Exploiting Local and Global Structure for Point Cloud Semantic Segmentation with Contextual Point Representations

Xu Wang[1], **Jingming He**[1], **Lin Ma**[2*]
1. Shenzhen University 2. Tencent AI Lab
wangxu@szu.edu.cn; hejingming519@gmail.com; forest.linma@gmail.com

1 Proposed Model

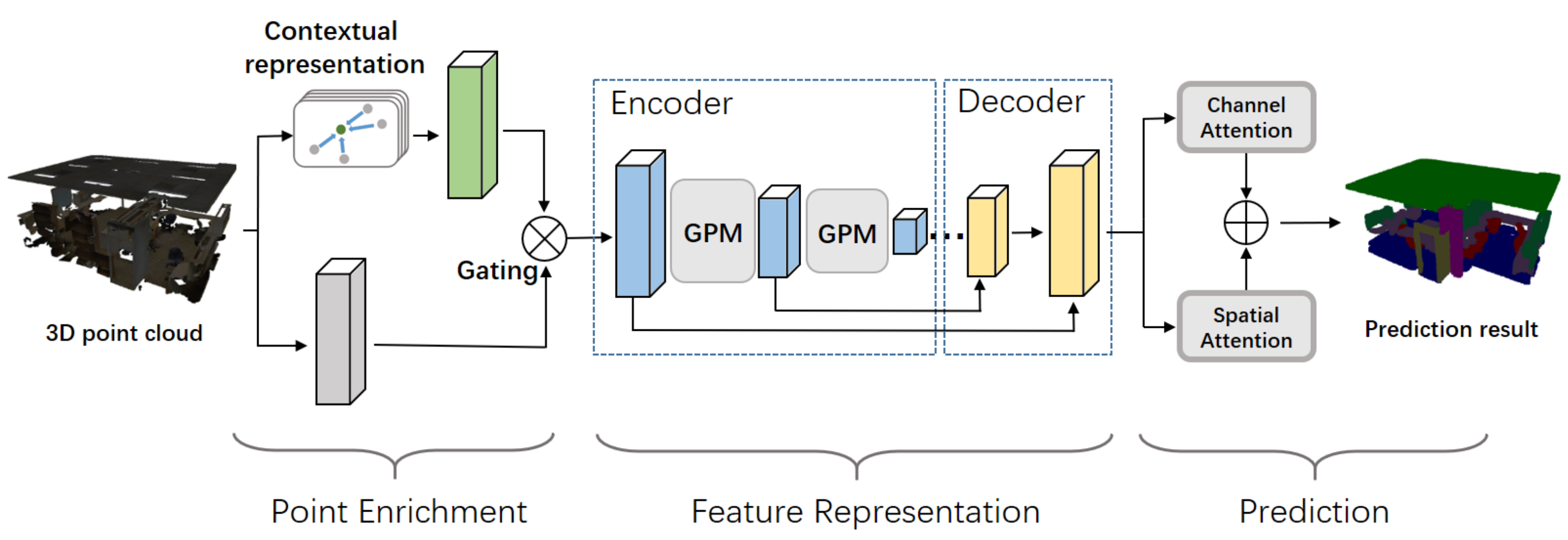

Figure 1: Our proposed model for the point cloud segmentation, consisting of three fully-coupled components. The point enrichment not only considers the point itself but also its contextual points to enrich the corresponding semantic representation. The feature representation relies on conventional encoder-decoder architecture with lateral connections to learn the feature representation for each point. Specifically, the GPM is proposed to dynamically compose and update each point representation via a GAB module. For the prediction, we resort to both channel-wise and spatial-wise attentions to exploit the global structure for the final semantic label prediction of each point.

