# Supplementary Material: Omni-supervised Point Cloud Segmentation via Gradual Receptive Field Component Reasoning

Jingyu Gong[1] Jiachen Xu[1] Xin Tan[1] Haichuan Song[2]
Yanyun Qu[3] Yuan Xie[2*] Lizhuang Ma[1,2*]
[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
[2]School of Computer Science and Technology, East China Normal University, Shanghai, China
[3]School of Informatics, Xiamen University, Fujian, China
{gongjingyu,xujiachen,tanxin2017}@sjtu.edu.cn hcsong@cs.ecnu.edu.cn
yyqu@xmu.edu.cn yxie@cs.ecnu.edu.cn ma-lz@cs.sjtu.edu.cn

## Abstract

*This supplementary material consists of five parts. In Sec 1, we study the impact of supervisions at different scales on the segmentation performance. We visualize the intermediate RFCC prediction to show the intermediate feature learning in Sec 2. Then, we compare the performance of omni-supervision on decoder and encoder in Sec 3. Later, we provide more visual results of ScanNet v2, S3DIS and Semantic3D in Sec 4. Finally, Sec 5 lists the detailed experimental results of these three datasets shown in the main paper.*

## 1. Supervisions at Different Layers

| Supervision Scales | | | | | mIoU |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | |
| ✓ | ✓ | ✓ | ✓ | | 76.3 |
| ✓ | ✓ | ✓ | | ✓ | 76.6 |
| ✓ | ✓ | | ✓ | ✓ | 76.9 |
| ✓ | | ✓ | ✓ | ✓ | 76.2 |
| ✓ | ✓ | ✓ | ✓ | ✓ | 77.8 |

Table 1. Ablation study on significance of supervisions at different scales.

We design an omni-scale supervision method for point cloud segmentation via the proposed gradual Receptive Field Component Reasoning in the main paper. All scales are supervised in the decoding stage to learn informative representation for semantic segmentation. In this section, we attempt to analyze the significance of supervisions at different scales. In this ablation study, deformable KPConv [19] is also taken as the backbone and performance is evaluated on the Semantic3D reduced-8 task. In the architecture of deformable KPConv network, there are 5 different scales as shown in Figure 1. So, we separately remove the supervisions for $l = 2, 3, 4, 5$. It is noteworthy that we always keep the supervision for the final layer ($l = 1$) because it directly guides the semantic label prediction, otherwise the network will give random prediction. The results is reported in Table 1. The results indicates supervision in the center-most layer ($l = 5$) plays an important role in the omni-scale supervision. That is because it can help the encoder to obtain representative global features which is quite important for the following reasoning. Meanwhile, the supervision before the final prediction $l = 2$ also contributes a lot because it can directly provide semantic informative

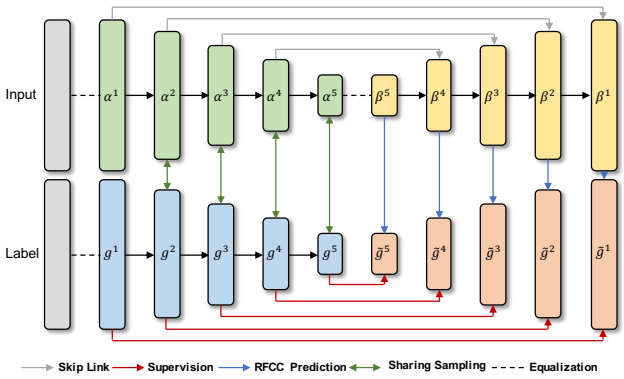


Figure 1. Illustration of framework using deformable KPConv as the backbone. In our method, all the five scales are supervised by the target RFCCs.

*Corresponding Author

features to the final segmentation.

## 2. Visualization of intermediate RFCC

We visualize the RFCC reasoning process and our predicted RFCCs in intermediate layers to implicitly show the intermediate feature learning in Figure 2. Meanwhile, the OA of RFCC prediction is 97.34% on the validation set of ScanNet v2, demonstrating good representation learning of intermediate features to some extent.

