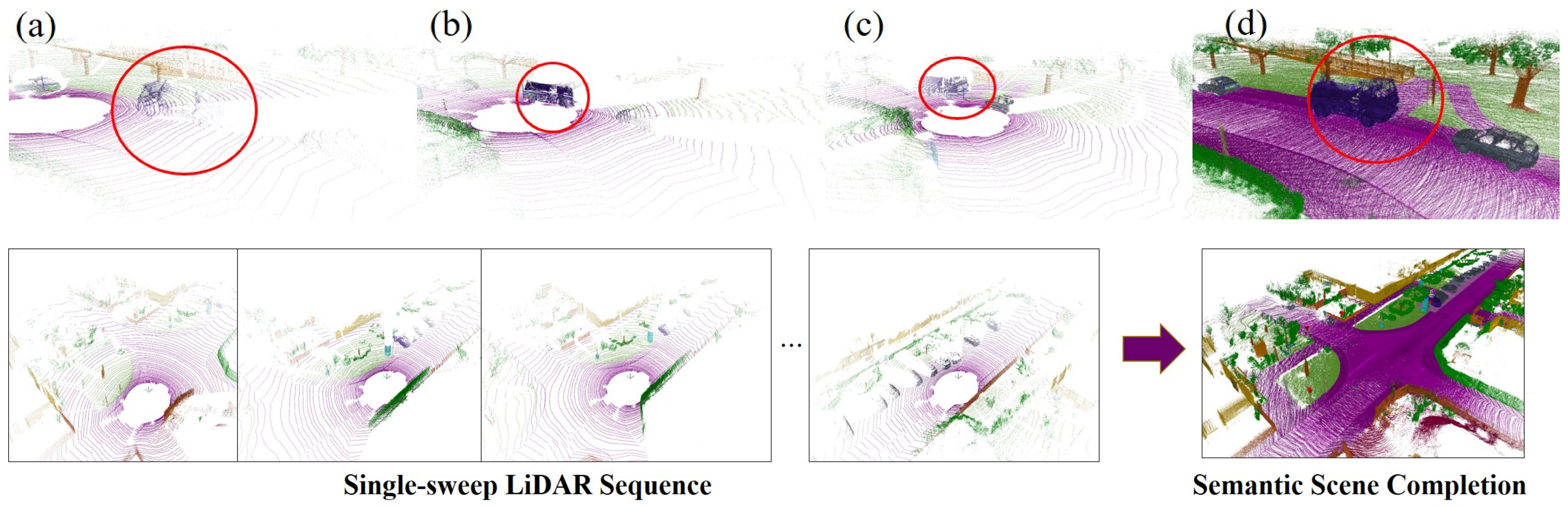
Sparse Single Sweep LiDAR Point Cloud Segmentation via Learning Contextual Shape Priors from Scene Completion

Xu Yan[1,2,†], Jiantao Gao[2,4,†], Jie Li[1,3], Ruimao Zhang[1,2], Zhen Li[1,2,⋆], Rui Huang[1,3], Shuguang Cui[1,2]

[1] The Chinese University of Hong Kong, Shenzhen , [2] Shenzhen Research Institute of Big Data
[3] Shenzhen Institute of Artificial Intelligence and Robotics for Society, [4] Shanghai University

Problem and Contributions

Motivation: For a single frame point cloud (a) with extremely sparse points, it seems impossible for previous methods to conduct accurate segmentation. Nevertheless, such segmentation would be possible, if we introduce the richer shape information from the other two frames, (b) and (c), to reconstruct a shape-complete object as shown in (d). Inspired by this, we present a novel enhanced sparse LiDAR point clouds semantic segmentation model assisted by learned contextual shape priors from Scene Completion.



Key Contributions:

- The proposed **JS3C-Net** is the first to achieve the enhanced sparse single sweep LiDAR semantic segmentation via auxiliary scene completion.
- For better trade-off between performance and effectiveness, our auxiliary components are designed in cascaded and disposable manners, and a novel point-voxel interaction (PVI) module is proposed for better feature interaction and fusion between the two tasks.
- Our method shows superior results in both Semantic Segmentation (SS) and Semantic Scene Completion (SSC) on two benchmarks, *i.e.,* SemanticKITTI and SemanticPOSS.

