Appendix

A. Details of Feature Extractor

The feature extractor of our GrowSP simply follows the successful SparseConv [12] architecture. Particularly, we use the existing implementation of MinkowskiEngine PyTorch package [8]. As illustrated in 7, the encoder uses Res16, and the decoder instead consists of 4 MLP layers which produce 128-dimensional features for interpolations to obtain the final per-point features.

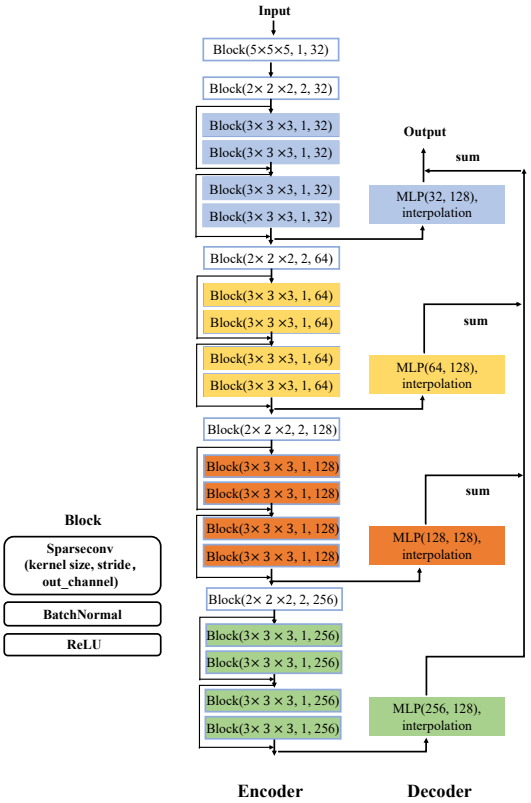

Figure 7. Details of feature extractor (SparseConv [12] with the Res16 architecture).

B. Details of K-means

In both progressive growing of superpoints and semantic primitive clustering, we adopt the simple K-means algorithm. Particularly, we use the existing package provided by Scikit-learn. *https://scikitlearn.org/stable/modules/generated/sklearn.cluster.KMeans.html*

C. Initial Superpoints Construction

As discussed in Section 3.2 of the main paper, initial superpoints are generated by VCCS [40] combined with a Region Growing algorithm [1], which jointly consider the spatial/normal/normalized RGB information of 3D points to construct initial superpoints for indoor 3D point clouds.

Details of VCCS [40]: VCCS incrementally expands superpoints from a set of seed points distributed evenly in 3D space on a voxel grid with resolution R_{seed}.

In our experiments of S3DIS and ScanNet datasets, we firstly voxelize input point clouds into $5 \times 5 \times 5$ cm voxel grids. Secondly, a set of seed points are distributed evenly in the voxelized point clouds with an interval of 50cm. For each seed, in its 50cm radius sphere, we set the seed point as an initial center and search its 27 neighbors, and compute a distance between each neighbor and the center as below:

$$\boldsymbol{D} = \sqrt{w_c D_c^2 + \frac{w_s D_s}{3R_{seed}^2} + w_n D_n} \tag{3}$$

where D_c, D_s, D_n are color/spatial/normal Euclidean distances. We assign the point with smallest distance into the superpoint associated with the current center. Iteratively, we set the newly added points as new centers to increase the superpoint until it meets the sphere boundary. In our experiments, the w_c, w_s, w_n are set as 0.2, 0.4, 1. The interval of 50cm is the main parameter controlling the superpoint size, we add ablation studies in the main paper Table 7. Figure 8 shows an example of initial superpoints obtained by VCCS.

In implementation, we simply use the existing Point Cloud Library: *https://pcl.readthedocs.io/projects/tutorials/en/latest/supervoxel_clustering.html*

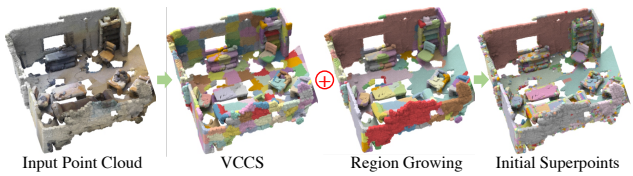

Figure 8. Example of VCCS, Region Growing, Initial Superpoints.

Details of Region Growing Algorithm [1]: This algorithm aims to merge points that are close enough in terms of local smoothness. The output of this algorithm is a set of clusters, where each cluster is a set of points that are considered to be a part of same smooth surface. The evaluation of smoothness is based on the similarity of point normals. Besides, the Region Growing algorithm begins its growth from the point that has the minimum curvature value. This is because the point with the minimum curvature is usually located in the flat area. Figure 8 shows an example.

In implementation, we simply use the existing Point Cloud Library: *https://pcl.readthedocs.io/projects/tutorials/en/latest/region_growing_segmentation.html*

For an input point cloud, both VCCS and the Region Growing algorithm split it into a large number of partitions. We then combine these partitions together. Specifically, for each partition obtained by VCCS, if half of its 3D points are included by a partition obtained by Region Growing, we then merge all points in the former partition to the latter.

Table 9. Quantitative results of our method and baselines on the Area-1 of S3DIS dataset.