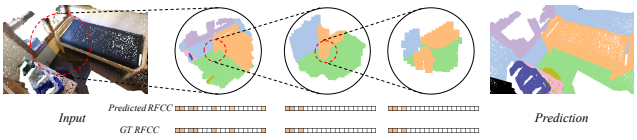


Figure 2. Visualization of intermediate RFCCs whose element color represents the probability of existence for each category.

## 3. Supervision on Decoder vs. Encoder

| Method | mIoU |
|---|---|
| KPConv *deform* | 73.1 |
| KPConv *deform* + [RFCR + FD][encoder] | 76.8 |
| KPConv *deform* + RFCR + FD | 77.8 |

Table 2. More ablation study on the strategy of omni-scale supervision.

In our implementation, all the supervisions are added in the decoder even the target RFCCs are generated according to the receptive fields of features in the encoder. That is because the features in the encoder can also be supervised through the skip links. In order to show the advantage of our strategy, we attempt to supervise the features in the encoder rather than the decoder according to the RFCCs, and Feature Densification is also applied on the corresponding features in the encoder. Compared with supervision in the decoding stage, guiding the feature extraction using RFCCs in the encoder is not able to effectively extract informative representation from global and local features in the decoding stage, such obtaining inferior result as reported in Table 2.

## 4. Visualization Results

In this section, we present more visualization results of our method on the three datasets described in the main paper. We present more visualization results of our baseline and our methods on the validation set of ScanNet v2 [1] in Figure 3. In Figure 4, we provide additional visualization results to show the qualitative improvement over the baseline in S3DIS Area 5. We also visualize more scenes in the validation set of Semantic3D in Figure 5.

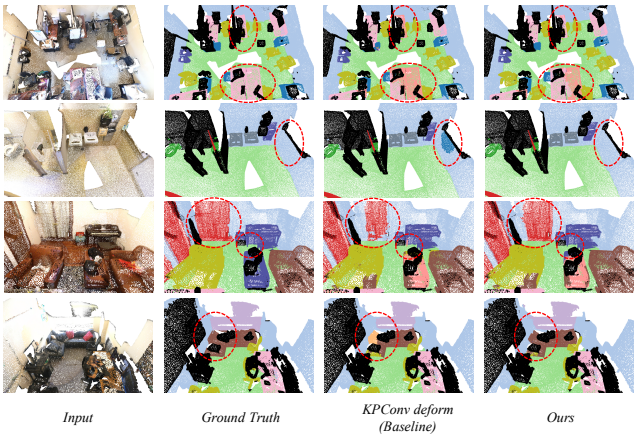


Figure 3. More visualization results on the validation dataset of ScanNet v2. The images from the left to right are input point clouds, semantic labels, predictions given by our baseline and our method, respectively.

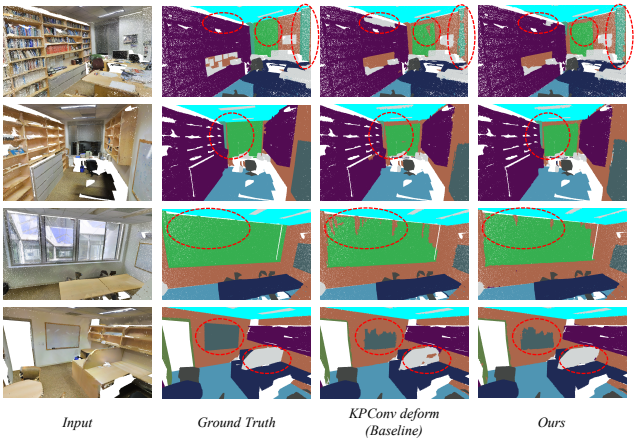


Figure 4. More visualization results on the test dataset of the S3DIS Area-5. The left-most images are inputs and the following images are segmentation ground truth, predictions of baseline and our method separately.

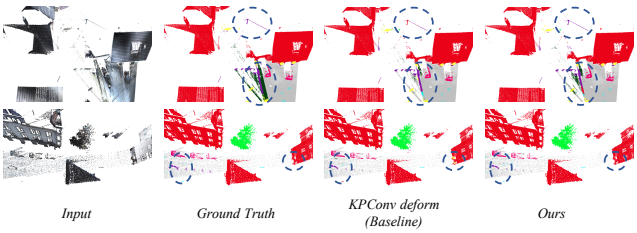


Figure 5. More visualization results on the validation dataset of Semantic3D. Input point clouds, semantic labels, results of our baseline and our method are presented respectively from left to right.

