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

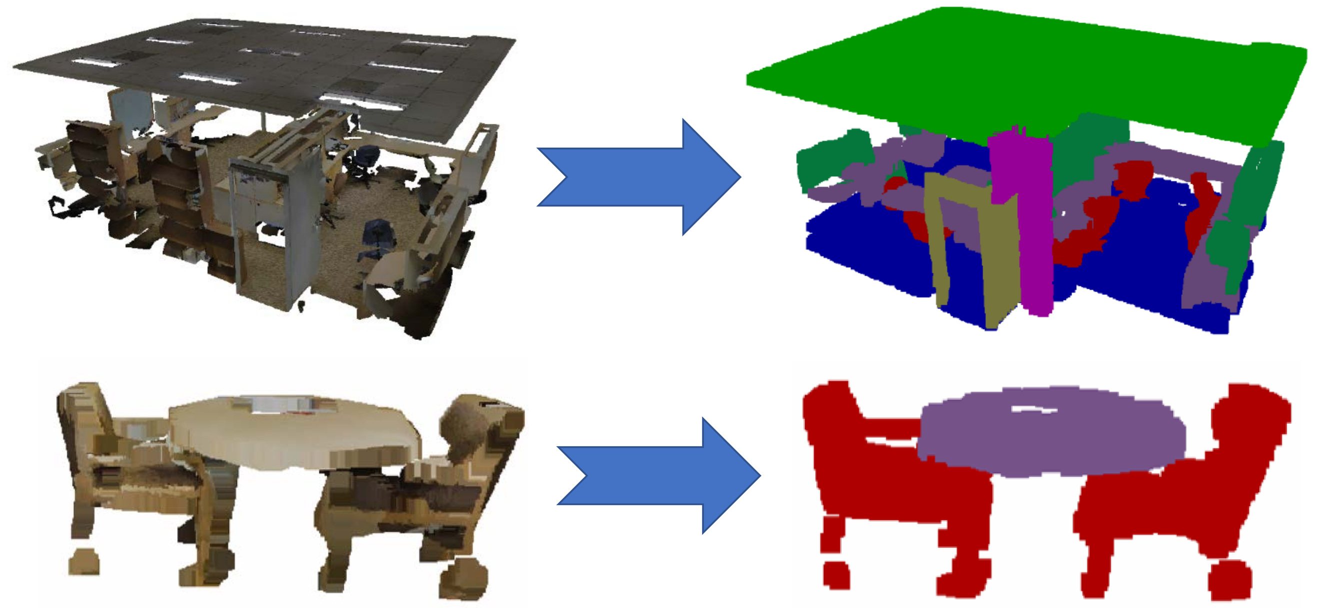
https://github.com/fly519/ELGS

Xu Wang[1], Jingming He[1], Lin Ma[2]

