Lidar Panoptic Segmentation and Tracking without Bells and Whistles

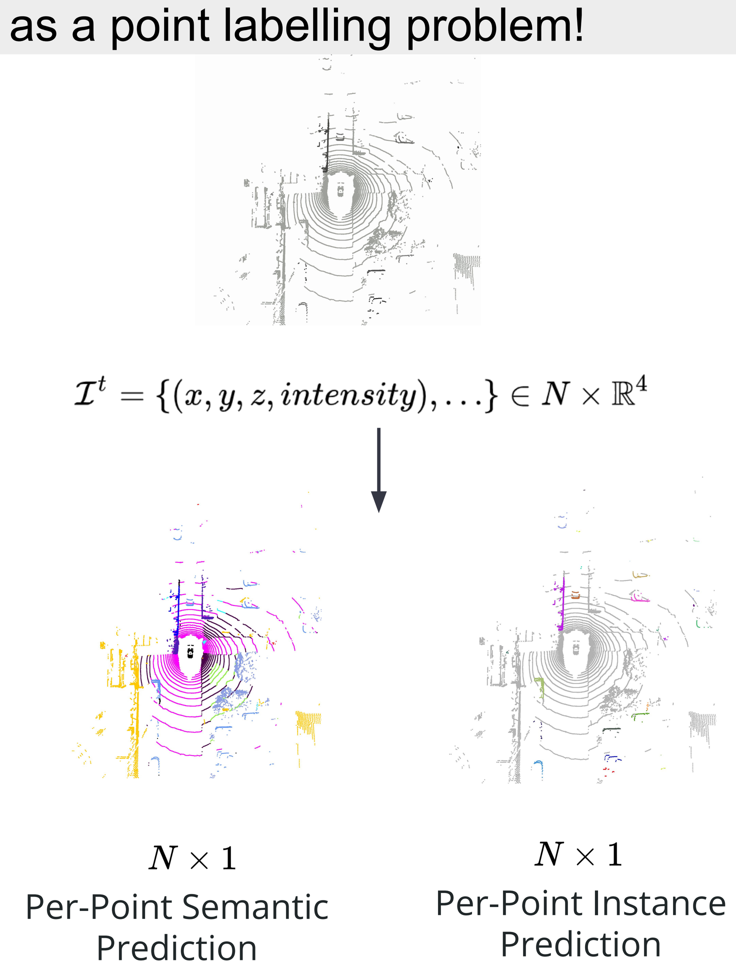
Abhinav Agarwalla[1], Xuhua Huang[1,2], Jason Ziglar[3], Francesco Ferroni[3], Laura Leal-Taixé[4], James Hays[5], Aljoša Ošep[1] and Deva Ramanan[1]

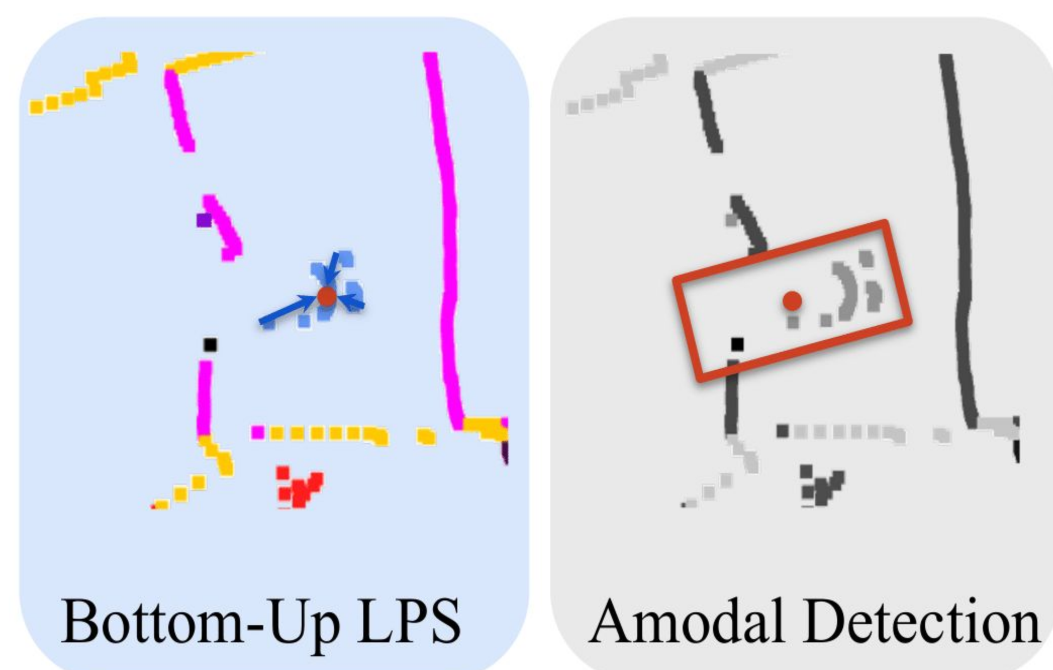
[1]Carnegie Mellon University, [2]Meta AI, [3]Argo AI, [4]TU Munich, [5]Georgia Institute of Technology