## 5. Detailed Experimental Results

In this section, we provide more quantitative details about our experimental results for better comparison with

| Method | mIoU | bath. | bed | bksf. | cab. | chair | ctr. | curt. | desk | door | floor | oth. | pic. | ref. | shw. | sink | sofa | tab. | toil. | wall | win. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PointNet++ (NIPS'17) [15] | 33.9 | 58.4 | 47.8 | 45.8 | 25.6 | 36.0 | 25.0 | 24.7 | 27.8 | 26.1 | 67.7 | 18.3 | 11.7 | 21.2 | 14.5 | 36.4 | 34.6 | 23.2 | 54.8 | 52.3 | 25.2 |
| PointCNN (NIPS'18) [12] | 45.8 | 57.7 | 61.1 | 35.6 | 32.1 | 71.5 | 29.9 | 37.6 | 32.8 | 31.9 | 94.4 | 28.5 | 16.4 | 21.6 | 22.9 | 48.4 | 54.5 | 45.6 | 75.5 | 70.9 | 47.5 |
| 3DMV (ECCV'18) [2] | 48.4 | 48.4 | 53.8 | 64.3 | 42.4 | 60.6 | 31.0 | 57.4 | 43.3 | 37.8 | 79.6 | 30.1 | 21.4 | 53.7 | 20.8 | 47.2 | 50.7 | 41.3 | 69.3 | 60.2 | 53.9 |
| PointConv (CVPR'19) [23] | 55.6 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| TextureNet (CVPR'19) [5] | 56.6 | 67.2 | 66.4 | 67.1 | 49.4 | 71.9 | 44.5 | 67.8 | 41.1 | 39.6 | 93.5 | 35.6 | 22.5 | 41.2 | 53.5 | 56.5 | 63.6 | 46.4 | 79.4 | 68.0 | 56.8 |
| HPEIN (ICCV'19) [7] | 61.8 | 72.9 | 66.8 | 64.7 | 59.7 | 76.6 | 41.4 | 68.0 | 52.0 | 52.5 | 94.6 | 43.2 | 21.5 | 49.3 | 59.9 | 63.8 | 61.7 | 57.0 | **89.7** | 80.6 | 60.5 |
| SegGCN (CVPR'20) [10] | 58.9 | 83.3 | 73.1 | 53.9 | 51.4 | 78.9 | 44.8 | 46.7 | 57.3 | 48.4 | 93.6 | 39.6 | 6.1 | 50.1 | 50.7 | 59.4 | 70.0 | 56.3 | 87.4 | 77.1 | 49.3 |
| SPH3D-GCN (TPAMI'20) [11] | 61.0 | 85.8 | 77.2 | 48.9 | 53.2 | 79.2 | 40.4 | 64.3 | 57.0 | 50.7 | 93.5 | 41.4 | 4.6 | 51.0 | 70.2 | 60.2 | 70.5 | 54.9 | 85.9 | 77.3 | 53.4 |
| FusionAwareConv (CVPR'20) [28] | 63.0 | 60.4 | 74.1 | 76.6 | 59.0 | 74.7 | **50.1** | 73.4 | 50.3 | 52.7 | 91.9 | 45.4 | 32.3 | 55.0 | 42.0 | 67.8 | 68.8 | 54.4 | 89.6 | 79.5 | 62.7 |
| FPConv (CVPR'20) [13] | 63.9 | 78.5 | 76.0 | 71.3 | 60.3 | 79.8 | 39.2 | 53.4 | 60.3 | 52.4 | 94.8 | 45.7 | 25.0 | 53.8 | 72.3 | 59.8 | 69.6 | 61.4 | 87.2 | 79.9 | 56.7 |
| DCM-Net (CVPR'20) [16] | 65.8 | 77.8 | 70.2 | 80.6 | 61.9 | 81.3 | 46.8 | 69.3 | 49.4 | 52.4 | 94.1 | 44.9 | 29.8 | 51.0 | 82.1 | 67.5 | 72.7 | 56.8 | 82.6 | 80.3 | 63.7 |
| PointASNL (CVPR'20) [25] | 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 |
| FusionNet (ECCV'20) [27] | 68.8 | 70.4 | 74.1 | 75.4 | 65.6 | 82.9 | **50.1** | 74.1 | 60.9 | 54.8 | 95.0 | **52.2** | **37.1** | 63.3 | 75.6 | **71.5** | 77.1 | 62.3 | 86.1 | 81.4 | **65.8** |
| SceneEncoder (IJCAI'20) [24] | 62.8 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| SceneEncoder + Ours | 65.9 | 69.1 | 72.4 | 69.6 | 63.2 | 81.5 | 47.7 | 75.4 | **64.6** | 50.9 | **95.2** | 42.8 | 28.4 | 56.6 | 76.1 | 62.6 | 71.1 | 61.0 | 88.9 | 79.3 | 61.0 |
| KPConv *deform* (ICCV'19) [19] | 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 |
| KPConv *deform* + Ours | **70.2** | **88.9** | 74.5 | **81.3** | **67.2** | 81.8 | 49.3 | **81.5** | 62.3 | **61.0** | 94.7 | 47.0 | 24.9 | 59.4 | **84.8** | 70.5 | 77.9 | **64.6** | 89.2 | **82.3** | 61.1 |