		OA(%)	mAcc(%)	mIoU(%)	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board.
Supervised Methods	PointNet [41]	75.4	74.8	55.0	88.3	93.2	69.2	49.5	37.8	74.5	65.6	41.2	42.5	22.3	35.4	40.9
	PointNet++ [42]	76.1	77.9	58.2	90.5	94.4	65.7	38.2	31.9	61.5	66.0	45.3	60.4	41.2	45.8	57.4
	SparseConv [12]	89.0	79.5	72.5	93.6	95.6	76.1	65.9	60.9	60.0	74.2	81.9	85.4	69.2	73.4	33.5
Unsupervised Methods	RandCNN	24.9±1.1	20.3±0.68	10.9±0.9	27.8±8.2	29.0±4.0	16.6±2.0	10.7±5.4	4.7±1.3	14.8±6.4	8.1±2.5	4.1±0.8	3.7±2.0	0.2±0	6.4±1.4	5.1±2.9
	van Kmeans	20.9±1.1	24.1±2.1	10.1±0.9	15.4±2.3	17.8±2.3	10.5±1.7	16.8±0.3	1.9±1.4	16.0±4.1	12.1±3.9	9.9±1.6	8.1±0.2	0.1±0.2	6.2±1.7	6.7±2.2
	van Kmeans-S	21.8±0.9	25.6±2.7	10.8±0.6	18.1±0.5	20.4±0.2	9.5±0.6	14.5±3.7	1.0±0.5	17.2±3.5	16.2±4.7	9.6±1.4	6.8±1.0	0.6±1.1	8.2±1.3	7.6±0.8
	van Kmeans-PFH	26.4±0	25.7±0.1	12.6±0	34.0±0.4	27.4±0.1	11.6±0.3	17.0±0.1	5.2±0.1	15.4±0.6	9.7±0.8	10.8±0	5.8±0.1	1.6±0	8.1±0.1	4.3±0.8
	van Kmeans-S-PFH	25.1±0.6	23.0±1.8	10.8±0.7	26.1±7.0	18.5±5.3	15.0±1.4	4.6±7.0	2.6±2.0	20.5±3.9	15.9±2.3	10.4±1.4	6.0±0.5	0.6±0.7	5.9±3.4	3.7±2.4
	IIC [24]	29.2±2.1	14.3±0.4	8.0±0.8	17.0±2.9	31.4±8.0	25.6±0.1	4.3±0.1	11.1±0.4	0±0	2.6±0.8	1.4±0.2	0.7±0.6	0±0	0.2±0.1	1.4±2.1
	IIC-S [24]	31.3±0.2	21.1±0.2	8.2±0.1	35.5±0.3	11.6±1.0	13.9±0.2	1.3±0.4	0.2±0.2	13.0±0.4	15.6±0.2	5.1±0.1	0.6±0	0.3±0.6	0.3±0	7.6±0.1
	IIC-PFH [24]	27.3±1.9	12.0±1.1	6.1±0.3	17.1±2.4	17.0±7.7	23.4±4.9	4.4±1.1	0.3±0.2	0.8±1.0	4.0±3.7	0.4±0.3	0.4±0.4	0±0	4.9±5.7	0.1±0.1
	IIC-S-PFH [24]	24.1±0.3	13.2±0.4	7.2±0.2	21.7±0.9	16.8±3.8	18.0±1.0	5.0±0.6	3.6±0.7	2.9±1.7	3.7±0.4	4.4±0.4	0.7±0.2	0±0	4.7±2.8	4.5±1.6
	PICIE [7]	45.7±2.9	28.3±2.4	19.4±1.7	77.2±5.4	63.1±10.9	24.5±2.4	15.8±0.6	3.3±2.8	4.4±0.2	9.6±0.6	10.2±2.4	14.7±0.6	0±0	9.9±0.8	0±0
	PICIE-S [7]	48.9±0.3	30.3±1.6	21.8±1.2	77.66±8.4	58.1±16.8	39.0±2.1	13.6±1.6	2.5±2.2	2.8±0.7	11.8±0.7	20.0±1.9	17.1±1.9	1.8±0.7	17.1±2.3	0.6±0.9
	PICIE-PFH [7]	51.9±3.3	38.4±2.3	27.1±2.1	59.8±1.5	78.7±9.1	37.5±5.5	6.8±1.8	9.5±1.3	4.9±2.5	26.4±2.0	50.2±1.3	29.5±3.0	0±0	22.8±1.3	0.1±0.2
	PICIE-S-PFH [7]	47.9±1.4	38.9±3.0	27.2±2.4	71.5±1.2	66.3±1.9	23.9±2.3	20.4±1.0	8.8±3.4	6.9±2.8	12.9±3.3	39.3±1.9	32.2±1.6	0.2±0.2	42.0±1.2	2.4±5.1
	GrowSP(Ours)	**72.9**±0.8	**60.4**±4.2	**45.6**±1.4	**94.2**±0.2	**90.8**±0.6	**52.7**±1.6	**36.7**±5.8	**19.7**±1.6	**33.3**±4.9	**35.8**±1.3	**66.5**±0.7	**72.6**±4.4	**13.1**±7.5	**31.2**±5.6	**16.7**±1.2

Table 10. Quantitative results of our method and baselines on the Area-2 of S3DIS dataset.