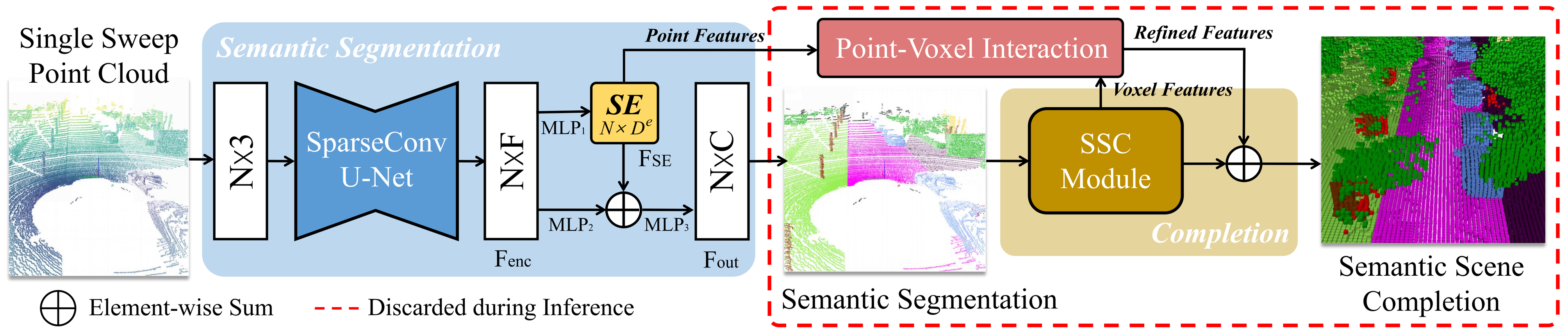
Method

Proposed Framework:

We propose an enhanced **J**oint single sweep LiDAR point cloud **S**emantic **S**egmentation by exploiting learned shape prior form **S**cene **C**ompletion **Net**work, *i.e.,* **JS3C-Net**.

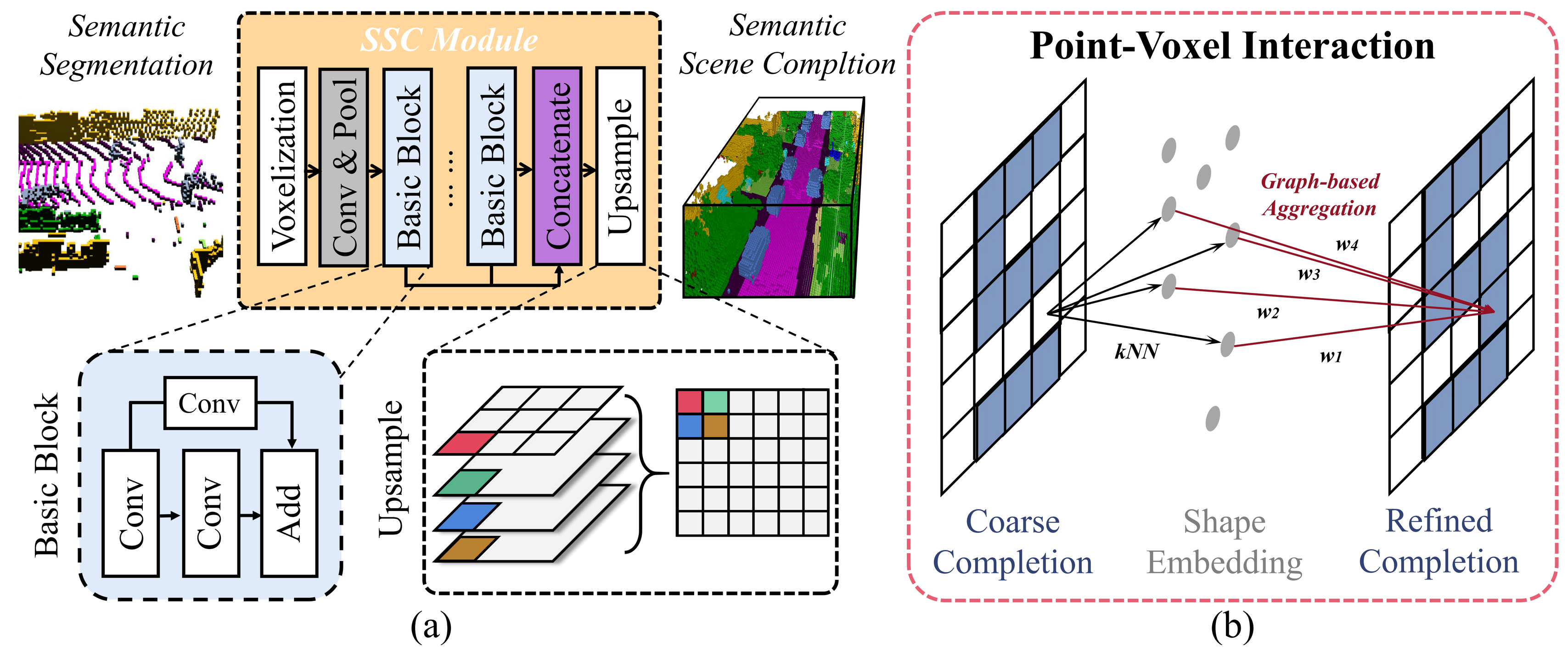
- We firstly use the general appealing point cloud segmentation network to obtain initial point semantic segmentation.
- The SSC module takes results of segmentation network as input and generates the completed voxel of the whole scene with dense convolution neural network.
- The Point-Voxel Interaction (PVI) module is proposed to conduct shape-aware knowledge transfer.