Table 3. Semantic segmentation results on ScanNet v2.

| Method | mIoU | ceil. | floor | wall | beam | col. | wind. | door | chair | table | book. | sofa | board | clut. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PointNet (CVPR'17) [15] | 41.09 | 88.80 | 97.33 | 69.80 | 0.05 | 3.92 | 46.26 | 10.76 | 58.93 | 52.61 | 5.85 | 40.28 | 26.38 | 33.22 |
| RSNet (CVPR'18) [6] | 51.93 | 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 |
| PointCNN (NIPS'18) [12] | 57.26 | 92.31 | 98.24 | 79.41 | 0.00 | 17.6 | 22.77 | 62.09 | 74.39 | 80.59 | 31.67 | 66.67 | 62.05 | 56.74 |
| ASIS (CVPR'19) [22] | 53.40 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| ELGS (NIPS'19) [21] | 60.06 | 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 |
| PAT (CVPR'19) [26] | 60.07 | 93.04 | 98.51 | 72.28 | **1.00** | **41.52** | **85.05** | 38.22 | 57.66 | 83.64 | 48.12 | 67.00 | 61.28 | 33.64 |
| SPH3D-GCN (TPAMI'20) [11] | 59.5 | 93.3 | 97.1 | 81.1 | 0.0 | 33.2 | 45.8 | 43.8 | 79.7 | 86.9 | 33.2 | 71.5 | 54.1 | 53.7 |
| PointASNL (CVPR'20) [25] | 62.6 | 94.3 | 98.4 | 79.1 | 0.0 | 26.7 | 55.2 | 66.2 | 83.3 | 86.8 | 47.6 | 68.3 | 56.4 | 52.1 |
| FPConv (CVPR'20) [13] | 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 |
| Point2Node (AAAI'20) [3] | 62.96 | 93.88 | 98.26 | 83.30 | 0.00 | 35.65 | 55.31 | 58.78 | 79.51 | 84.67 | 44.07 | 71.13 | 58.72 | 55.17 |
| SegGCN (CVPR'20) [10] | 63.6 | 93.7 | **98.6** | 80.6 | 0.0 | 28.5 | 42.6 | **74.5** | 80.9 | **88.7** | 69.0 | 71.3 | 44.4 | 54.3 |
| DCM-Net (CVPR'20) [16] | 64.0 | 92.1 | 96.8 | 78.6 | 0.0 | 21.6 | 61.7 | 54.6 | 78.9 | **88.7** | 68.1 | 72.3 | 66.5 | 52.4 |
| FusionNet (ECCV'20) [27] | 67.2 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| RandLA (CVPR'20) [4] | 62.42 | 91.19 | 95.66 | 80.11 | 0.00 | 25.24 | 62.27 | 47.36 | 75.78 | 83.17 | 60.82 | 70.82 | 65.15 | 53.95 |
| RandLA+Ours | 65.09 | 92.66 | 97.43 | 82.40 | 0.00 | 37.04 | 59.72 | 52.30 | 77.49 | 86.95 | 63.48 | 71.99 | 70.54 | 54.13 |
| KPConv *deform* (ICCV'19) [19] | 67.1 | 92.8 | 97.3 | 82.4 | 0.0 | 23.9 | 58.0 | 69.0 | 91.0 | 81.5 | 75.3 | **75.4** | 66.7 | 58.9 |
| KPConv *deform*+Ours | **68.73** | 94.18 | 98.33 | **84.34** | 0.00 | 28.45 | 62.36 | 71.17 | **91.95** | 82.60 | **76.13** | 71.14 | **71.60** | **61.25** |