		OA(%)	mAcc(%)	mIoU(%)	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board.
Supervised Methods	PointNet [41]	72.5	55.5	36.6	79.2	87.4	64.9	14.5	8.2	14.8	39.6	28.8	64.0	7.8	24.4	5.1
	PointNet++ [42]	72.1	62.3	39.9	85.8	69.6	71.2	24.9	27.5	32.5	43.6	27.4	51.3	6.0	26.8	12.4
	SparseConv [12]	87.9	69.5	57.3	89.5	93.8	77.0	29.1	32.5	65.5	45.7	67.9	88.8	34.9	54.5	8.15
Unsupervised Methods	RandCNN	20.4±1.2	17.0±2.1	7.6±0.8	16.8±2.4	19.2±4.3	12.8±0.2	1.6±0.3	1.1±0.4	1.5±2.1	9.4±1.7	2.9±0.3	16.1±2.9	0.3±0.4	6.4±0.7	**13.0**±1.0
	van Kmeans	17.6±0.6	16.6±0.6	6.4±0.3	16.4±1.5	15.6±2.3	11.3±0.9	3.3±0.5	0.9±0.1	0.4±0.6	6.8±00.8	3.7±0.6	11.0±0.3	1.4±0.3	4.6±0.7	1.5±0.7
	van Kmeans-S	17.3±0.4	17.6±0.7	6.3±0.2	14.7±1.4	15.0±0.9	12.1±0.3	3.0±1.6	0.9±0.3	0±0	7.0±0.2	3.6±0.4	11.4±0.4	1.9±0.1	4.3±0.8	2.0±1.6
	van Kmeans-PFH	21.8±0.4	22.3±0.3	9.0±0	31.0±0.4	21.1±1.5	10.1±0.2	3.0±0.1	2.6±0	**3.2**±1.0	10.7±0.1	5.3±1.8	9.8±0.2	4.0±0.2	7.2±0.4	0.6±0.9
	van Kmeans-S-PFH	21.5±1.5	18.0±2.1	8.0±0.3	23.7±3.0	21.1±3.3	14.1±1.1	3.8±0.9	1.8±0.2	1.8±1.6	5.4±2.5	4.3±1.5	10.2±0.6	2.5±0.2	6.5±1.4	0.9±0.8
	IIC [24]	41.6±0.3	16.8±0.3	10.6±0.1	33.0±0.6	43.7±3.2	27.6±0.7	1.7±0.7	0±0	0±0	5.6±1.0	0.1±0.1	13.0±2.2	0±0	2.8±0.1	0±0
	IIC-S [24]	30.3±0	11.9±0.1	6.2±0.1	29.3±0.1	7.6±0.2	22.6±0.1	0.5±0.5	1.5±0	0±0	3.5±0	0.3±0.2	2.2±0.2	1.8±1.6	3.4±0.1	1.1±0
	IIC-PFH [24]	27.3±3.0	10.5±0.2	5.8±0.3	16.0±0.6	13.9±3.9	24.9±2.6	0.3±0.2	0.1±0.1	0±0	4.0±0.3	0.6±0.4	8.3±4.2	0.1±0.1	0.8±0.1	0.1±0.1
	IIC-S-PFH [24]	26.8±0.1	12.8±0.3	6.8±0.1	30.5±0.4	12.5±0.7	16.5±0.9	1.1±0.1	0.4±0.2	1.1±0.1	7.0±0.1	1.1±0.1	5.7±0.4	0.4±0.5	4.0±0.3	1.2±0.1
	PICIE [7]	48.3±2.3	27.2±1.5	17.4±0.6	72.4±1.5	44.2±6.9	39.6±1.6	6.2±1.2	1.7±0	0.5±0	7.7±0.3	4.1±0.9	20.1±4.1	0±0	7.7±0.2	3.6±0.4
	PICIE-S [7]	55.5±0.9	30.4±1.0	21.5±0.4	69.1±1.2	68.7±3.3	46.8±2.0	3.1±0.3	1.1±0.9	0.4±0.3	11.6±1.1	6.6±4.4	38.2±2.9	3.0±0.6	7.7±0.7	1.8±1.2
	PICIE-PFH [7]	55.8±3.4	37.2±1.9	25.7±1.4	51.7±6.0	82.6±6.3	41.8±4.8	2.5±1.0	5.3±0.3	1.0±0.6	13.7±1.4	40.6±1.5	55.0±11.7	0±0	12.5±1.3	2.1±0.6
	PICIE-S-PFH [7]	55.4±1.1	39.2±0.8	26.2±0.7	65.8±6.1	77.7±2.2	29.3±4.3	4.6±0.9	2.7±0.8	0.5±0.5	12.4±0.8	26.7±3.5	62.2±7.0	1.0±0.1	28.0±1.0	3.7±0.3
	GrowSP(Ours)	**79.0**±0.8	**51.8**±3.1	**39.1**±1.6	**85.7**±1.2	**88.2**±1.5	**67.0**±1.1	**12.0**±2.0	**24.8**±7.0	0±0	**24.2**±1.6	**51.2**±5.6	**77.1**±4.7	**4.1**±5.7	**24.5**±8.4	0.2±0.3

Table 11. Quantitative results of our method and baselines on the Area-3 of S3DIS dataset.

