

# Bi-Directional Attention for Joint Instance and Semantic Segmentation in Point Clouds

## Supplementary Material

Guangnan Wu<sup>1</sup>[0000-0002-0841-5791], Zhiyi Pan<sup>1</sup>[0000-0002-0635-0349],  
Peng Jiang<sup>1\*</sup>[0000-0002-7342-7940], and Changhe Tu<sup>1\*</sup>[0000-0002-1231-3392]

<sup>1</sup>Shandong University, China.  
{wuguangnan1006, panzhiyi1996, sdujump, changhe.tu}@gmail.com

In this supplementary, we will firstly give the detailed explanation of our loss function in Sec. 1. Then, the train and test details of three datasets are in Sec. 2. And we also provide code tutorial in Sec. 3. More visual results of our method on the S3DIS, PartNet and ScanNetV2 datasets are shown in Sec. 4. Finally, more instance and semantic similarity matrices are visualized in Sec. 5.

### 1 Loss Function

As mentioned in the main manuscript, our loss function  $\mathcal{L}$  is defined as:

$$\mathcal{L} = \mathcal{L}_{sem} + \mathcal{L}_{ins}, \quad (1)$$

where  $\mathcal{L}_{sem}$  is classical cross entropy loss for semantic segmentation and  $\mathcal{L}_{ins}$  is instance segmentation loss.  $\mathcal{L}_{ins}$  is defined as:

$$\mathcal{L}_{ins} = \alpha \cdot \mathcal{L}_{var} + \beta \cdot \mathcal{L}_{dist} + \gamma \cdot \mathcal{L}_{reg}, \quad (2)$$

where  $\mathcal{L}_{var}$  aims to pull the points belonging to the same instance towards the mean embedding i.e. the instance center, while  $\mathcal{L}_{dist}$  aims to push different instance centers far away from each other. And  $\mathcal{L}_{reg}$  is a regularization term to keep the embedding values bounded.  $\alpha$ ,  $\beta$  and  $\gamma$  are the weights of the three terms. Specifically, each term is formulated as:

$$\mathcal{L}_{var} = \frac{1}{C} \sum_{i=1}^C \frac{1}{N_i} \sum_{j=1}^{N_i} [\|\mu_i - x_j\|_1 - \delta_v]_+^2, \quad (3)$$

$$\mathcal{L}_{dist} = \frac{1}{C(C-1)} \sum_{c_A=1}^C \sum_{\substack{c_B=1 \\ c_A \neq c_B}}^C [2\delta_d - \|\mu_{c_A} - \mu_{c_B}\|_1]_+^2, \quad (4)$$

$$\mathcal{L}_{reg} = \frac{1}{C} \sum_{c=1}^C \|\mu_c\|_1, \quad (5)$$

---

\* Corresponding author.

2 G. Wu, Z. Pan, P. Jiang, C. Tu.

045 where  $C$  is the number of ground-truth instances,  $N_c$  is the number of points  
046 in instance  $c$ ,  $\mu_i$  is the mean embedding of instance  $c$ ,  $x_j$  is an embedding of a  
047 point.  $\|\cdot\|_1$  is the  $l_1$  distance.  $\delta_v$  and  $\delta_d$  are margins.  $[x]_+ = \max(0, x)$  means the  
048 hinge. In our experiment, we set  $\alpha = 1$ ,  $\beta = 1$ ,  $\gamma = 0.001$ ,  $\delta_v = 0.5$ ,  $\delta_d = 1.5$ .  
049

## 050 2 Train and Test Details 051

052 We evaluated BAN on the three prevalent datasets. For S3DIS, we carry out  
053 training (100 epochs) and testing with the 6-fold cross-validation and split the  
054 rooms into  $1m \times 1m$  overlapped blocks (each containing 4096 points) on the  
055 ground plane, as used in PointNet. 056

057 While for ScanNetV2, we use the official training and testing set partition.  
058 We split the rooms as S3DIS and train the network for 20 epochs. As for PartNet,  
059 following SGPN [1], we train and test on each object category separately and  
060 report the evaluation results as the mean of metric values over all the objects.  
061 Each object is represented by a point cloud with 4,096 points. 062

## 063 3 Code Tutorial 064

065 We provide source code and use S3DIS dataset as example to prove the trust-  
066 worthy and reproducible of our method. Please check "README.md" file in  
067 "BAN" folder for the details. 068

