Joint Semantic-Instance Segmentation of 3D Point Clouds: Instance Separation and Semantic Fusion

Min Zhong, Gang Zeng
Key Laboratory on Machine Perception, Peking University

Introduction:

- **3D Semantic and Instance Segmentation** Input: 3D point clouds. Output: instance labels and semantic labels for each point.
- **Joint 3D Semantic-Instance Segmentation**. points with different semantic labels must belong to different instances. points within the same instance must share the same semantics.

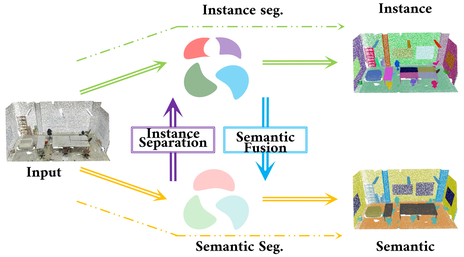


Fig. 1: Joint two tasks with two interaction modules.

The overall framework:

- **Two task branches**: In Fig. 2. A point cloud encoder (e.g. PointNet) used to extract feature F. Then one branch for semantic seg. (orange) and the other for instance seg. (green).
- **Two modules set up cooperation.** Instance Separation Module(purple) estimates object centroid O, which concat with S to create semantic-aware instance feature. The Semantic Fusion Module(blue) uses instance embedding to produce attention map and get more instance consistent semantic feature.

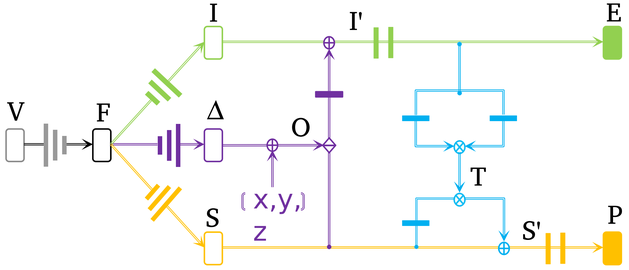


Fig. 2: The overall network.

Two interaction modules

- **Instance Separation.** Semantic feature can help separate different semantic instances. However, the same semantics may contain different instances. So we futher supplement the semantic feature with instance-specific position information.
- **Semantic Fusion.** Points within the same instance must have the same semantics. We produce an attention map between each pair of points in the instance embedding space, and then used to aggregate the semantic information in the semantic feature space.

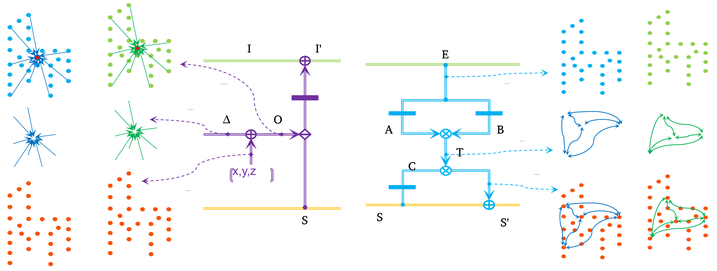


Fig. 3: Two interaction modules: instance separation and semantic fustion

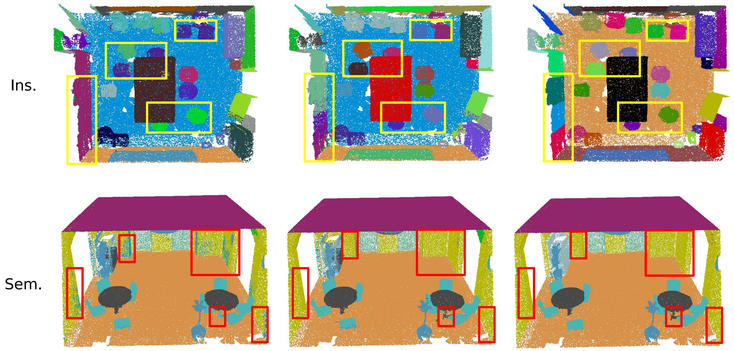
Experiments:

- **Compare with other methods**.

Backbone	Method	mCov	mWCov	mPrec	mRec
Test on Area 5					
PN	SGPN [28]	32.7	35.5	36.0	28.7
	ASIS [29]	40.4	43.3	44.5	37.4
PN++	ASIS [29]	44.6	47.8	55.3	42.4
	Ours	**48.7**	**51.8**	**58.2**	**46.6**
Test on 6-fold CV					
PN	SGPN [28]	37.9	40.8	38.2	31.2
	ASIS [29]	44.7	48.2	53.2	40.7
PN++	ASIS [29]	51.2	55.1	63.6	47.5
	Ours	**54.2**	**58.1**	**65.3**	**50.8**

Backbone	Method	mAcc	mIoU	oAcc
Test on Area 5				
PN	PN (*RePr*)	52.1	43.4	83.5
	ASIS [29]	55.7	46.4	84.5
PN++	ASIS [29]	60.9	53.4	86.9
	Ours	**62.7**	**55.3**	**87.7**
Test on 6-fold CV				
PN	PN (*RePr*)	60.3	48.9	80.3
	ASIS [29]	62.3	51.1	81.7
PN++	ASIS [29]	70.1	59.3	86.2
	Ours	**71.6**	**60.9**	**86.7**

Tab. 1: Compare with other methods.



(a) ASIS (b) Ours (c) GT

Fig. 4: The semantic and instance seg. results. compare with the ASIS method.

- **Ablation studies on two modules.**

Method	+FC	+FS	+FF	mCov	mWCov	mPrec	mRec	mAcc	mIoU	oAcc
Test on Area 5										
Baseline				42.6	45.7	53.4	40.6	58.3	50.8	86.7
FC-Layer	✓			45.0	48.0	54.9	42.3	61.0	53.1	87.2
FS-Module		✓		47.3	50.2	56.8	45.0	61.3	53.2	86.9
FF-Module			✓	45.5	48.5	54.8	42.2	61.7	54.5	87.7
Ours-Full		✓	✓	**48.7**	**51.8**	**58.2**	**46.6**	**62.7**	**55.3**	**87.7**

Tab. 2: Ablation studies. FC is FS without position information; FS is Feature separation modulse; FF is Feature Fusion module.

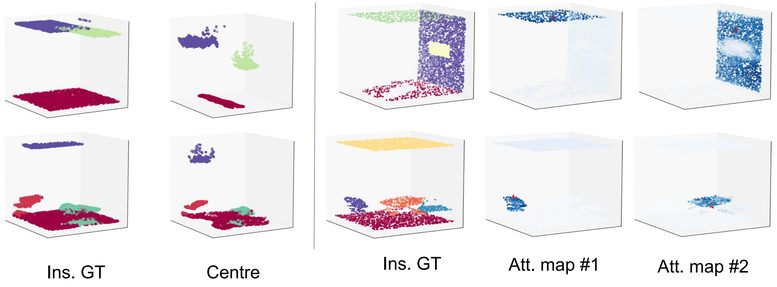


Fig. 5: Visulization of the estimates object centroid O (left) and the attention maps (left)