		OA(%)	mAcc(%)	mIoU(%)	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board.
Supervised Methods	PointNet [41]	78.2	74.9	57.7	90.3	96.9	66.9	55.5	15.1	60.0	67.7	51.8	54.8	27.6	56.0	50.0
	PointNet++ [42]	79.8	85.9	65.8	91.4	98.0	68.5	50.1	15.2	74.8	74.7	63.2	70.1	53.6	54.0	76.5
	SparseConv [12]	91.3	86.8	78.6	93.1	96.2	80.4	74.7	63.3	77.2	69.5	80.1	85.5	89.5	80.1	52.2
Unsupervised Methods	RandCNN	25.3±3.7	21.7±1.9	10.8±1.6	25.5±9.6	32.6±12.3	17.2±0.5	4.9±2.8	3.3±1.0	10.6±3.2	9.1±3.1	4.2±1.9	2.9±2.0	0.9±0.9	9.3±3.1	8.8±1.1
	van Kmeans	21.3±02	22.1±0.9	9.43±0.1	20.2±2.2	20.6±0.96	13.3±0.6	5.7±2.0	1.3±1.8	2.3±0.9	14.1±1.7	6.8±0.3	6.8±0.9	3.7±0.7	9.7±0.7	8.6±1.5
	van Kmeans-S	21.2±1.2	21.9±2.3	9.3±0.8	19.8±0.9	18.8±2.8	14.4±0.6	5.0±0.9	1.6±1.4	1.5±0.6	14.4±2.0	7.4±0.8	6.6±0.6	4.2±0.1	10.3±0.2	7.7±2.7
	van Kmeans-PFH	24.7±0.5	27.3±1.8	11.6±0.4	33.6±1.3	23.0±2.3	11.0±1.1	8.9±0.4	3.5±0.3	10.8±2.0	12.0±2.4	8.4±0.7	6.9±1.1	3.7±0.6	12.6±0.2	4.9±1.0
	van Kmeans-S-PFH	23.3±1.7	21.7±2.9	9.7±1.2	24.7±4.0	20.1±3.4	14.1±1.2	5.4±4.1	2.3±1.2	5.2±5.1	12.3±1.3	5.1±2.4	7.3±0.2	3.1±0.9	9.3±3.2	8.0±1.1
	IIC [24]	32.1±1.4	15.4±1.9	8.4±0.7	20.5±2.4	25.5±6.9	31.4±1.0	1.0±0.9	6.9±5.7	0.2±0.4	3.2±1.1	1.6±0.2	0.3±0.1	0±0	10.6±3.6	0±0
	IIC-S [24]	30.3±0.1	14.8±0.1	7.1±0	33.9±0.2	10.1±0	16.1±0.3	2.0±0	0.4±0	5.1±0	8.8±0.3	0.6±0	0.3±0.5	2.7±0.5	2.1±0.5	5.5±0.2
	IIC-PFH [24]	34.5±4.5	11.8±2.1	6.3±1.8	22.6±2.1	13.0±2.0	32.4±1.9	0.1±0.1	0±0	0.1±0.2	1.9±2.3	0.3±0.2	0.3±0.3	1.0±0.6	2.3±3.1	1.9±0.4
	IIC-S-PFH [24]	25.5±1.3	12.8±0.5	6.6±0.1	26.1±1.3	8.7±1.6	18.8±2.4	1.3±0.6	2.9±1.7	0.8±0.6	5.3±1.9	1.6±1.9	4.8±2.5	1.7±0.6	5.3±2.0	2.6±1.1
	PICIE [7]	40.4±1.8	29.2±0.9	16.2±0.5	50.5±3.7	49.6±14.5	33.7±1.7	13.2±1.7	3.0±0.3	1.8±1.6	6.5±0.99	8.9±7.9	7.5±1.1	3.5±3.0	16.2±1.5	0.4±0.7
	PICIE-S [7]	52.8±1.5	32.3±0.2	22.4±0.4	79.4±2.2	80.0±1.7	38.6±2.9	9.1±0.6	2.3±0.7	3.9±1.9	10.7±1.7	15.9±9.0	11.0±9.2	0±0	17.6±1.4	0.6±0.2
	PICIE-PFH [7]	57.8±3.7	36.7±1.0	26.4±1.6	60.5±1.5	78.8±11.1	53.9±3.5	4.1±2.7	4.8±1.5	0.3±0.6	16.0±0.9	50.0±1.0	25.3±15.0	1.0±1.3	26.4±4.5	0.6±0.4
	PICIE-S-PFH [7]	51.8±1.7	43.3±2.2	29.3±2.5	85.0±0.9	72.6±2.9	23.7±1.6	14.2±0.4	4.0±0.5	5.7±2.0	12.7±2.5	48.5±10.2	32.2±1.9	2.8±1.8	48.3±4.2	2.1±0.3
	GrowSP(Ours)	**74.2**±0.8	**68.4**±1.3	**47.7**±1.2	**92.9**±0.1	**91.7**±0.1	**48.3**±3.3	**49.3**±5.9	**15.8**±2.3	**21.1**±2.0	**38.7**±1.8	**60.6**±1.4	**66.5**±2.9	**28.5**±5.4	**59.2**±6.6	0±0

Table 12. Quantitative results of our method and baselines on the Area-4 of S3DIS dataset.