## 069 4 Visual Results 070

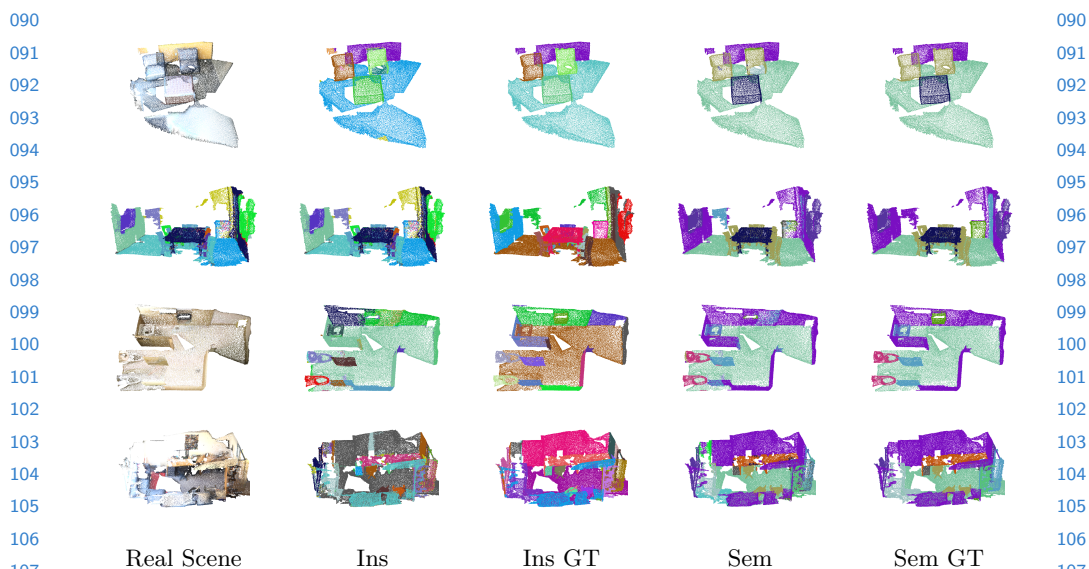
071 Here, more visual results of our method are shown in Fig. 3, Fig. 2 and Fig. 1.  
072

## 073 5 Similarity Matrix Visualization 074

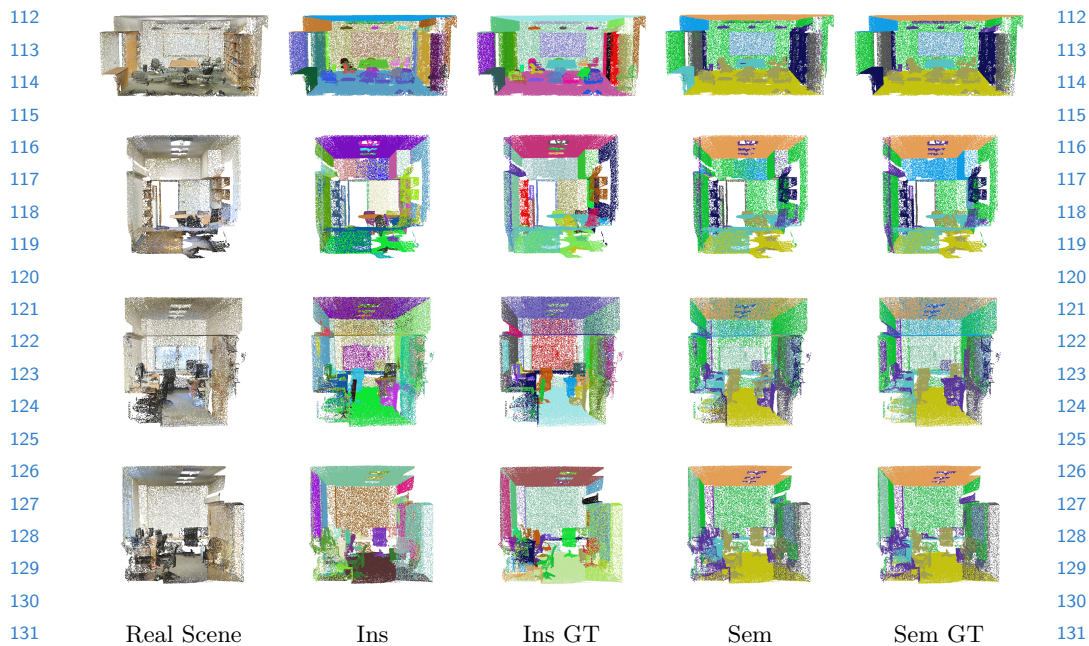
075 Here, more instance and semantic similarity matrices are visualized in Fig. 4.  
076

