Supplementary: Virtual Multi-view Fusion for 3D Semantic Segmentation

Abhijit Kundu, Xiaoqi Yin, Alireza Fathi, David Ross,
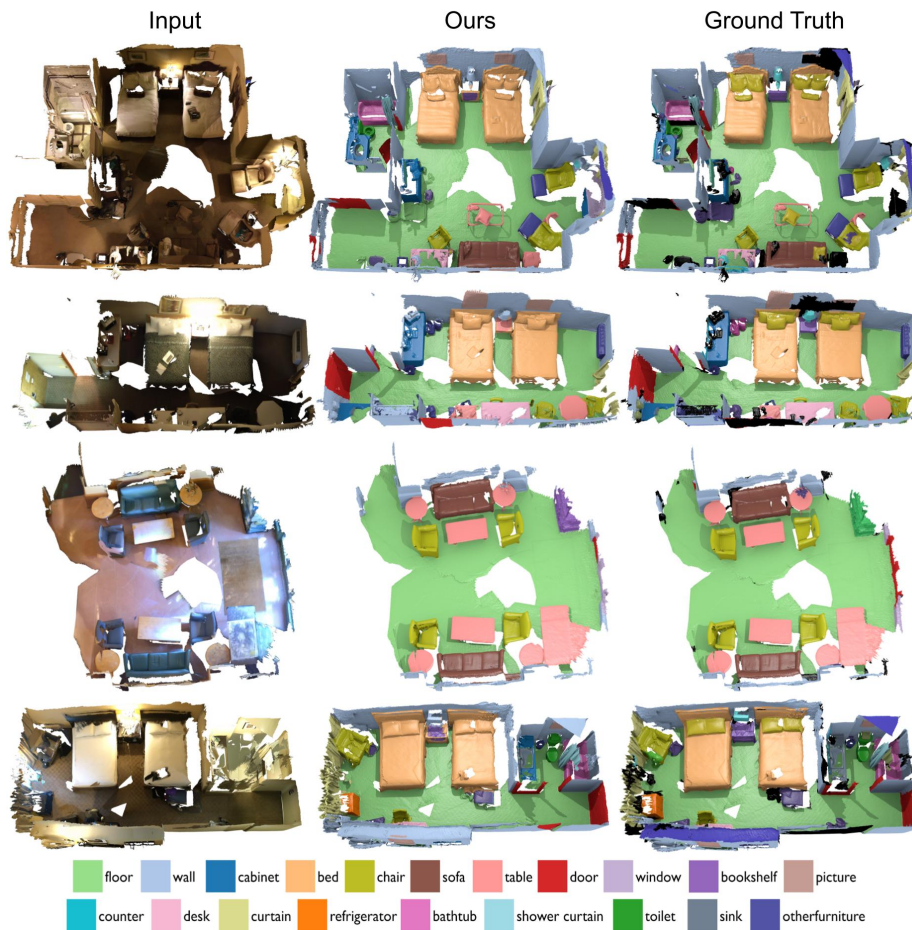Brian Brewington, Thomas Funkhouser, and Caroline Pantofaru

Google Research

Fig. 1: Qualitative 3D semantic segmentation results on ScanNet validation set for scenes scene0645_00, scene0221_00, scene0549_00, and scene0435_01 respectively.
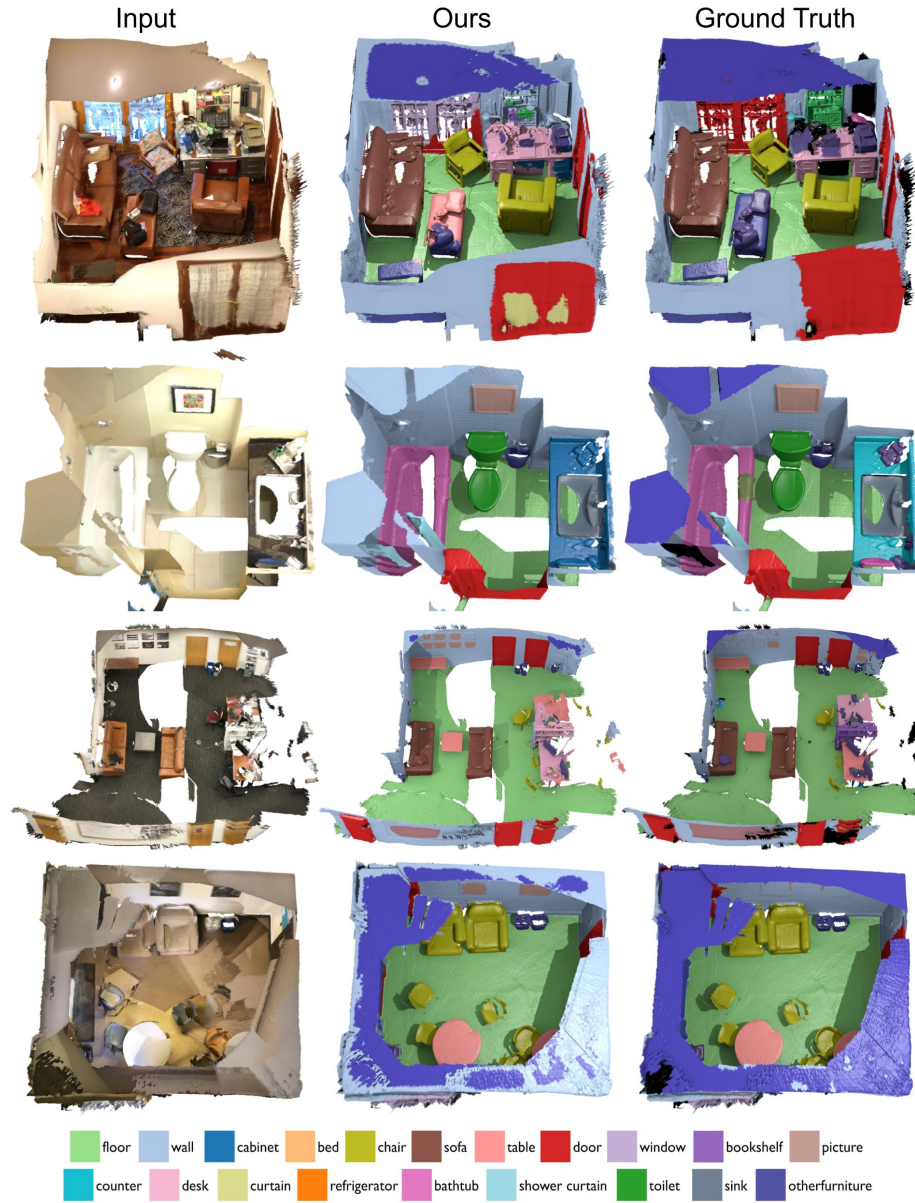
Fig. 2: Qualitative 3D semantic segmentation results on ScanNet validation set for scenes scene0050_00, scene0664_01, scene0329_02, and scene0616_01 respectively.