		OA(%)	mAcc(%)	mIoU(%)	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board.
Supervised Methods	PointNet [41]	73.0	58.6	41.6	81.3	95.7	68.4	1.3	22.4	29.0	44.8	39.3	42.5	17.6	36.6	20.1
	PointNet++ [42]	74.8	66.4	47.7	85.5	96.1	68.9	4.4	23.8	27.0	50.5	44.9	54.0	35.6	38.4	43.8
	SparseConv [12]	88.3	76.2	65.5	93.0	94.9	78.2	53.3	57.9	43.4	59.1	69.4	76.6	55.1	73.8	30.9
Unsupervised Methods	RandCNN	22.6±1.9	18.3±2.0	9.0±1.1	22.0±9.7	28.2±2.8	16.4±1.7	0.4±0 3	4.2±0.5	5.6±0.9	11.9±0.7	4.5±1.3	4.3±0.5	1.4±0.8	6.4±0.9	**2.3**±0.9
	van Kmeans	17.9±0.5	19.9±0.5	7.8±0.2	18.6±1.2	18.2±0.2	10.6±1.0	0.9±0.3	3.8±1.6	5.2±1.3	11.7±1.3	5.8±0.3	7.4±1.8	2.4±0.7	8.7±1.0	0.4±0.6
	van Kmeans-S	17.3±0.5	19.6±1.0	7.4±0.3	16.7±0.5	14.8±1.1	10.2±1.0	0.9±0.3	4.4±1.8	5.1±0.8	10.5±0.8	6.1±0.4	7.5±1.4	2.8±0.2	9.8±0.3	0±0
	van Kmeans-PFH	21.7±0.5	22.3±0.8	9.3±0.3	28.9±1.0	21.3±0.9	12.6±0.7	1.2±0.1	5.9±1.0	4.4±0.8	11.1±0.6	6.6±0.5	8.6±0	2.6±0.7	7.4±0.7	0.7±0.2
	van Kmeans-S-PFH	23.1±3.6	18.3±2.3	8.9±1.3	24.4±6.5	26.2±7.6	14.5±2.5	1.7±1.0	5.0±1.1	3.2±2.0	11.7±0.8	3.8±0.8	6.1±1.3	0.4±0.7	9.0±0.4	1.1±0.3
	IIC [24]	33.0±0.3	13.5±0.5	8.2±0.2	14.9±1.0	25.9±3.9	35.1±0.4	0±0	1.1±0.4	1.4±1.4	3.7±1.0	4.1±2.0	0.9±0.1	0±0	10.8±0.7	0±0
	IIC-S [24]	26.7±0.2	13.5±0.1	6.1±0.1	27.6±0	4.9±0.9	15.3±0	0.1±0.1	1.0±0.1	5.3±0.1	12.0±0	1.5±1.9	2.3±1.9	1.1±0.9	1.6±0	0±0
	IIC-PFH [24]	31.0±7.6	11.4±1.5	6.3±1.4	21.0±1.5	10.9±2.1	30.0±6.2	0.1±0.2	1.3±1.4	3.0±4.0	2.3±0.4	0.8±0.1	1.3±1.2	0±0	4.6±4.1	0.1±0.1
	IIC-S-PFH [24]	27.3±2.0	13.2±0.4	6.9±0.1	29.4±1.9	9.9±0.4	20.0±4.9	0.1±0	2.6±1.5	3.1±1.2	4.0±0.9	3.7±1.1	2.5±1.1	0.6±0.1	6.3±1.9	0.8±0.5
	PICIE [7]	43.2±0.6	29.4±1.2	17.8±0.5	62.2±1.3	72.7±2.1	22.6±0.7	2.5±0.6	3.4±0.2	3.5±0.4	8.8±1.1	4.1±7.0	17.4±1.1	0±0	15.5±1.7	0.7±0.1
	PICIE-S [7]	50.0±1.2	32.2±1.1	22.3±0.4	72.7±2.1	77.7±2.8	37.4±1.6	1.6±1.1	2.4±0.7	3.3±1.3	11.3±1.1	25.8±1.4	19.7±1.5	3.5±1.5	12.3±3.5	0±0
	PICIE-PFH [7]	58.6±2.6	42.6±4.2	28.5±1.3	69.5±6.8	83.6±7.3	48.3±3.2	2.6±1.6	9.4±6.4	5.8±1.0	16.4±1.1	44.1±0.8	36.3±3.4	1.6±0.6	23.9±3.3	0.1±0.2
	PICIE-S-PFH [7]	47.1±1.5	44.8±0.8	27.6±0.5	67.2±3.5	67.8±3.5	25.9±2.2	2.6±0.2	14.5±0.5	5.2±0.8	10.2±0.2	47.9±1.1	44.1±1.0	0.9±0.2	44.2±0.7	0.4±0.7
	GrowSP(Ours)	**76.0**±0.9	**59.8**±1.0	**42.8**±1.1	**90.6**±0.7	**91.5**±0.7	**64.4**±1.1	**15.9**±5.3	**7.60**±3.3	**27.4**±1.9	**31.5**±0.6	**52.0**±1.2	**67.4**±1.6	**16.8**±6.6	**48.5**±5.1	0±0

Table 13. Quantitative results of our method and baselines on the Area-5 of S3DIS dataset.

