	**69.8**	60.5	44.9
	DBSCAN [17]	13.4	22.8	32.6	26.7	42.0	42.6	55.3
	OGC(Ours)	**54.4**	**42.4**	**52.4**	**47.3**	58.8	**63.7**	**93.6**

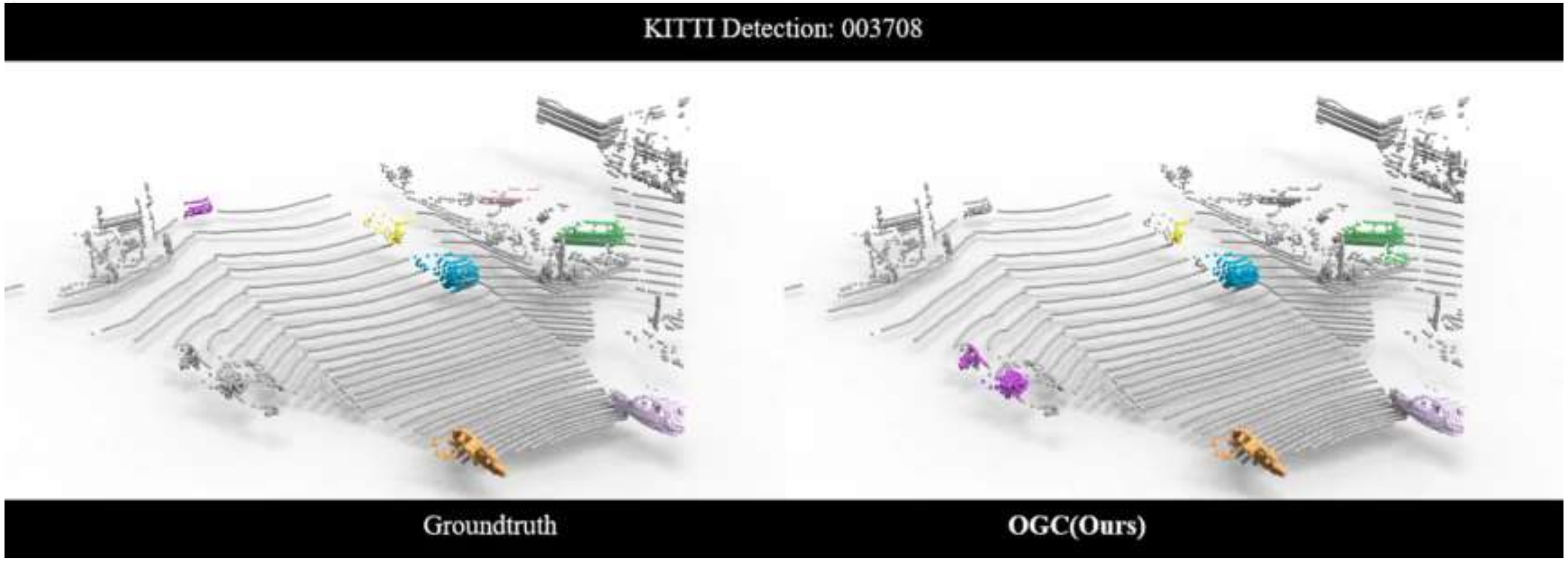


Experiments

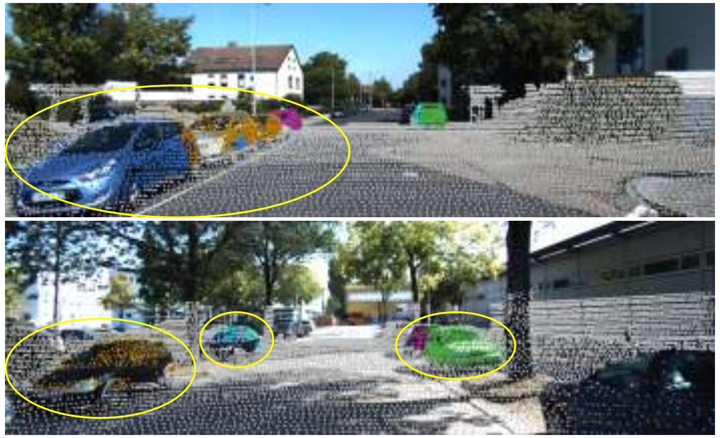
Generalization to single-frame LiDAR data: KITTI-Det dataset

		AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
Supervised Methods	PointRCNN [59]	95.7	80.1	88.9	81.3	98.0	91.4	97.2
	PV-RCNN [58]	95.4	77.3	84.4	73.7	98.8	92.7	97.1
	Voxel-RCNN [14]	95.8	79.6	87.3	78.1	98.9	92.6	97.3
	OGC_{sup}	80.0	68.5	78.3	72.7	84.8	84.0	96.9
	$OGC*_{sup}$	51.4	41.0	49.1	43.7	56.0	66.2	91.0
Unsupervised Method	**OGC*(Ours)**	40.5	30.9	37.0	30.8	46.5	60.6	86.4

(* trained on KITTI-SF)



OGC can segment static cars:

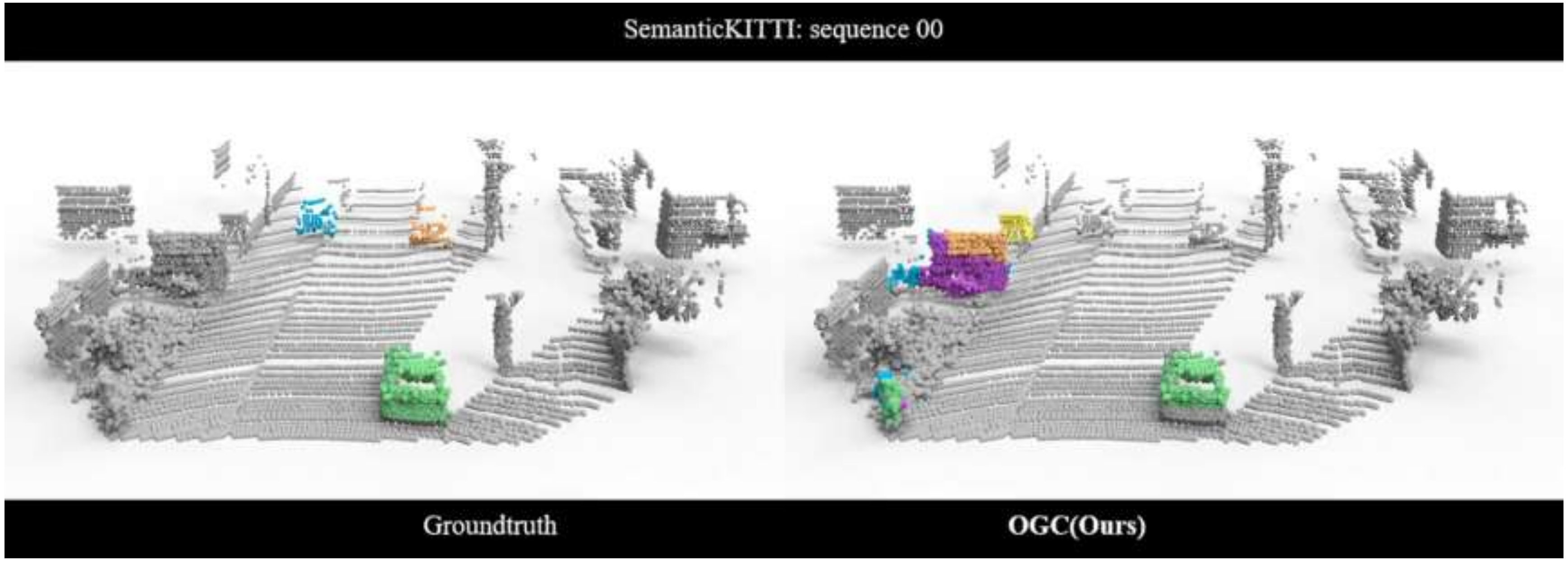


Experiments

Generalization to large-scale LiDAR data: SemanticKITTI (23201 frames)