[1]Shenzhen University; [2]Tencent AI Lab

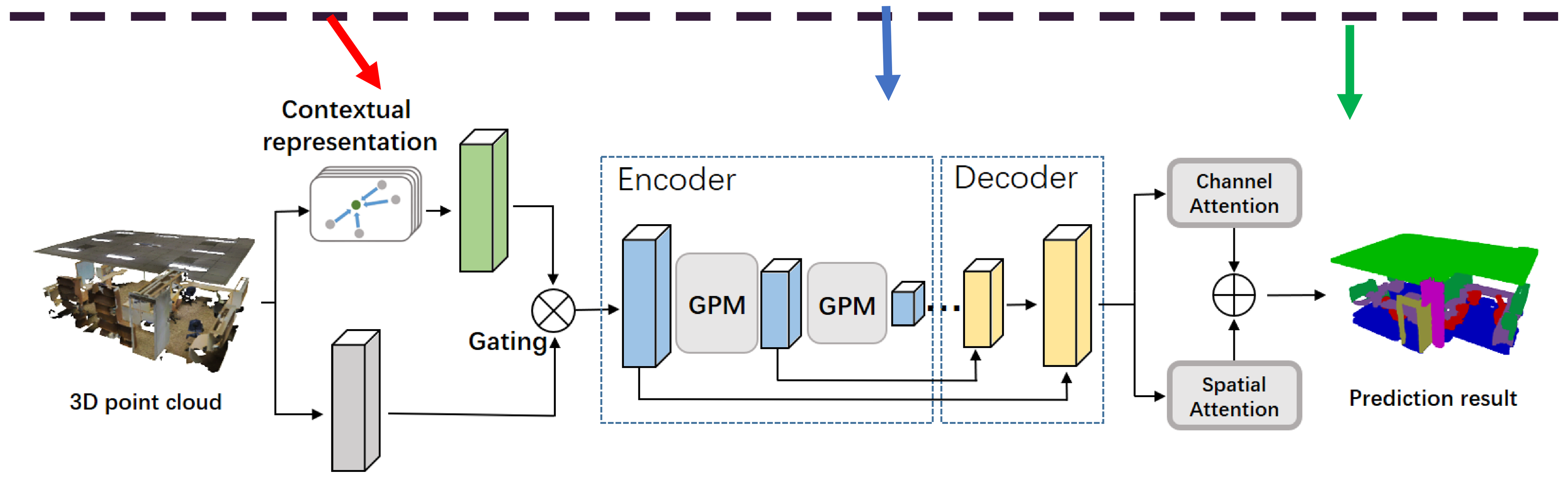
wangxu@szu.edu.cn; {hejingming519,forest.linma}@gmail.com

1 Point Cloud Semantic Segmentation



The point cloud semantic segmentation aims to take the 3D point cloud as input and assign one semantic class label for each point.

2 Proposed Model: Our proposed network consists of three components, (1) point enrichment, (2) feature representation, (3) prediction

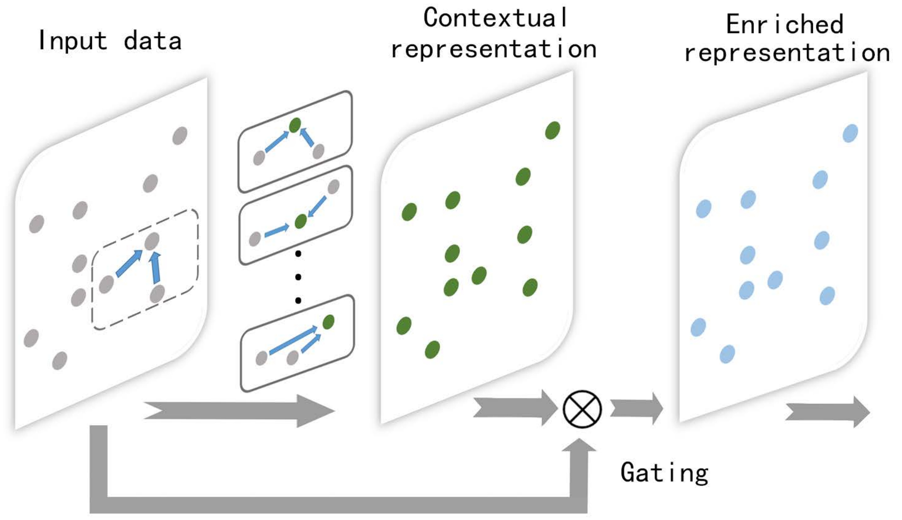


Previous set-based methods that only consider the raw coordinate and attribute information of each single point, we pay more attentions on the spatial context information within neighbor points.