Table 4. Results of indoor scene semantic segmentation on S3DIS Area-5.

| Method | mIoU | man-made. | natural. | high veg. | low veg. | buildings | hard scape | scanning. | cars |
|---|---|---|---|---|---|---|---|---|---|
| SegCloud (3DV'17) [17] | 61.3 | 83.9 | 66.0 | 86.0 | 40.5 | 91.1 | 30.9 | 27.5 | 64.3 |
| RF_MSSF (3DV'18) [18] | 62.7 | 87.6 | 80.3 | 81.8 | 36.4 | 92.2 | 24.1 | 42.6 | 56.6 |
| SPG (CVPR'18) [9] | 73.2 | 97.4 | **92.6** | 87.9 | 44.0 | 93.2 | 31.0 | 63.5 | 76.2 |
| ShellNet (ICCV'19) [29] | 69.4 | 96.3 | 90.4 | 83.9 | 41.0 | 94.2 | 34.7 | 43.9 | 70.2 |
| GACNet (CVPR'19) [20] | 70.8 | 86.4 | 77.7 | **88.5** | 60.6 | 94.2 | 37.3 | 43.5 | 77.8 |
| FGCN (CVPR'20) [8] | 62.4 | 90.3 | 65.2 | 86.2 | 38.7 | 90.1 | 31.6 | 28.8 | 68.2 |
| PointGCR (WACV'20) [14] | 69.5 | 93.8 | 80.0 | 64.4 | **66.4** | 93.2 | 39.2 | 34.3 | **85.3** |
| RandLA (CVPR'20) [4] | 77.4 | 95.6 | 91.4 | 86.6 | 51.5 | **95.7** | **51.5** | 69.8 | 76.8 |
| KPConv *rigid* (ICCV'19) [19] | 74.6 | 90.9 | 82.2 | 84.2 | 47.9 | 94.9 | 40.0 | 77.3 | 79.7 |
| KPConv *deform* + Ours | 77.6 | **97.0** | 90.9 | 86.7 | 50.8 | 94.5 | 37.3 | **79.7** | 84.1 |
| KPConv *deform* (ICCV'19) [19] | 73.1 | - | - | - | - | - | - | - | - |
| KPConv *deform* + Ours | **77.8** | 94.2 | 89.1 | 85.7 | 54.4 | 95.0 | 43.8 | 76.2 | 83.7 |

Table 5. Semantic segmentation results on Semantic3D (reduced-8).

other competitors. In Table 3, we present the mean IoU (mIoU) over categories and the IoUs for different classes for ScanNet v2. We also list the category scores for S3DIS Area-5 in Table 4. It's noteworthy that all the methods do not have good performance on the segmentation of beams in Area 5 because there is a large difference between the beams in Area 5 (test set) and those in Area 1, 2, 3, 4, and 6 (training set). Finally, Table 5 shows the IoUs of various

classes for Semantic3D reduced-8 task.

## References


[1] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5828–5839, 2017.

[2] Angela Dai and Matthias Nießner. 3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 452–468, 2018.

[3] Wenkai Han, Chenglu Wen, Cheng Wang, Xin Li, and Qing Li. Point2node: Correlation learning of dynamic-node for point cloud feature modeling. In *Thirty-fourth AAAI conference on artificial intelligence*, 2020.

