FPConv: Learning Local Flattening for Point Convolution Supplementary Material

The supplementary material contains:

- The results of the proposed fusion model between FPConv and PointConv [9] on Scannet [2].
- The results of the proposed fusion model between FPConv and KPConv-*deform* [8] on S3DIS [1].
- Comparison of trainable parameters between different convolution operators.
- More qualitative and quantitative results on large-scale scene segmentation tasks.

A. More results of the proposed fusion strategy

a. Fusing FPConv and PointConv on ScanNet

We conduct experiments on fusion of FPConv with PointConv [9] on ScanNet [2]. The results are reported in Table 1, where all methods are performed under same settings (architecture, hyper parameters, etc.). Note that we reduce sampled points to 8k in a block of 1.5m $\times$ 1.5m for all experiments.

Method	mIoU	mA	oA
PointConv [9]	55.6	-	-
PointConv†	60.3	72.3	83.6
FPConv (ours)	63.0	75.6	85.3
FPConv $\otimes$ PointConv	**64.2**	**76.1**	**86.0**

Table 1: Quantitative results of the segmentation task on evaluation dataset of ScanNet. PointConv† indicates our re-implementation of PointConv [9].

b. Fusing FPConv and KPConv-*deform* on S3DIS

We further report the results of the proposed fusion model between FPConv and KPConv-*deform* [8] on S3DIS [1] in Table.3, where the results of each class are also shown. As seen, the proposed fusion model wins all existing methods, reaching the state-of-the-art.

B. Parameter Comparison

We compared the trainable parameters of PointConv, FPConv and their fusion forms in Table.2. We can see that fusion of convolution operators of same type cannot bring significant improvement and even get worse. While for the fusion of different types (FPConv $\otimes$ PointConv), even if we reduce the channel size of the fusion block, it still performs much better than before the fusion.

Method	mIoU	parameters
PointConv†	60.3	4.5
FPConv (ours)	63.4	4.8
FPConv $\otimes$ PointConv	64.2	7.7
FPConv $\otimes$ PointConv + mid ch / 2	**65.1**	**3.8**
PointConv $\otimes$ PointConv	60.7	7.6
FPConv $\otimes$ FPConv	62.9	7.8

Table 2: Comparison of trainable parameters between different convolution operators on ScanNet evaluation dataset. † indicates our implementation. + mid ch /2 is halving the middle channel size of bottleneck in residual block.

C. More Results on Segmentation Tasks

We provide more details of our experimental results. As shown in Table.4, we compare our FPConv with other popular methods on S3DIS [1] 6-fold cross validation, which shows that FPConv can achieve higher score on flat-shaped objects, such like ceiling, floor, table, board, etc. While KPConv [8], a volumetric-style method, performs better on complex structures. More visual results are shown in Fig.1 and Fig.2.

References

[1] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2016. 1

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-

Method	mIoU	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board	clut.
KPConv-*rigid* [8]	65.4	92.6	97.3	81.4	0.0	16.5	54.4	**69.5**	80.2	90.1	66.4	74.6	63.7	58.1
KPConv-*deform* [8]	67.1	92.8	97.3	**82.4**	0.0	**23.9**	58.0	69.0	**81.5**	**91.0**	**75.4**	**75.3**	66.7	**58.9**
FPConv (ours)	62.8	**94.6**	**98.5**	80.9	0.0	19.1	**60.1**	48.9	80.6	88.0	53.2	68.4	**68.2**	54.9
FP $\oplus$ KP-*rigid*	66.7	**94.5**	**98.6**	**83.9**	0.0	24.5	61.1	**70.9**	81.6	89.4	60.3	73.5	**70.8**	57.8
FP $\oplus$ KP-*deform*	**68.2**	94.2	98.5	83.7	0.0	**24.7**	**63.0**	66.6	**82.5**	**91.9**	**76.7**	**75.6**	70.5	**59.1**

Table 3: Fusion results on S3DIS area 5. $\oplus$ indicates fusing in final feature level.