2.1 Point Enrichment

(1) The local region centered on i-th point

$$R_i = \underset{j \in \mathcal{N}_i}{\|} P_j$$



(2) Gated fusion strategy

For each point, we have two different representations R_i and P_i

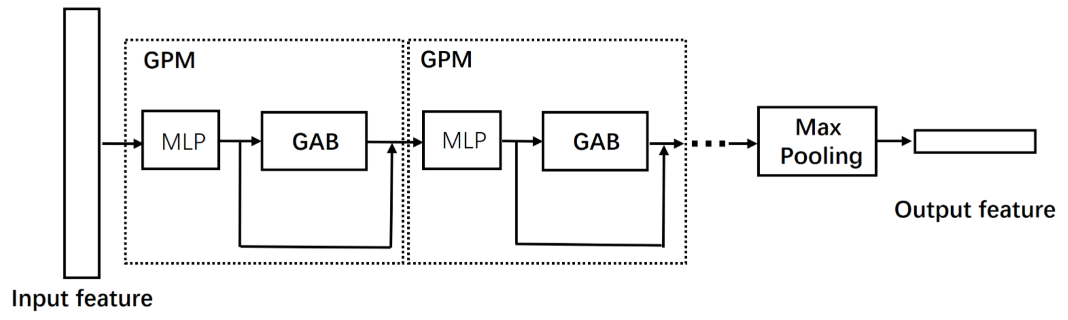
The gated fusion operation is performed:

$$g_i = \sigma(w_i R_i + b_i), \qquad \hat{P}_i = g_i \odot \tilde{P}_i,$$

$$g_i^R = \sigma(w_i^R \tilde{P}_i + b_i^R), \qquad \hat{R}_i = g_i^R \odot R_i,$$

As such, the i-th point representation is then enriched by concatenating them together as $\hat{P}_i \| \hat{R}_i$

2.2 Feature Representation



Graph Pointnet Module

(1) Similarity α_{ij} between point i and point j

$$\alpha_{ij} = \hat{G}_i \cdot \hat{G}_j.$$

(2) Influence factor of point j on point i:

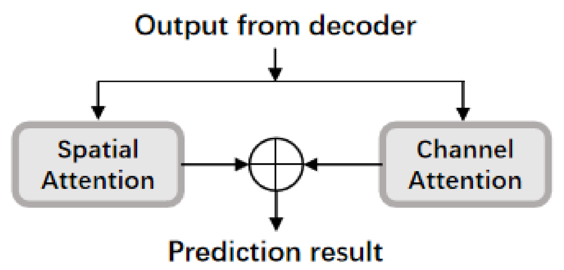
$$\beta_{ij} = \mathrm{softmax}_j(\mathrm{LeakyReLU}(\alpha_{ij})),$$

(3) Update representation by attentively aggregating the point representations with β_{ij}

$$\tilde{G}_i = \sum_{j=1}^{N_e} \beta_{ij} \hat{G}_j.$$

2.3 Prediction

Spatial-wise & Channel-wise Attention



To capture:

- the global context information for each point
- the inter-dependencies between feature channels

$$v_{ij} = \mathrm{softmax}_j(A_i \cdot B_j),$$

$$\hat{F}_i = \sum_{j=1}^{N_d} (v_{ij} D_j) + F_i.$$

3 Experimental Results

Table 1: Results of S3DIS dataset on "Area 5" and over 6 fold in terms of OA and mIoU. † and ‡ indicate that the PointNet performances are directly copied from [8] and [3], respectively. * indicates that the PointNet++ performances are produced with the publicly available code.

Test Area	Method	OA	mIoU
Area5	PointNet† [13]	-	41.09
	SEGCloud [19]	-	48.92
	RSNet [6]	-	51.93
	PointNet++* [14]	86.43	54.98
	SPGraph [8]	86.38	58.04
	Ours	**88.43**	**60.06**
6 fold	PointNet‡ [13]	78.5	47.6
	SGPN [23]	80.8	50.4
	Engelmann et al. [3]	81.1	49.7
	A-SCN [26]	81.6	52.7
	SPGraph [8]	85.5	62.1
	DGCNN [24]	84.3	56.1
	Ours	**87.6**	**66.3**