TL;DR: In the absence of amodal (cuboid) annotations, we regress modal centroids and object extent using trajectory-level and point-level supervision, which cannot be inferred from single scan due to occlusions and the sparse nature of the lidar data. The resulting model works really well on 3D/4D panoptic segmentation tasks.

Motivation and Introduction

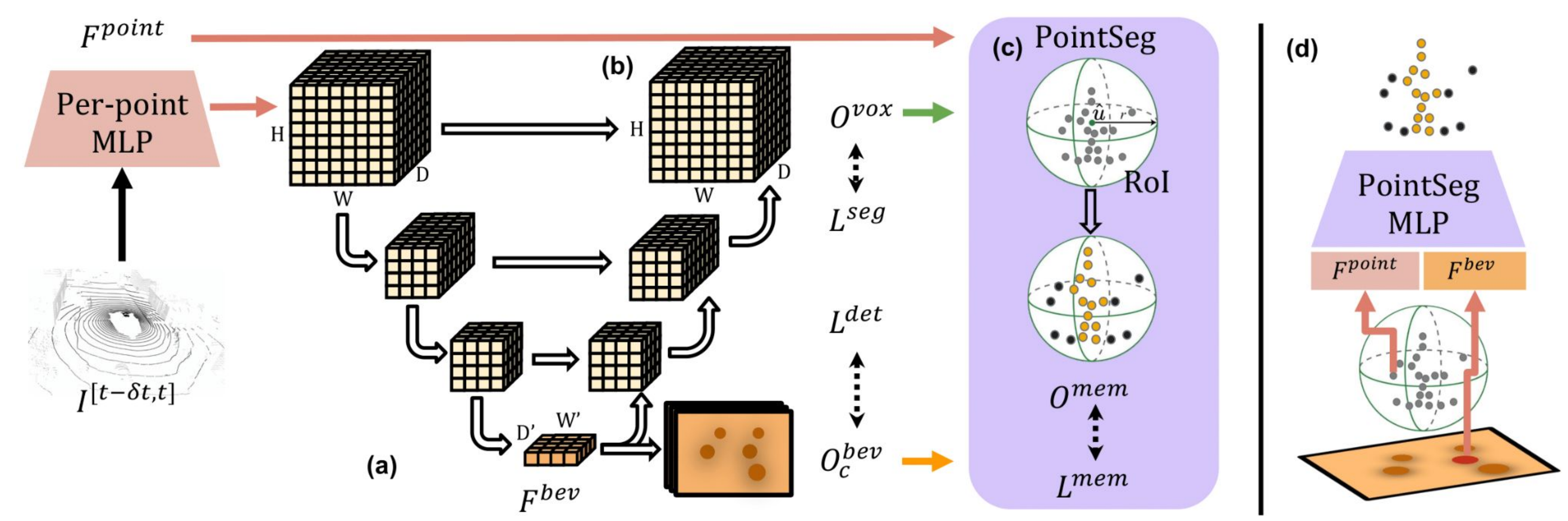
Lidar Panoptic Segmentation (LPS) as a point labelling problem!