		OA(%)	mAcc(%)	mIoU(%)	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board.
Supervised Methods	PointNet [41]	77.5	59.1	44.6	85.2	97.4	72.3	0.1	10.6	54.9	18.5	48.4	39.5	12.4	55.5	40.2
	PointNet++ [42]	77.5	62.6	50.1	83.1	97.2	66.4	0	8.1	55.6	15.2	60.4	64.5	36.6	58.3	55.7
	SparseConv [12]	88.4	69.2	60.8	92.6	95.9	77.2	0.1	36.7	37.6	59.8	77.2	83.9	59.7	78.5	30.39
Unsupervised Methods	RandCNN	23.3±2.6	17.3±1.1	9.2±1.2	25.3±3.4	24.5±3.4	17.4±1.5	0±0	2.3±0.8	12.5±2.4	6.5±1.2	5.7±1.7	3.0±0.3	0.3±0.2	10.1±2.0	2.2±0.9
	van Kmeans	21.4±0.6	21.2±1.6	8.7±0.3	18.7±2.6	18.0±1.2	16.7±0.2	0.2±0	2.5±0.5	12.0±02	5.7±02.3	8.7±0.6	5.6±1.0	0±0	13.6±1.0	2.3±1.3
	van Kmeans-S	21.9±0.5	22.9±0.4	9.0±0.2	19.3±1.1	18.1±0.7	17.0±1.3	0.2±1.3	2.1±0.2	11.8±0	4.5±1.1	8.9±0.4	6.6±0.7	0.2±0.4	14.0±1.8	4.8±0.3
	van Kmeans-PFH	23.2±0.7	23.6±1.7	10.2±1.4	32.0±0.6	20.5±1.5	10.3±0.9	0.1±0.1	3.6±0.7	15.2±1.3	7.1±0.5	9.9±0.2	6.2±0.1	0.7±0.1	12.4±0.2	**4.9**±0.1
	van Kmeans-S-PFH	22.8±1.7	20.6±0.7	9.2±0.9	25.2±3.9	26.5±8.0	12.7±0.7	0.4±0.1	2.0±0.7	8.7±0.7	8.1±3.6	5.8±1.7	5.7±1.0	0±0	12.7±1.8	3.3±0.6
	IIC [24]	28.5±0.2	12.5±0.2	6.4±0	6.1±0.8	19.8±0.7	27.9±0.5	0±0	2.1±0.1	0.1±0.1	3.4±0.1	7.9±0.2	0.4±0.3	0±0	8.6±0.5	0±0
	IIC-S [24]	29.2±0.5	13.0±0.2	6.8±0	28.9±0.7	12.3±0.3	18.7±0.2	0±0	0.1±0	3.6±0.1	1.3±0.2	3.8±0.3	0.6±0	0±0	8.1±0.1	3.8±0
	IIC-PFH [24]	28.6±0.1	16.8±0.1	7.9±0.4	23.7±0.1	24.9±0.1	17.7±0.3	**5.9**±0.1	1.4±0.1	12.6±0.1	0.2±0	5.3±0.1	0.6±0	0±0	2.4±0.1	0±0
	IIC-S-PFH [24]	31.2±0.2	16.3±0.1	9.1±0.1	43.2±0.1	23.6±0.2	14.9±0.5	0±0	1.6±0	3.9±0.1	2.4±0.1	3.6±0	1.5±0	0.8±0	9.5±0.2	4.5±0.1
	PICIE [7]	61.6±1.5	25.8±1.6	17.9±0.9	65.7±7.4	61.4±12.3	58.4±0.4	0±0	0.3±0.4	2.2±2.7	1.7±1.2	12.1±8.8	0±0	0±0	12.4±2.0	0±0
	PICIE-S [7]	49.6±2.8	28.9±1.0	20.0±0.6	64.2±5.8	75.1±3.7	42.4±2.4	0.1±0	1.2±0.1	4.6±0.5	7.4±2.6	18.7±0.9	9.2±4.6	1.0±0.2	16.0±2.2	0.4±0.2
	PICIE-PFH [7]	54.0±0.8	36.8±1.7	24.4±0.6	58.4±5.0	68.6±4.5	49.9±2.0	0.1±0.2	7.6±0.7	5.3±0.8	14.2±1.8	45.1±1.2	16.3±1.0	0.1±0	27.1±5.6	0.6±0.5
	PICIE-S-PFH [7]	48.4±0.9	40.4±1.6	25.2±1.2	59.6±2.4	72.5±1.7	26.0±1.3	0.2±0	8.5±3.2	5.9±0.1	8.7±0.7	46.0±3.7	26.9±4.6	0.4±0.1	46.8±4.1	0.3±0
	GrowSP(Ours)	**78.4**±1.5	**57.2**±1.7	**44.5**±1.1	**90.45**±2.1	**90.1**±2.1	**66.7**±1.7	0±0	**14.8**±6.6	**27.6**±3.6	**45.6**±1.2	**59.4**±1.6	**71.9**±2.8	**10.7**±4.0	**56.0**±4.2	0.2±0.1

Table 14. Quantitative results of our method and baselines on the Area-6 of S3DIS dataset.