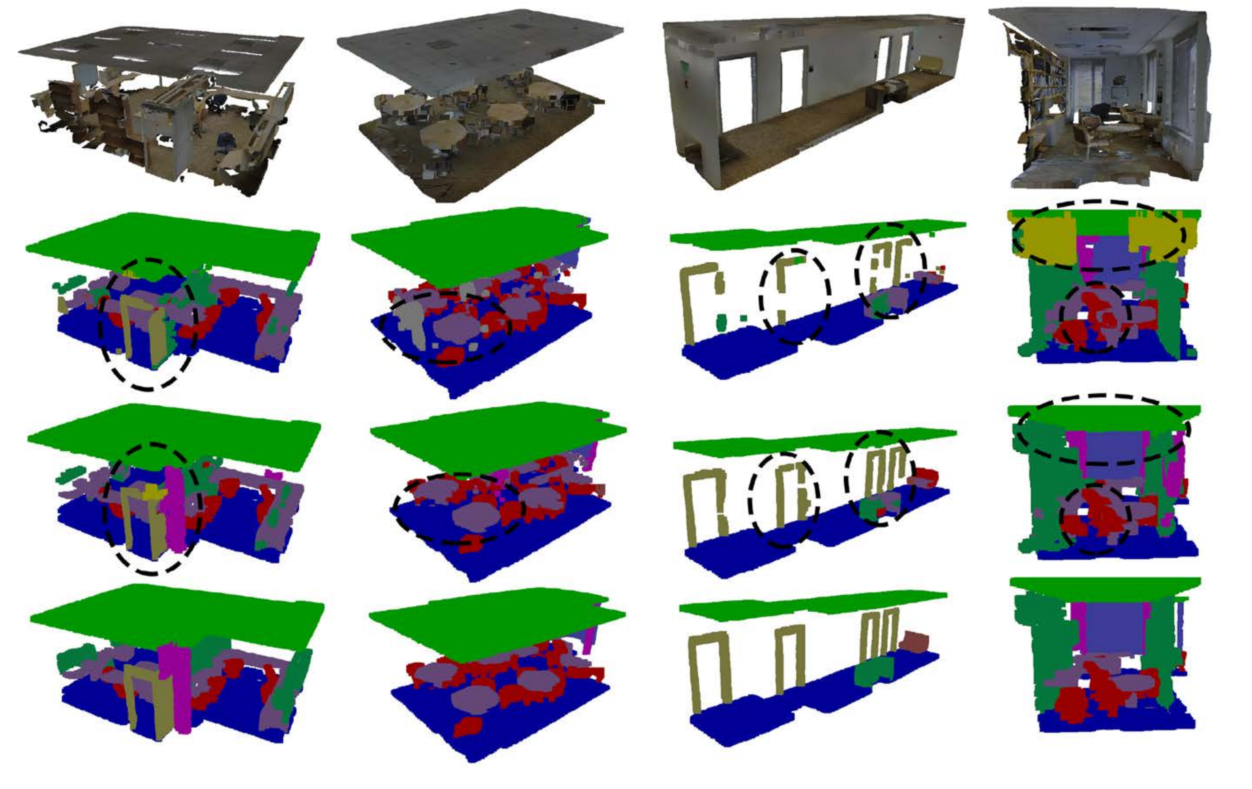


Figure 4: Qualitative results from the S3DIS dataset. All the walls are removed for better visualization. From top to bottom are the result of the Point Cloud, PointNet++, Ours, and Ground Truth, respectively. The segmentation results of our proposed model is closer to the ground truth than that of PointNet++.

Table 2: The segmentation results of S3DIS dataset in terms of IoU for each category.

Test Area	Method	ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter
Area5	PointNet [13] in [8]	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 [19]	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 [6]	**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++ [14]	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 [8]	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 [13] in [3]	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.* [3]	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 [8]	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 4: Ablation studies in terms of OA and mIoU.