Method	mean IoU	bathtub	bed	bookshelf	cabinet	chair	counter	curtain	desk	door	floor	other furniture	picture	refrigerator	shower curtain	sink	sofa	table	toilet	wall	window
PointNet [5]	55.7	73.5	66.1	68.6	49.1	74.4	39.2	53.9	45.1	37.5	94.6	37.6	20.5	40.3	35.6	55.3	64.3	49.7	82.4	75.6	51.5
PointConv [8]	66.6	78.1	75.9	69.9	64.4	82.2	47.5	77.9	56.4	50.4	95.3	42.8	20.3	58.6	75.4	66.1	75.3	58.8	90.2	81.3	64.2
PointASNL [9]	66.6	70.3	78.1	75.1	65.5	83.0	47.1	76.9	47.4	53.7	95.1	47.5	27.9	63.5	69.8	67.5	75.1	55.3	81.6	80.6	70.3
KPConv [6]	68.4	84.7	75.8	78.4	64.7	81.4	47.3	77.2	60.5	59.4	93.5	45.0	18.1	58.7	80.5	69.0	78.5	61.4	88.2	81.9	63.2
JSENet [4]	69.9	88.1	76.2	82.1	66.7	80.0	52.2	79.2	61.3	60.7	93.5	49.2	20.5	57.6	85.3	69.1	75.8	65.2	87.2	82.8	64.9
SparseConvNet [3]	72.5	64.7	82.1	84.6	72.1	86.9	53.3	75.4	60.3	61.4	95.5	57.2	32.5	71.0	87.0	72.4	82.3	62.8	93.4	86.5	68.3
MinkowskiNet [1]	73.6	85.9	81.8	83.2	70.9	84.0	52.1	85.3	66.0	64.3	95.1	54.4	28.6	73.1	89.3	67.5	77.2	68.3	87.4	85.2	72.7
Ours	**74.6**	77.1	81.9	84.8	70.2	86.5	39.7	89.9	69.9	66.4	94.8	58.8	33.0	74.6	85.1	76.4	79.6	70.4	93.5	86.6	72.8

Table 1: Per class 3D semantic labeling results on the ScanNet test split.

Method	mean IoU	bathtub	bed	bookshelf	cabinet	chair	counter	curtain	desk	door	floor	other furniture	picture	refrigerator	shower curtain	sink	sofa	table	toilet	wall	window
3DMV [2]	49.8	48.1	61.2	57.9	45.6	34.3	38.4	62.3	52.5	38.1	84.5	25.4	26.4	55.7	18.2	58.1	59.8	42.9	76.0	66.1	44.6
AdapNet++ [7]	50.3	61.3	72.2	41.8	35.8	33.7	37.0	47.9	44.3	36.8	90.7	20.7	21.3	46.4	52.5	61.8	65.7	45.0	78.8	72.1	40.8
SSMA [7]	57.7	69.5	71.6	43.9	56.3	31.4	44.4	71.9	55.1	50.3	88.7	34.6	34.8	60.3	35.3	70.9	60.0	45.7	90.1	78.6	59.9
Ours	**74.5**	86.1	83.9	88.1	67.2	51.2	42.2	89.8	72.3	71.4	95.4	45.4	50.9	77.3	89.5	75.6	82.0	65.3	93.5	89.1	72.8

Table 2: Per class 2D semantic labeling results on the ScanNet test split.

References

1. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3075–3084 (2019)
2. Dai, A., Nießner, M.: 3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 452–468 (2018)
3. Graham, B., Engelcke, M., van der Maaten, L.: 3d semantic segmentation with submanifold sparse convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9224–9232 (2018)
4. Hu, Z., Zhen, M., Bai, X., Fu, H., Tai, C.l.: Jsenet: Joint semantic segmentation and edge detection network for 3d point clouds. In: ECCV (2020)
5. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
6. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 6411–6420 (2019)
7. Valada, A., Mohan, R., Burgard, W.: Self-supervised model adaptation for multimodal semantic segmentation. International Journal of Computer Vision pp. 1–47 (2019)
8. Wu, W., Qi, Z., Fuxin, L.: Pointconv: Deep convolutional networks on 3d point clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9621–9630 (2019)
9. Yan, X., Zheng, C., Li, Z., Wang, S., Cui, S.: Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. arXiv preprint arXiv:2003.00492 (2020)