		OA(%)	mAcc(%)	mIoU(%)	ceil.	floor	wall	beam	col.	wind.	door	table	chair	sofa	book.	board.
Supervised Methods	PointNet [41]	79.0	79.6	60.9	85.7	96.5	71.8	59.4	47.4	67.4	74.3	56.2	48.9	20.9	50.0	52.5
	PointNet++ [42]	82.0	89.3	69.0	87.5	96.3	76.8	66.4	54.4	72.1	77.4	64.3	66.5	43.7	51.8	70.2
	SparseConv [12]	91.6	87.3	80.5	97.4	95.0	83.4	83.0	75.1	81.1	74.9	81.3	84.3	79.0	80.7	61.4
Unsupervised Methods	RandCNN	22.1±2.6	16.0±1.1	8.5±1.2	17.6±8.9	24.2±3.3	19.2±1.5	0±0	1.7±0.8	12.2±2.4	7.6±1.2	6.2±1.7	2.6±0.3	0.2±0.2	8.9±2.0	1.7±8
	van Kmeans	21.0±0.2	25.0±2.4	10.4±1.0	18.6±2.5	17.6±2.4	8.9±1.6	11.3±5.4	0.6±0.2	14.8±0.8	17.6±1.2	12.0±0.7	8.7±1.3	0.3±0.6	7.8±0.4	**6.2**±1.2
	van Kmeans-S	20.6±1.5	26.3±2.2	10.2±0.8	16.6±1.2	17.8±2.5	9.0±0.9	13.4±6.4	0.6±0.2	15.0±0.8	17.0±2.8	10.7±2.5	8.3±1.4	1.8±0.8	7.4±0.6	5.1±1.3
	van Kmeans-PFH	25.8±0.2	27.2±0.1	12.8±0.1	32.0±0.3	25.1±0.9	12.2±0	15.0±0.1	3.6±0.2	19.7±1.0	14.9±0.6	9.4±0.1	10.9±0	2.1±0	5.3±0	3.5±0.1
	van Kmeans-S-PFH	23.9±1.6	23.0±2.4	10.5±1.3	22.2±5.6	20.8±4.2	15.2±2.8	14.9±7.5	0.3±0.3	9.5±6.0	13.6±1.9	10.9±2.8	7.4±3.6	1.6±1.3	5.0±1.3	4.9±0.3
	IIC [24]	32.5±2.1	15.9±1.2	9.2±0.9	21.9±2.7	33.8±6.9	29.1±0.5	3.1±0.2	**15.2**±0.8	0±0	2.7±0.3	0.74±0.3	0±0	0±0	1.5±0.1	1.8±1.6
	IIC-S [24]	27.0±0.2	16.5±0.2	7.6±0.1	28.9±0.4	8.5±0.1	12.7±0.5	0.5±0.1	0.2±0	0±0	14.1±0.1	10.4±0.4	0.6±0	0±0	0.96±0.1	3.8±0
	IIC-PFH [24]	28.4±0.3	16.7±0.2	7.8±0.1	23.6±0.2	24.5±0.2	17.4±0.4	5.8±0.2	1.5±0	12.6±0.1	0.2±0	4.8±0.2	0.6±0.1	0±0	2.4±0.1	0±0
	IIC-S-PFH [24]	22.9±0.2	13.1±0	6.7±0.1	28.8±0.2	9.0±0.2	12.6±0.1	0.6±0	4.8±0	2.1±0.1	6.6±0.1	4.4±0.1	1.2±0.1	0.3±0	5.4±0.1	4.9±0.2
	PICIE [7]	39.3±4.8	28.5±0.2	17.8±1.5	56.9±1.9	61.7±17.5	18.6±0.5	20.5±1.9	4.2±1.2	6.0±0.4	8.7±0	14.7±1.7	15.9±0.2	1.1±1.8	5.7±0.9	0±0
	PICIE-S [7]	47.4±0.9	30.7±0.3	21.8±0.2	65.5±1.1	67.5±3.4	37.2±4.3	16.6±1.9	2.0±1.8	4.8±1.2	10.6±3.2	23.7±1.5	23.4±1.2	23.4±1.5	1.7±0.9	0±0
	PICIE-PFH [7]	51.8±2.6	41.3±1.0	27.9±0.9	63.1±14.1	56.5±6.5	39.0±4.0	11.4±3.5	10.0±0.9	5.3±0.4	19.1±3.0	63.8±0.6	50.4±0.5	0.5±0.9	37.2±4.3	0.9±1.5
	PICIE-S-PFH [7]	44.0±0.5	36.1±0.9	24.7±0.4	58.2±2.5	60.4±1.2	24.5±1.0	17.9±3.0	10.1±2.8	8.1±1.1	12.7±2.6	44.0±0.2	5.5±0.1	0.5±0.9	54.3±1.3	0.3±0.6
	GrowSP(Ours)	**75.6**±0.8	**58.5**±1.3	**47.6**±0.9	**89.4**±0.3	**88**±3.3	**57.7**±0.5	**70.6**±0.7	2.0±2.2	**32.4**±2.5	**36.7**±1.5	**63.2**±0.9	**69.8**±0.8	1.5±1.4	**58.9**±11.8	0.2±0.3

D. Point Feature Histograms Descriptors

In our semantic primitive clustering module, we adopt PFH feature [48] which explicitly measures the surface normal distributions to augment neural features for better semantic primitive clustering.

Given a center point with its neighbouring points, PFH iteratively selects any two of them and computes a set of angles from point normals. In our implementation, we adopt a simplified version to save computation. Specifically, for a superpoint that contains a set of points, we compute the cosine distances between normals of any two points in this suerpoint, and compute the distribution of cosine distances to form a histogram in the range $[-1, 1]$ with an interval 0.2. In this way, we can get a 10-dimensional vector to describe the normal distribution of a superpoint, and regard it an additional features for semantic primitives clustering.