Note that SSC module and PVI module can be ***discarded*** during inference to prevent introducing computing burden for segmentation.

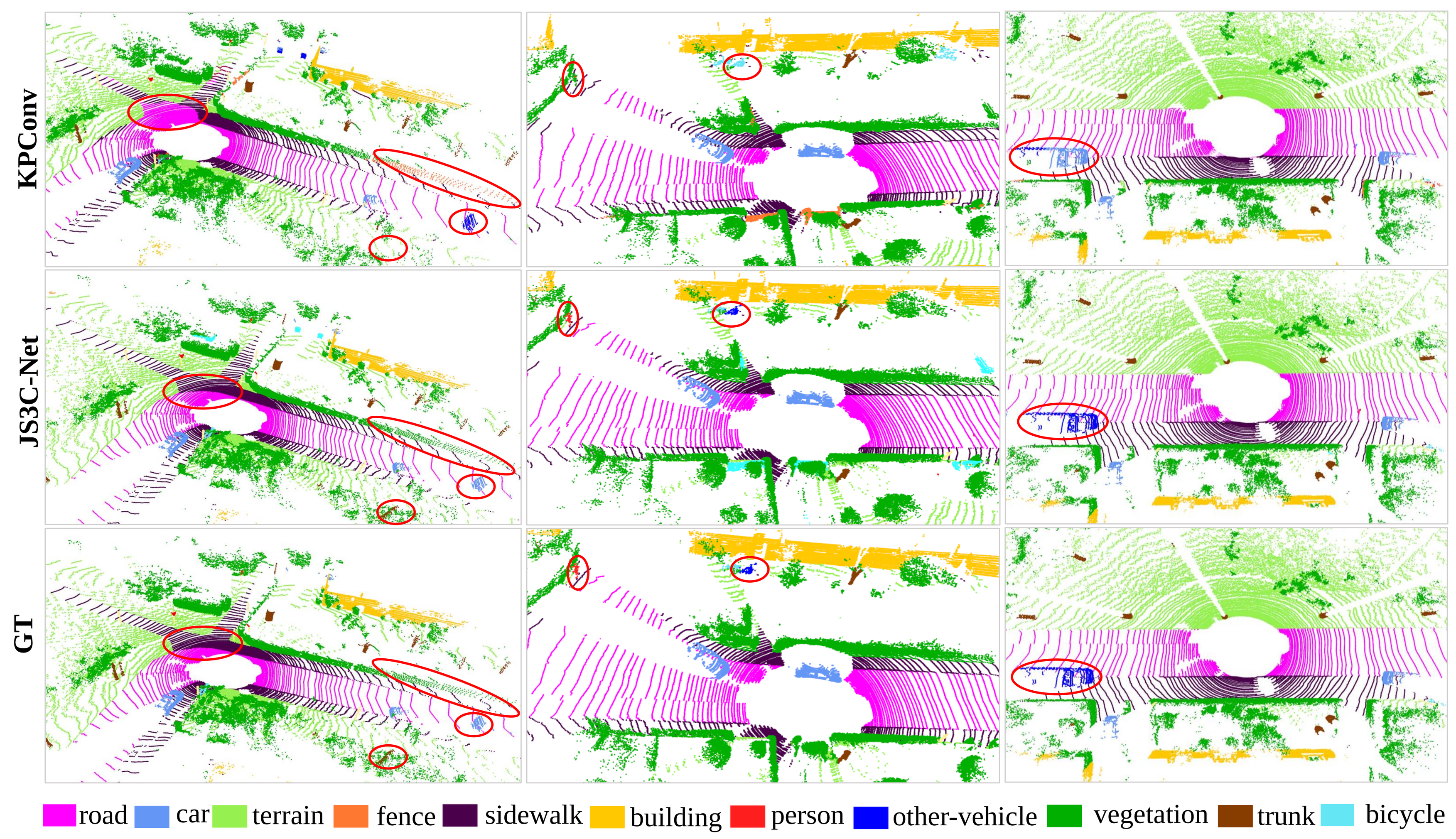