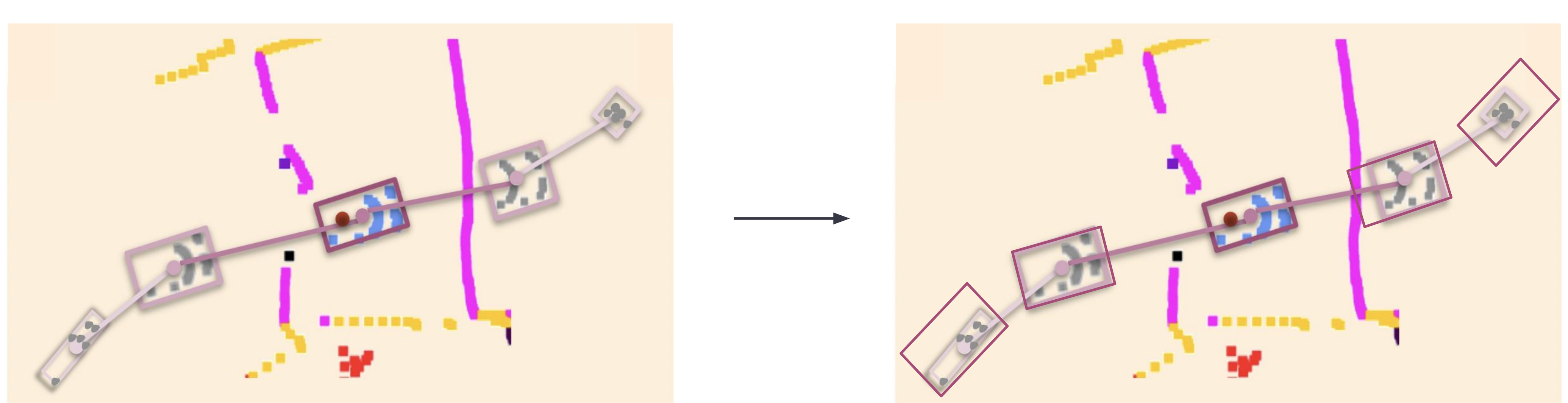
- → Bottom-Up LPS: employs clustering which can lead to over and under-segmentation
- → Amodal Detection: very good object detector but lacks point precise object boundaries

Model Architecture and Training Methodology



- → Modal Detection: Detect modal centers and extents in a 2D BEV space to construct ROIs
- → Semantic Segmentation: get semantic class predictions at the voxel levels
- → PointSegMLP: Binary membership/mask prediction to obtained fine-grained boundaries

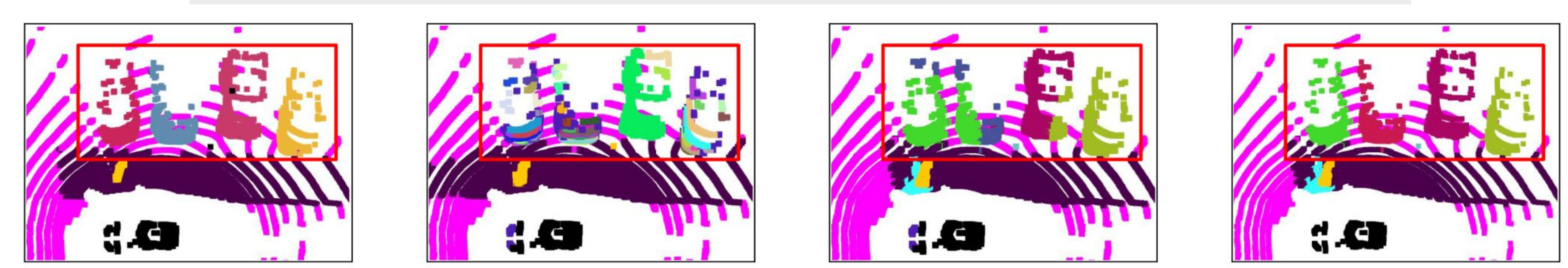
Obtaining Modal Boxes



- → A simple way to obtain modal boxes is to put a tight fitting box around the visible points
- → but, this leads to **large variations** and **small object extents** in object sizes across frames
- → Aggregation by taking maximum observed extents across frames is the key!