1.1 Point Enrichment

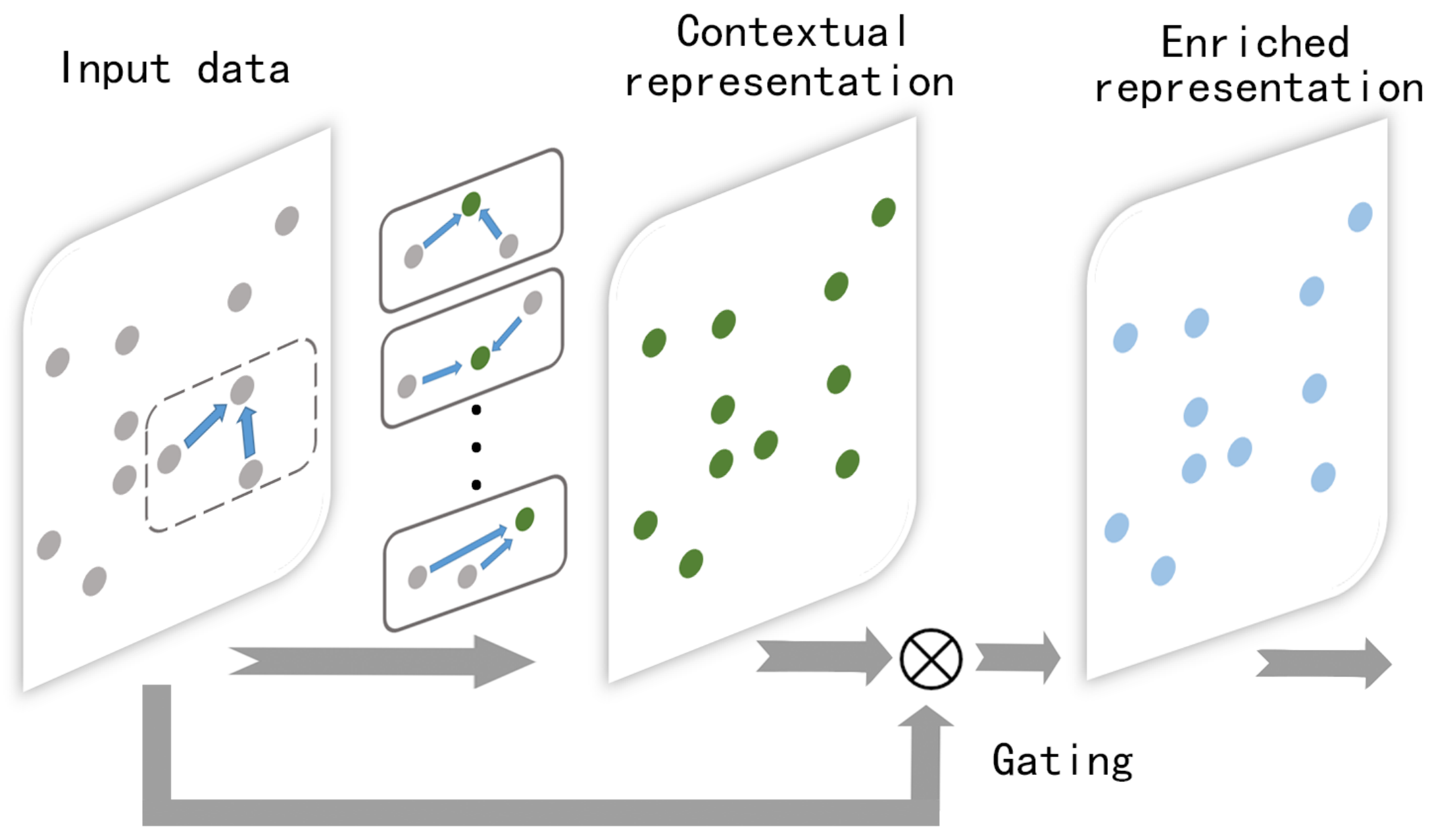

Figure 2: The point enrichment process relies on our proposed gated fusion strategy to enrich the point representation by considering both the neighbouring and contextual points of each point.

To make accurate class prediction for each point within the complicated point cloud structure, we need to not only consider the information of each point itself but also its neighboring or contextual points. Different from the existing approaches, relying on the information of each point itself, such as the geometry, color, etc., we proposed a point enrichment layer to enrich each point representation by taking its neighboring or contextual points into consideration. With the incorporated contextual information, each point is able to sense the complicated point cloud structure information.