Implementation & Results

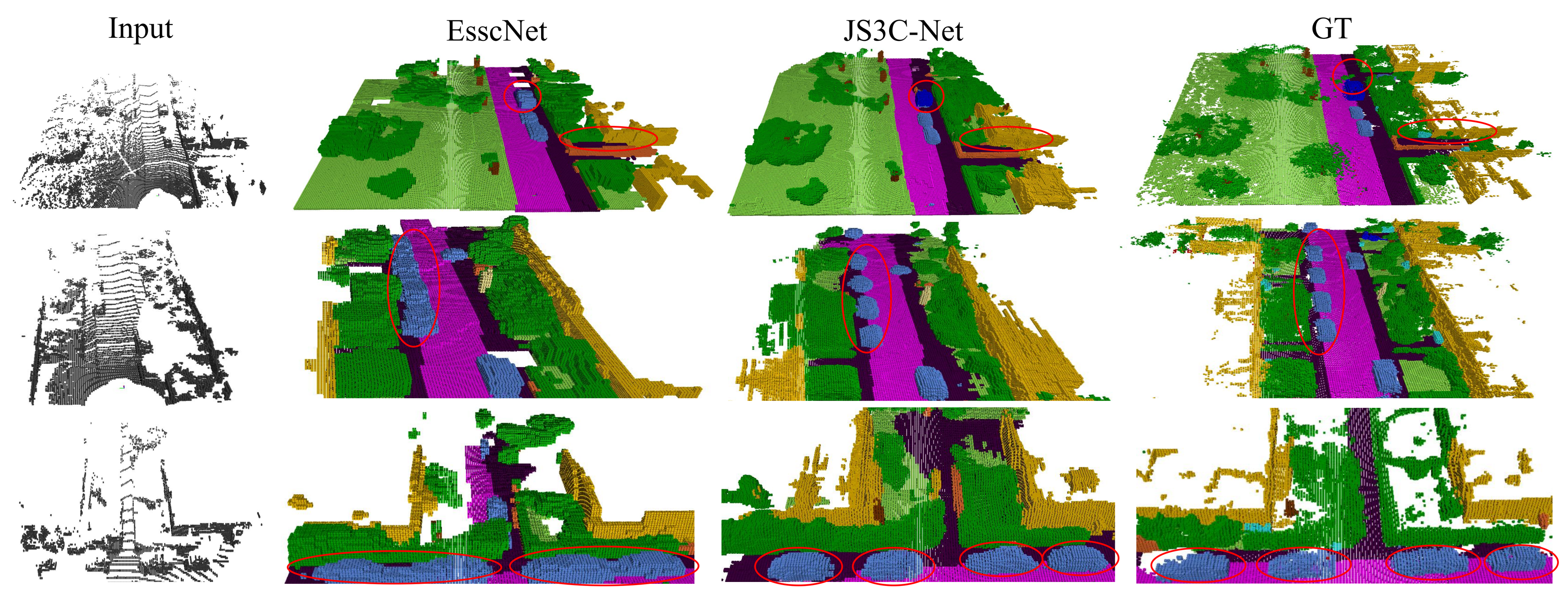
Inner structure of Semantic Scene Completion (SSC) module and Point-Voxel Interaction (PVI) module.