[4] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[5] Jingwei Huang, Haotian Zhang, Li Yi, Thomas Funkhouser, Matthias Nießner, and Leonidas J Guibas. Texturenet: Consistent local parametrizations for learning from high-resolution signals on meshes. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4440–4449, 2019.

[6] Qiangui Huang, Weiyue Wang, and Ulrich Neumann. Recurrent slice networks for 3d segmentation of point clouds. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2626–2635, 2018.

[7] 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 *The IEEE International Conference on Computer Vision (ICCV)*, pages 10433–10441, 2019.

[8] Saqib Ali Khan, Yilei Shi, Muhammad Shahzad, and Xiao Xiang Zhu. Fgcn: Deep feature-based graph convolutional network for semantic segmentation of urban 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[9] 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 (CVPR)*, pages 4558–4567, 2018.

[10] Huan Lei, Naveed Akhtar, and Ajmal Mian. Seggcn: Efficient 3d point cloud segmentation with fuzzy spherical kernel. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11611–11620, 2020.

[11] Huan Lei, Naveed Akhtar, and Ajmal Mian. Spherical kernel for efficient graph convolution on 3d point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2020.

[12] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 820–830, 2018.

[13] Yiqun Lin, Zizheng Yan, Haibin Huang, Dong Du, Ligang Liu, Shuguang Cui, and Xiaoguang Han. Fpconv: Learning local flattening for point convolution. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4293–4302, 2020.

[14] Yanni Ma, Yulan Guo, Hao Liu, Yinjie Lei, and Gongjian Wen. Global context reasoning for semantic segmentation of 3d point clouds. In *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2931–2940, 2020.

[15] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, 2017.

[16] Jonas Schult, Francis Engelmann, Theodora Kontogianni, and Bastian Leibe. Dualconvmesh-net: Joint geodesic and euclidean convolutions on 3d meshes. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[17] Lyne Tchapmi, Christopher Choy, Iro Armeni, JunYoung Gwak, and Silvio Savarese. Segcloud: Semantic segmentation of 3d point clouds. In *2017 international conference on 3D vision (3DV)*, pages 537–547. IEEE, 2017.

[18] Hugues Thomas, François Goulette, Jean-Emmanuel Deschaud, Beatriz Marcotegui, and Yann LeGall. Semantic classification of 3d point clouds with multiscale spherical neighborhoods. In *2018 International Conference on 3D Vision (3DV)*, pages 390–398. IEEE, 2018.

[19] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *The IEEE International Conference on Computer Vision (ICCV)*, pages 6411–6420, 2019.

[20] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10296–10305, 2019.

[21] Xu Wang, Jingming He, and Lin Ma. Exploiting local and global structure for point cloud semantic segmentation with contextual point representations. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 4571–4581, 2019.

[22] Xinlong Wang, Shu Liu, Xiaoyong Shen, Chunhua Shen, and Jiaya Jia. Associatively segmenting instances and semantics in point clouds. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4096–4105, 2019.

[23] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9621–9630, 2019.

[24] Jiachen Xu, Jingyu Gong, Jie Zhou, Xin Tan, Yuan Xie, and Lizhuang Ma. Sceneencoder: Scene-aware semantic segmentation of point clouds with a learnable scene descriptor.

In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.

[25] Xu Yan, Chaoda Zheng, Zhen Li, Sheng Wang, and Shuguang Cui. Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[26] Jiancheng Yang, Qiang Zhang, Bingbing Ni, Linguo Li, Jinxian Liu, Mengdie Zhou, and Qi Tian. Modeling point clouds with self-attention and gumbel subset sampling. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3323–3332, 2019.

[27] Feihu Zhang, Jin Fang, Benjamin Wah, and Philip Torr. Deep fusionnet for point cloud semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[28] Jiazhao Zhang, Chenyang Zhu, Lintao Zheng, and Kai Xu. Fusion-aware point convolution for online semantic 3d scene segmentation. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4534–4543, 2020.

[29] Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit Yeung. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In *The IEEE International Conference on Computer Vision (ICCV)*, pages 1607–1616, 2019.