Contextual Representation:

$$R_i = \mathop{\Big\|}_{j \in \mathcal{N}_i} P_j \in \mathbb{R}^{C_f k}, \tag{1}$$

where $\mathcal{N}_i$ is the k-nearest neighbor set within the local region centered on point i

Gated Fusion:

$$\begin{aligned} g_i &= \sigma(w_i R_i + b_i), & \hat{P}_i &= g_i \odot \tilde{P}_i, \\ g_i^R &= \sigma(w_i^R \tilde{P}_i + b_i^R), & \hat{R}_i &= g_i^c \odot R_i, \end{aligned} \tag{2}$$

where $w_i, w_i^R \in \mathbb{R}^{C_f k \times C_f}$ and $b_i, b_i^R \in \mathbb{R}^{C_f k}$ are the learnable parameters. σ is the non-linear sigmoid function. $\odot$ is the element-wise multiplication.

1.2 Feature Representation

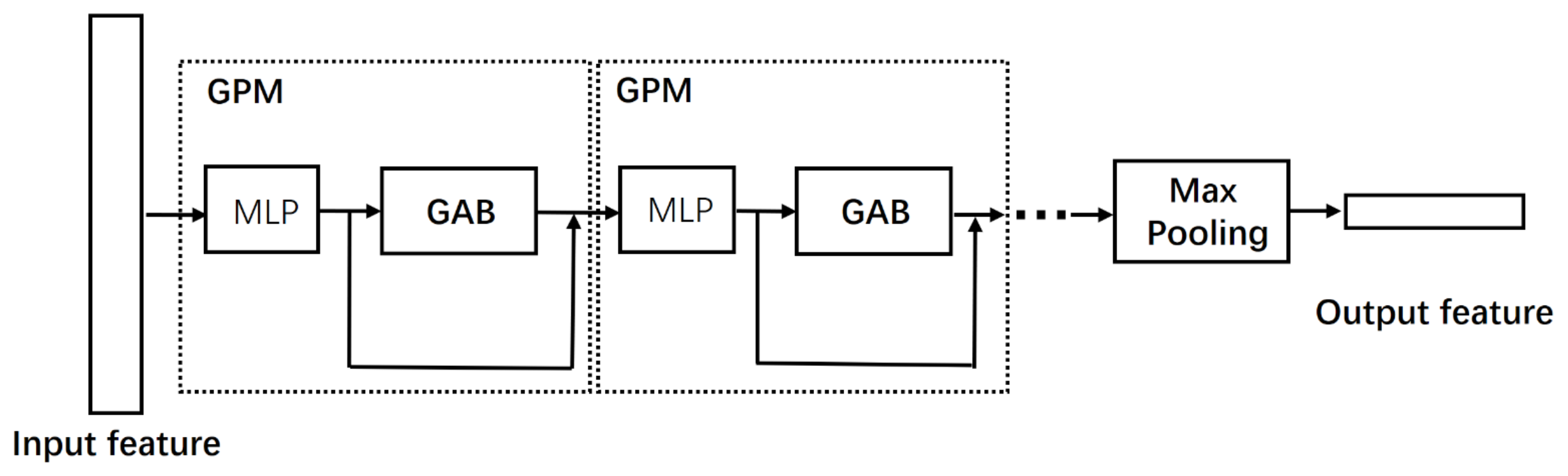

Figure 3: The architecture of our proposed GPM, which stacks MLP and the GAB to exploit the point relationships within the local structure.

With the enriched point representation, we resort to the conventional encoder-decoder architecture with lateral connections to learn the feature representation for each point. To further exploit local structure information of the point cloud, the GMP is employed in the encoder, which relies on the GAB to dynamically compose and update the feature representation of each point within its local regions. The decoder with lateral connections works on the compacted representation obtained from the encoder, to generate the semantic feature representation for each point.