## 077 References 078

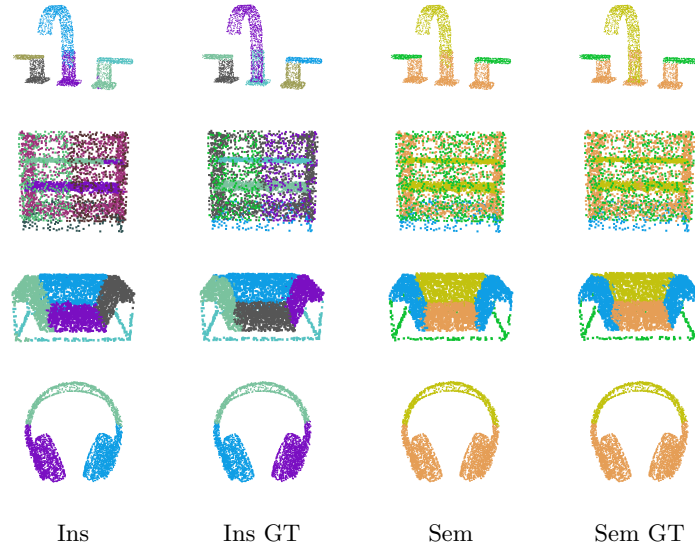
- 079  
080 1. Wang, W., Yu, R., Huang, Q., Neumann, U.: Sgpn: Similarity group proposal  
081 network for 3d point cloud instance segmentation. In: Proceedings of the IEEE  
082 Conference on Computer Vision and Pattern Recognition. (2018) 2569–2578  
083  
084  
085  
086  
087  
088  
089



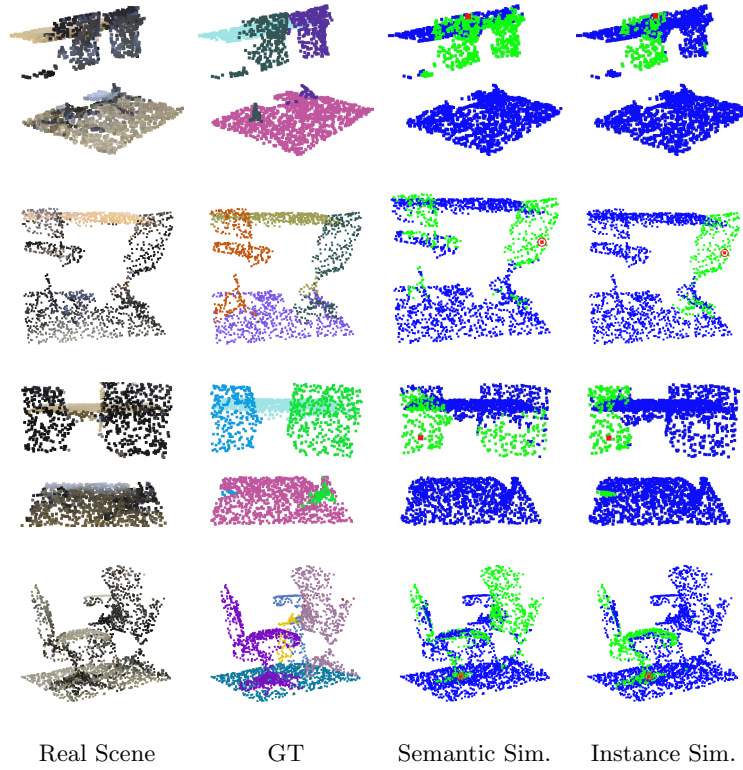
**Fig. 1.** Visual results on the ScanNetV2 dataset.



**Fig. 2.** Visual results on the S3DIS dataset.



**Fig. 3.** Visual results on the PartNet dataset.



**Fig. 4.** Visualization of instance and semantic similarity matrices.