Qualitative results of the Semantic Segmentation (SS) task on *SemanticKITTI* dataset. Red circles show that our method performs better in many details than recent state-of-the-art.



Qualitative results of the Semantic Scene Completion (SSC) task on *SemanticKITTI* dataset. Red circles show that our method performs better in many details than recent state-of-the-art.



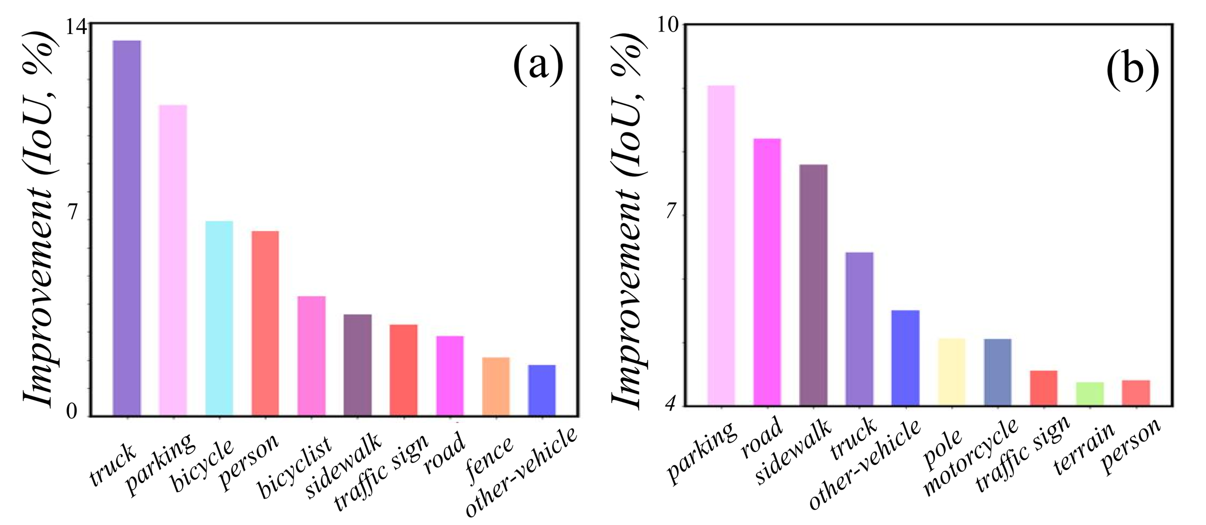
Semantic Segmentation results on the *SemanticKITTI* benchmark. The upper, medium and bot tom parts of the table contain projection-based, point-based and voxel-based methods, respectively.

Method	Selected 3 classes *Truck*	*Bicycle*	*Person*	all classes *mIoU*
SqueezeSegV2	13.4	18.5	20.1	39.7
DarkNet53Seg	25.5	24.5	36.2	49.9
RangeNet53++	25.7	25.7	38.3	52.2
3D-MiniNet	28.5	42.3	47.8	55.8
SqueezeSegV3	29.6	38.7	45.6	55.9
PointNet++	0.9	1.9	0.9	20.1
TangentConv	15.2	2.7	23.0	40.9
PointASNL	39.0	0.0	34.2	46.8
RandLA-Net	43.9	29.8	48.4	55.9
KPConv	33.4	30.2	61.5	58.8
PolarNet	22.9	40.3	43.2	54.3
SparseConv	43.5	51.0	60.4	61.8
JS3C-Net (Ours)	**54.3**	**59.3**	**69.5**	**66.0**

Semantic Scene Completion results on the *SemanticKITTI* benchmark. Only the recent published approaches are compared.

Method	precision	recall	IoU	mIoU
SSCNet	31.7	83.4	29.8	9.5
TS3D	31.6	84.2	29.8	9.5
TS3D[2]	25.9	**88.3**	25.0	10.2
EsscNet	62.6	55.6	41.8	17.5
TS3D[3]	**80.5**	57.7	50.6	17.7
JS3C-Net (Ours)	71.5	73.5	**56.6**	**23.8**

Top-10 mIoU gains between JS3C-Net and split-trained single task (SS or SSC) on the *SemanticKITTI* dataset.



Scan the Qr-Code:
Github & Preprint & Contact Us