Graph Pointnet Module: GAB defines one fully connected undirected graph to measure the similarities between any two points with such local structure. Given the output feature map $\mathbf{G} \in \mathbb{R}^{C_e \times N_e}$ of the MLP layer in the GPM module, we first linearly project each point to one common space by a learnable shared weight matrix $\mathbf{W}_g \in \mathbb{R}^{C_e \times C_e}$. The similarity α_{ij} between point i and point j:

$$\alpha_{ij} = \mathbf{W}_g G_i \cdot \mathbf{W}_g G_j. \tag{3}$$

Afterwards, we calculate the influence factor of point j on point i:

$$\beta_{ij} = \mathrm{softmax}_j(\mathrm{LeakyReLU}(\alpha_{ij})), \tag{4}$$

where β_{ij} is regarded as the normalized attentive weight, representing how point j relates to point i. The representation of each point is updated by attentively aggregating the point representations with reference to β_{ij}:

$$\tilde{G}_i = \sum_{j=1}^{N_e} \beta_{ij} \mathbf{W}_g G_j. \tag{5}$$

1.3 Prediction

Based on the obtained semantic representations, we resort to both the **channel-wise attention** and **spatial-wise attention** modules to further exploit the global structure of the point cloud. Afterwards, the semantics label is produced for each point.