Results

Qualitative comparison of MOST with other approaches with the same architecture



(a) GT (b) DS-Net (c) NN-Baseline (d) *MOST*

		Method	PQ	PQ†	RQ	SQ	PQTh	RQTh	SQTh	PQSt	RQSt	SQSt	mIoU
Nuscenes	Test	Panoptic-PHNet [9]	**80.1**	**82.8**	**87.6**	**91.1**	**82.1**	**88.1**	**93.0**	**76.6**	**86.6**	**87.9**	80.2
		Efficient-LidarPanopticSegmentation [45]	62.4	66.0	74.1	83.7	57.2	68.2	83.6	71.1	84.0	83.8	66.7
		PolarSeg-Panoptic [32]	63.6	67.1	75.1	84.3	59.0	69.8	84.3	71.3	83.9	84.2	67.0
		Ours	76.1	79.5	85.1	88.9	77.4	85.5	90.3	73.9	84.5	86.7	**80.4**

		Method	PQ	PQ†	RQ	SQ	PQTh	RQTh	SQTh	PQSt	RQSt	SQSt	mIoU
SemanticKITTI	Test	Panoptic-PHNet [9]	**61.5**	67.9	**72.1**	**84.8**	63.8	70.4	**90.7**	59.9	73.3	80.5	66.0
		SCAN [46]	**61.5**	67.5	**72.1**	**84.5**	61.4	69.3	**88.1**	61.5	74.1	81.8	67.7
		PolarSeg-Panoptic [32]	54.1	60.7	65.0	81.4	53.3	60.6	87.2	54.8	68.1	77.2	59.5
		DS-Net [7]	55.9	62.5	66.7	82.3	55.1	62.8	87.2	56.5	69.5	78.7	61.6
		Efficient-LidarPanopticSegmentation [45]	57.4	63.2	68.7	83.0	53.1	60.5	87.8	60.5	74.6	79.5	61.4
		MaskPLS [33]	58.2	63.3	68.6	83.9	55.7	61.7	89.2	60.0	73.7	80.0	62.5
		GP-S3Net [8]	60.0	**69.0**	**72.1**	82.0	**65.0**	**74.5**	86.6	56.4	70.4	78.7	**70.8**
		Ours	61.0	66.8	72.0	84.4	58.1	66.0	88.1	**63.2**	**76.3**	**81.7**	66.1

Extension to 4D Panoptic Segmentation / Tracking

- → tracking-by-detection paradigm
- → associate boxes greedily via back projection velocity estimates
- → assign a unique id to a tracklet

	Method	$LSTQ$	PAT	S_{assoc}	S_{cls}	PTQ	PQ
SemanticKITTI	RangeNet++ [21] + PP + MOT	35.5	-	24.1	52.4	-	-
	KPConv [17] + PP + SFP	38.5	-	26.6	55.9	-	-
	4D-PLS [5]	56.9	-	56.4	57.4	-	-
	Contrastive Association [35]	63.1	-	65.7	60.6	-	-
	4D-StOP [50]	**63.9**	-	**69.5**	58.8	-	-
	Ours	60.3	-	57.8	**62.8**	-	-
nuScenes	PanopticTrackNet [34]	44.8	45.7	36.7	58.9	51.6	51.7
	4D-PLS [5]	57.8	60.5	53.6	62.3	55.6	56.6
	E-LPS [45] + Kalman	63.7	67.1	60.2	67.4	62.3	63.6
	E-LPT [45]	66.4	70.4	-	69.5	67.5	67.9
	Ours	**73.2**	**74.9**	**66.6**	**80.4**	**72.0**	**76.0**

abhinavagarwalla/most-lps

https://mostlps.github.io/

Acknowledgements

This project was funded by, in parts, by Sofja Kovalevskaja Award of the Humboldt Foundation.