Method	mIoU	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board	clut.
PointNet [7]	47.6	88.0	88.7	69.3	42.4	23.1	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
RSNet [3]	56.5	92.5	92.8	78.6	32.8	34.4	51.6	68.1	59.7	60.1	16.4	50.2	44.9	52.0
SPGraph [5]	62.1	89.9	95.1	76.4	62.8	47.1	55.3	68.4	69.2	73.5	45.9	63.2	8.7	52.9
PointCNN [6]	65.4	94.8	97.3	75.8	63.3	51.7	58.4	57.2	69.1	71.6	61.2	39.1	52.2	58.6
HPEIN [4]	67.8	-	-	-	-	-	-	-	-	-	-	-	-	-
KPConv-*rigid* [8]	69.6	93.7	92.0	82.5	62.5	49.5	65.7	**77.3**	64.0	57.8	71.7	68.8	60.1	59.6
KPConv-*deform* [8]	**70.6**	93.6	92.4	**83.1**	**63.9**	**54.3**	**66.1**	76.6	64.0	57.8	**74.9**	**69.3**	61.3	60.3
FPConv (ours)	68.7	**94.8**	**97.5**	82.6	42.8	41.8	58.6	73.4	**71.0**	**81.0**	59.8	61.9	**64.2**	**64.2**

Table 4: Detailed semantic segmentation scores on S3DIS 6-fold cross validation.

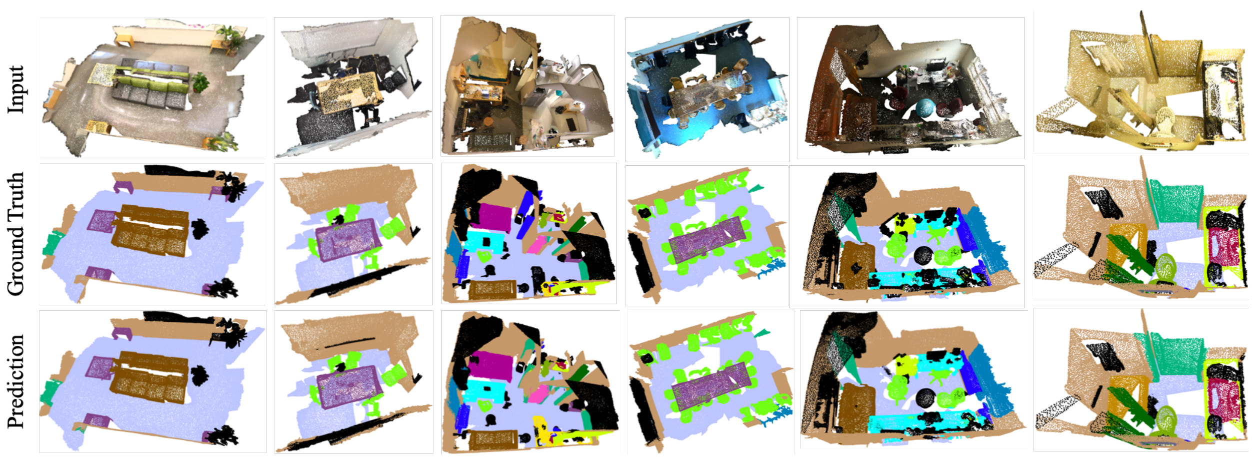

Figure 1: Visualization of semantic segmentation results of FPConv on ScanNet.

annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5828–5839, 2017. 1

[3] Qiangui Huang, Weiyue Wang, and Ulrich Neumann. Recurrent slice networks for 3d segmentation of point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2626–2635, 2018. 2

[4] Li Jiang, Hengshuang Zhao, Shu Liu, Xiaoyong Shen, Chi-Wing Fu, and Jiaya Jia. Hierarchical point-edge interaction network for point cloud semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10433–10441, 2019. 2

[5] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4558–4567, 2018. 2

[6] Yangyan Li, Rui Bu, Mingchao Sun, and Baoquan Chen. Pointcnn. *arXiv preprint arXiv:1801.07791*, 2018. 2

[7] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017. 2

[8] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. *arXiv preprint arXiv:1904.08889*, 2019. 1, 2