2 Results

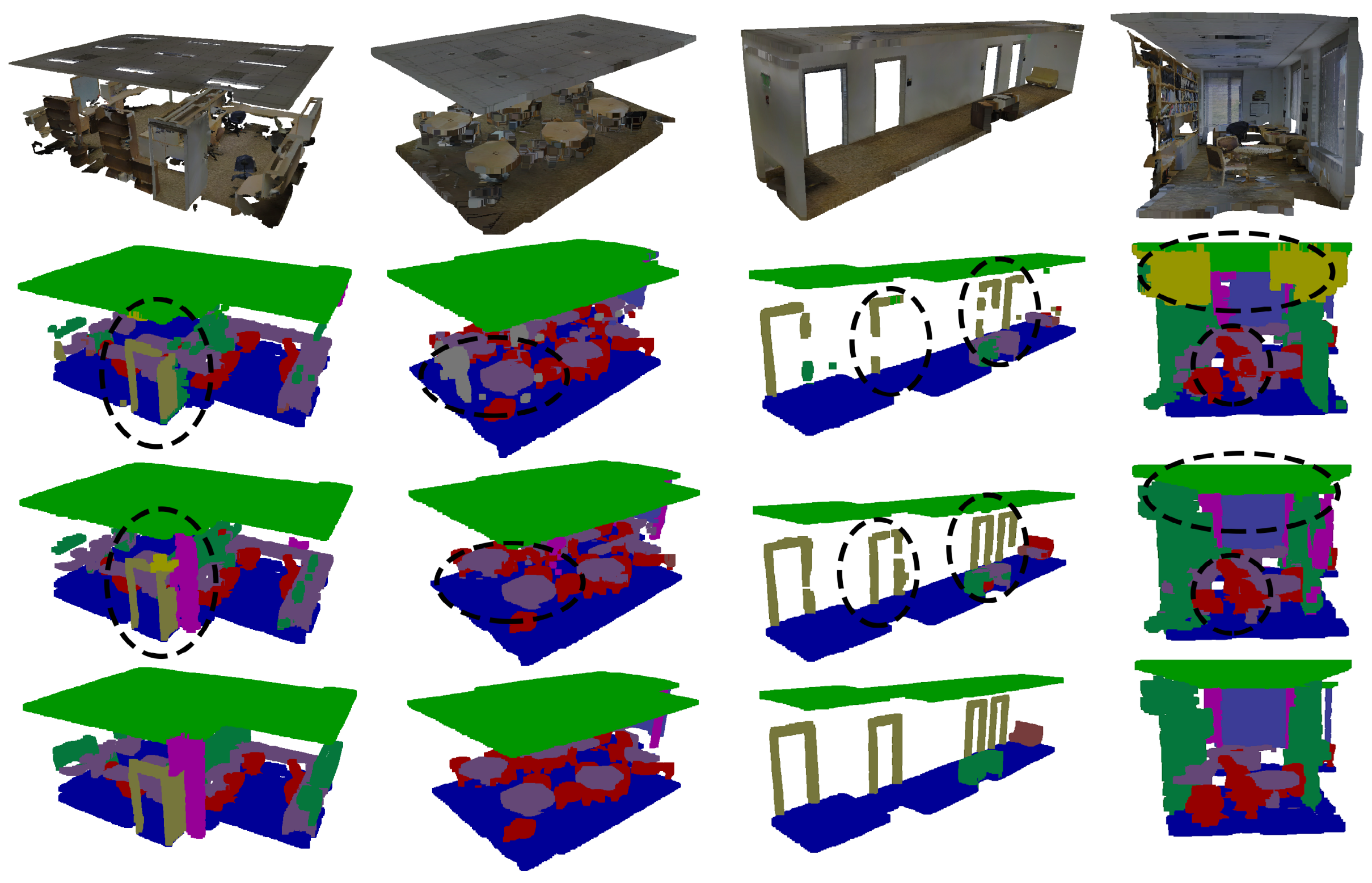

Figure 4: Examples from S3DIS dataset: All the walls are removed for better visualization of interior structure. The segmentation results of our proposed method is closer to the ground truth than that of PointNet++. For top to bottom are the Point Cloud, PointNet++, Ours, Ground Truth, respectively.