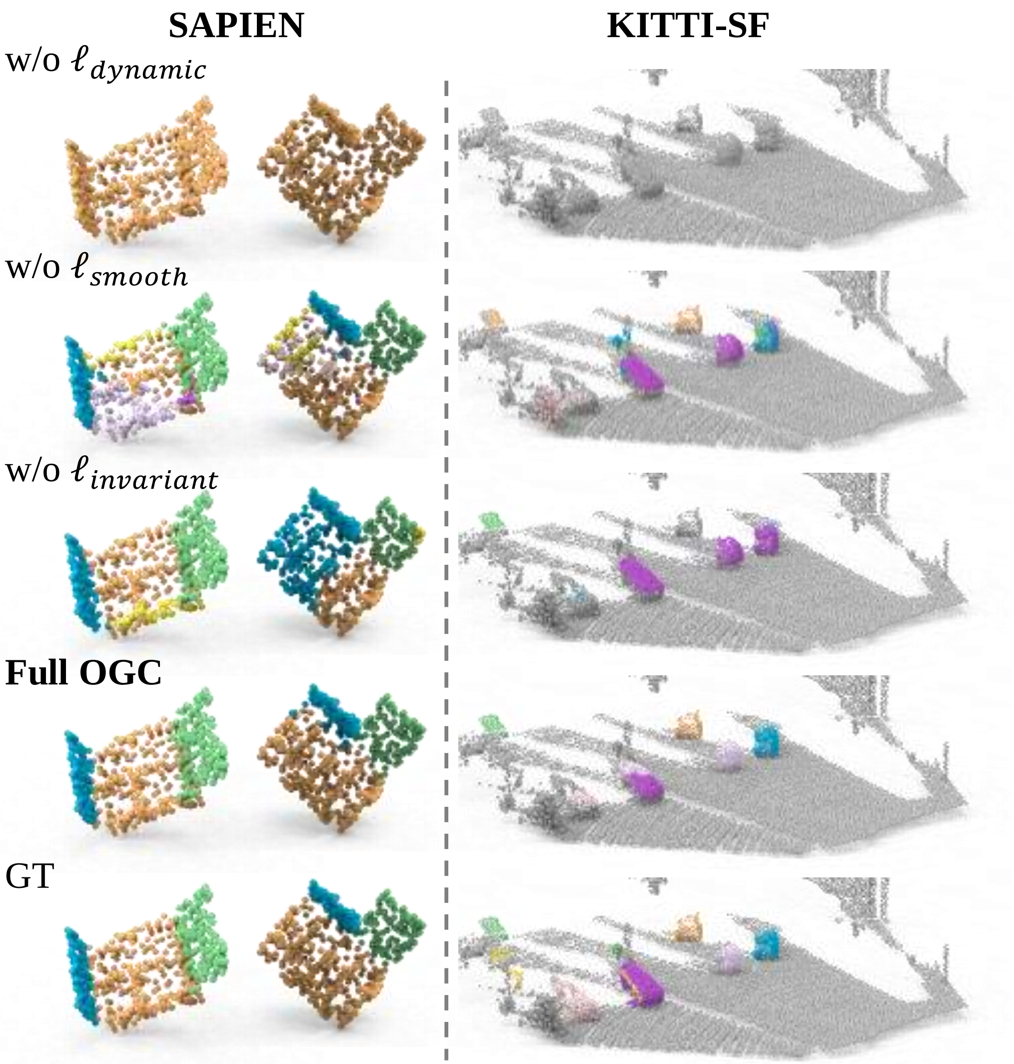
Sequences	Methods	AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
00∼10	OGC*$_{sup}$	53.8	41.3	48.1	40.1	60.0	68.3	90.0
	OGC*(Ours)	42.6	30.2	35.3	28.2	47.3	60.3	86.0
00∼07 & 09∼10	OGC*$_{sup}$	55.3	41.8	48.4	40.1	61.1	69.9	90.3
	OGC*(Ours)	43.6	30.5	35.5	28.1	48.2	62.1	86.3
08	OGC*$_{sup}$	49.4	39.2	46.6	40.0	55.8	60.3	88.3
	OGC*(Ours)	38.6	29.1	34.7	28.6	44.0	51.8	84.3

(* trained on KITTI-SF)



Experiments

Ablation studies



Our **loss design** and **iterative optimization strategy** are validated

	AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
w/o $\ell_{dynamic}$	35.4	35.3	54.1	**91.1**	38.5	28.6	52.7
w/o ℓ_{smooth}	21.8	18.5	26.9	19.1	45.4	52.4	63.7
w/o $\ell_{invariant}$	48.9	46.1	61.3	61.9	60.7	57.9	70.3
Full OGC	**55.6**	**50.6**	**65.1**	65.0	**65.2**	**60.9**	**73.4**

	Object Segmentation						
$\#R$	AP↑	PQ↑	F1↑	Pre↑	Rec↑	mIoU↑	RI↑
1	45.9	47.7	62.3	60.2	64.5	60.2	72.3
2	55.6	50.6	65.1	65.0	65.2	60.9	73.4
3	56.3	50.7	65.4	65.1	65.8	61.1	73.7

Conclusion & Future Directions

Our contributions:

- First unsupervised multi-object segmentation
- OGC losses to supervise with motions
- Promising results

Future directions:

- Combination with supervision
- Leveraging multi-frame inputs (if available)

Thanks

paper: https://arxiv.org/abs/2210.04458
code: https://github.com/vLAR-group/OGC
demo: https://www.youtube.com/watch?v=dZBjvKWJ4K0