[9] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of*

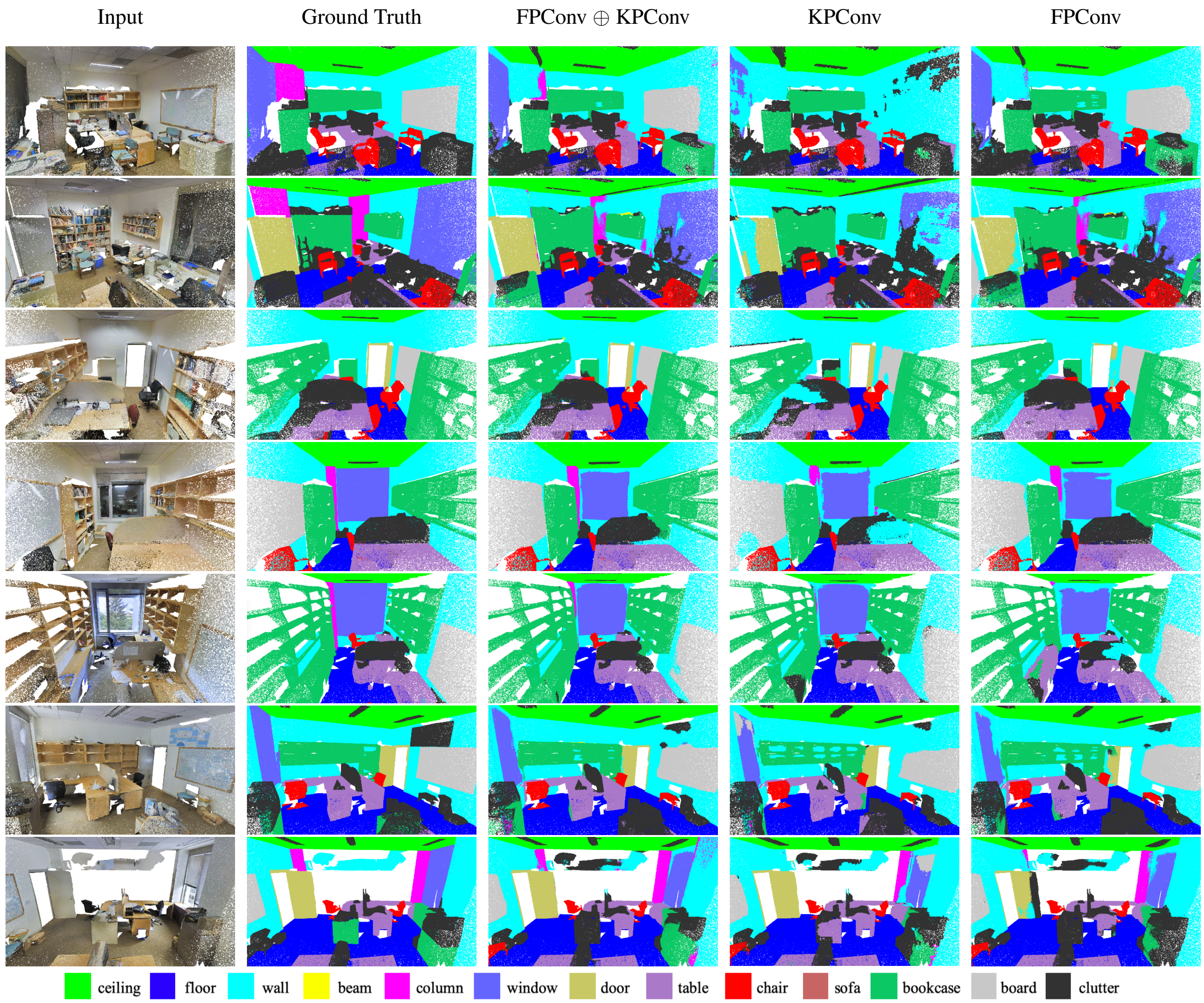

Figure 2: Qualitative comparisons of semantic segmentation tasks on S3DIS area 5. $\oplus$ indicates fusing in final feature level.

the IEEE Conference on Computer Vision and Pattern Recognition, pages 9621–9630, 2019. 1