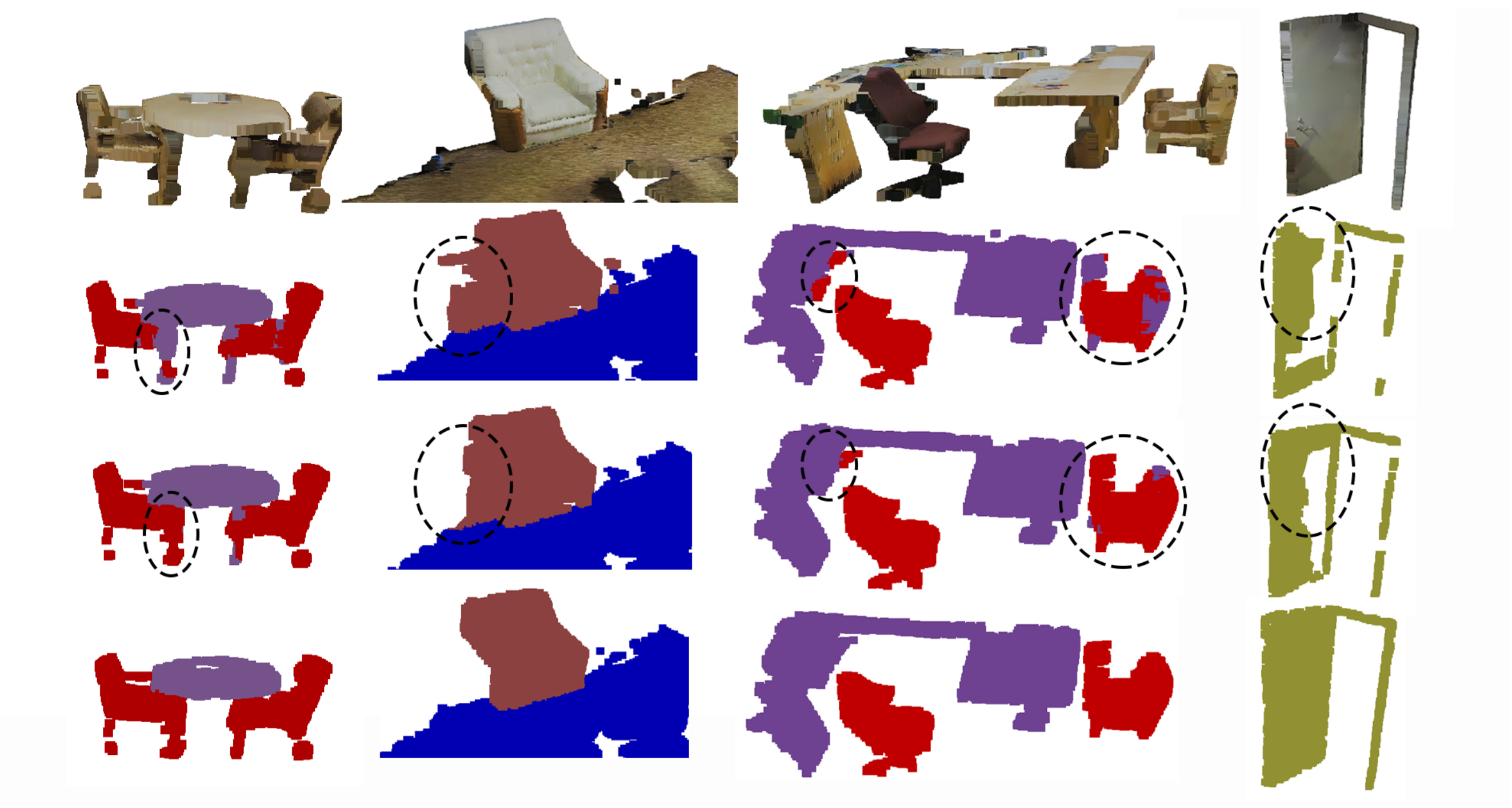

Figure 5: Examples from S3DIS dataset: The segmentation results of our proposed method is closer to the ground truth than that of PointNet++. For top to bottom are the Point Cloud, PointNet++, Ours, Ground Truth, respectively.

Test Area	Method	ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter
Area5	PointNet	88.80	97.33	69.80	**0.05**	3.92	46.26	10.76	52.61	58.93	40.28	5.85	26.38	33.22
	SEGCloud	90.06	96.05	69.86	0.00	18.37	38.35	23.12	75.89	70.40	58.42	40.88	12.96	41.60
	RSNet	**93.34**	98.36	79.18	0.00	15.75	45.37	50.10	65.52	67.87	22.45	52.45	41.02	43.64
	PointNet++	91.41	97.92	69.45	0.00	16.27	66.13	14.48	72.32	81.10	35.12	59.67	**59.45**	51.42
	SPGraph	89.35	96.87	**78.12**	0.00	**42.81**	48.93	**61.58**	**84.66**	75.41	**69.84**	52.60	2.10	52.22
	Ours	92.80	**98.48**	72.65	0.01	32.42	**68.12**	28.79	74.91	**85.12**	55.89	**64.93**	47.74	**58.22**
6fold	PointNet	88.0	88.7	69.3	42.4	23.1	47.5	51.6	42.0	54.1	38.2	9.6	29.4	35.2
	Engelmann *et al.*	90.3	92.1	67.9	44.7	24.2	52.3	51.2	47.4	58.1	39.0	6.9	30.0	41.9
	SPGraph	89.9	95.1	76.4	**62.8**	**47.1**	55.3	68.4	**73.5**	69.2	**63.2**	45.9	8.7	52.9
	Ours	**93.7**	**95.6**	**76.9**	42.6	46.7	**63.9**	**69.0**	70.1	**76.0**	52.8	**57.2**	**54.8**	**62.5**

Table 1: Results of S3DIS dataset: IoU of every category.

Method	ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter
Ours(w/o CR)	92.62	98.69	69.65	0.00	7.81	66.02	21.92	74.64	84.38	29.94	62.53	66.52	55.19
Ours(w/o AM)	92.30	97.91	70.98	0.00	21.40	65.43	31.58	75.16	83.26	48.80	62.68	56.84	56.45
Ours(w/o GPM)	92.17	98.75	72.29	0.00	14.89	72.30	19.70	75.78	84.61	36.48	62.73	68.01	54.25
Ours	92.80	98.48	72.65	0.01	32.42	68.12	28.79	74.91	85.12	55.89	64.93	47.74	58.22

Table 2: Ablation studies and analysis: IoU for each category