Method	OA	mean IoU
Ours(w/o CR)	87.91	56.15
Ours(w/o GPM)	87.74	57.84
Ours(w/o AM)	87.90	58.67
Ours(CR with concatenation)	88.21	59.14
Ours	**88.43**	**60.06**

Table 3: The segmentation results of ScanNet dataset in terms of both OA and mIoU.

Method	OA	mIoU
3DCNN [2]	73.0	-
PointNet [13]	73.9	-
PointNet++ [14]	84.5	38.28
RSNet [6]	-	39.35
PointCNN [10]	85.1	-
Ours	**85.3**	**40.6**

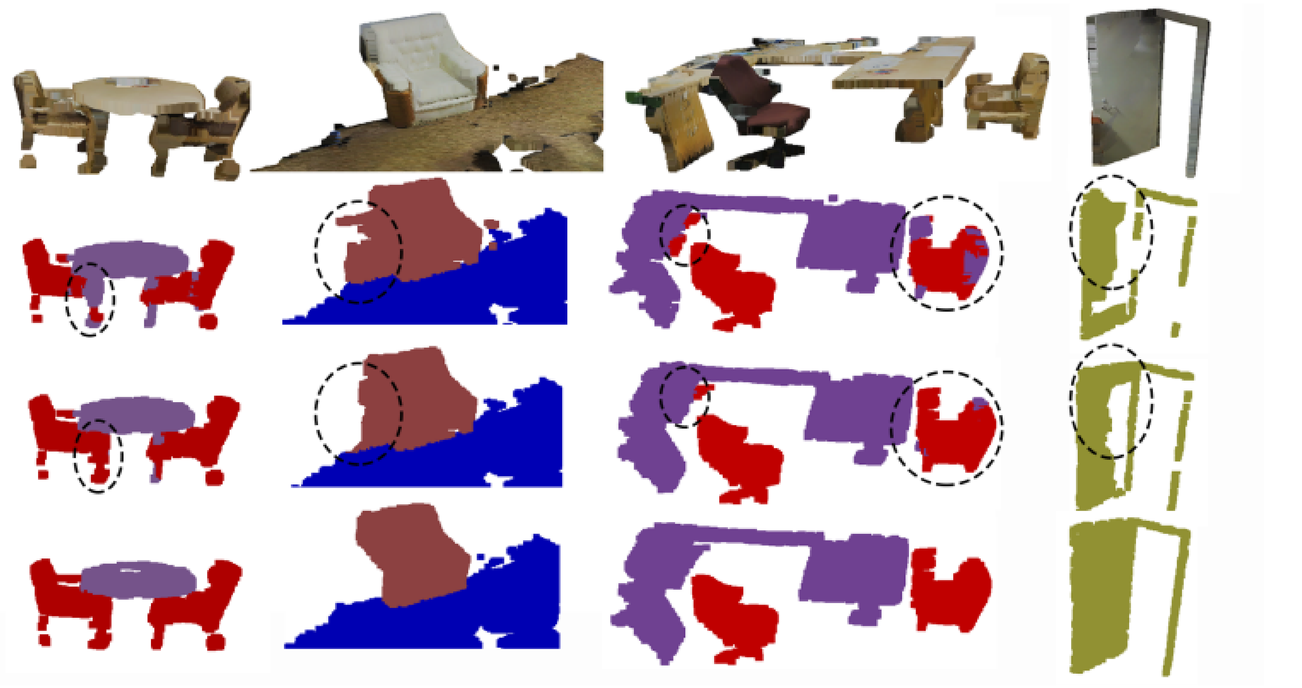


Figure 5: Qualitative results from the S3DIS dataset. From top to bottom are the result of the Point Cloud, PointNet++, Ours, and Ground Truth, respectively. The segmentation results of our proposed model is closer to the ground truth than that of PointNet++.