E. Evaluation on S3DIS

In S3DIS, we use the same hyper-parameters for all 6 Areas. Following SparseConv [12], we firstly use a grid size of 0.01m to sub-grid downsample input point clouds. We then use a 5cm voxel size to voxelize each point cloud. The cross-entropy loss is minimized by SGD optimizer with a batch size of 10, a momentum of 0.9, an initial learning rate of 0.1. The learning rate is decreased by Poly scheduler, and we train 1300 epochs for each area of S3DIS. During training, we cluster 300 semantic primitives for all epochs. The semantic primitive clustering, superpoint growing, and pseudo labels updating are conducted every 10 epochs.

Because the category *clutter* does not have consistent geometry, we exclude it during training. We never use these points to compute losses or apply K-means. Tabs. 9 to 14 present the per-category results on Area-1/2/3/4/5/6. Our **GrowSP** consistently outperforms all baselines and achieve comparable performance with the fully-superviesed PointNet [41]. More qualitative results can be seen in Figure 9.

F. Evaluation on ScanNet

In ScanNet, the hyper-parameters are the same as in S3DIS. We also use a 5cm voxel size to voxelize each point cloud. The cross-entropy loss is minimized by SGD optimizer with a batch size of 8, a momentum of 0.9, and an initial learning rate of 0.1. The learning rate is decreased by Poly scheduler. We train a total of 800 epochs. Since ScanNet has an *undefined* category which does not have consistent geometry, we exclude(masked) these points during training. We feed these points into the network but neither compute their losses nor apply K-means to them.

The per-category results on both validation and hidden test sets are presented in Tables 15 & 16. Our method outperforms all baselines and achieve comparable results with

Table 15. Per-category quantitative results on the validation split of ScanNet dataset.

	OA(%)	mAcc(%)	mIoU(%)	wall.	floor.	cab.	bed.	chair.	sofa.	table	door.	wind.	books.	pic.	counter.	desk.	curtain.	fridge.	shower.	toilet.	sink.	bathtub.	otherf.
RandCNN	11.9±0.4	8.4±0.1	3.2±0	9.3±0.4	10.0±0.5	3.5±0.5	2.5±0.7	6.8±0.9	2.0±0.4	4.8±0.7	5.1±0.3	3.9±0.2	3.1±1.0	1.7±0.5	0.8±0.3	2.3±0.4	2.8±0.3	1.0±0.1	0.2±0.2	0.2±0.1	0.1±0.1	0.6±0.1	3.6±0.5
van Kmeans	10.1±0.1	10.0±0.1	3.4±0	9.0±0.4	9.8±1.1	3.2±0.6	2.9±0.2	5.5±0.4	3.3±0.1	4.3±0.2	3.5±0.5	5.5±0.4	3.3±0.5	**2.6**±0.4	0.8±0.5	2.9±0.4	4.3±0.7	0.8±0.2	0.8±0.4	0.7±0	0.3±0.2	0.9±0.2	4.0±0.4
van Kmeans-S	10.2±0.1	9.8±0.3	3.4±0.1	8.9±0.5	10.3±0.6	3.4±0.3	3.2±0.3	5.5±0.2	3.4±0.2	4.2±0.6	3.4±0.1	5.2±1.4	3.1±0.7	2.6±0.4	0.7±0.2	2.8±0.4	4.2±0.3	0.6±0.2	0.8±0.5	0.7±0	0.2±0.1	1.0±0.2	4.1±0.3
van Kmeans-PFH	10.4±0.2	10.3±0.7	3.5±0.2	8.6±0.6	12.7±0.1	2.9±0.2	2.8±0.1	4.5±0.1	3.2±0.1	3.6±0.2	3.7±0.3	6.3±0.1	4.0±0.3	2.4±0.4	1.0±0.1	2.9±0.1	3.2±0.8	1.0±0.3	1.0±0.4	0.6±0.2	0.4±0.1	1.1±0.7	3.5±0.5
van Kmeans-S-PFH	12.2±0.6	9.3±0.5	3.6±0.1	11.3±0.4	12.3±1.4	2.9±1.0	2.4±0.6	5.4±0.8	2.8±0.4	4.2±0.8	3.8±0.5	5.8±0.7	3.8±0.6	2.3±0.7	1.2±0.3	2.4±0.4	2.9±1.7	0.9±0.4	1.4±0.7	0.6±0.1	0.1±0.1	1.1±0.3	4.1±0.7
IIC [24]	27.7±2.7	6.1±1.2	2.9±0.8	25.3±3.9	20.5±2.6	0.6±1.0	0.3±0.4	3.7±4.7	0.4±0.6	1.3±1.6	1.3±1.4	1.1±1.5	1.9±2.6	0.2±0.1	0.1±0.2	0.6±0.8	0.3±0.4	0.4±0.6	0±0	0±0	0±0	0.2±0.3	0.5±0.6
IIC-S [24]	18.3±2.6	6.7±0.6	3.4±0.1	18.3±2.6	16.0±1.6	2.6±0.9	2.3±0.4	4.4±1.2	2.0±0.3	5.4±2.0	3.2±1.6	2.9±0.8	3.3±1.4	0.7±0.1	0.4±0.2	1.4±0.6	1.6±0.7	0.7±0.2	0.1±0.2	0.3±0.2	0.1±0.1	0±0	2.6±0.8
IIC-PFH [24]	25.4±0.1	6.3±0	3.4±0	29.6±0.2	14.9±0.2	1.1±0.1	1.0±0	5.6±0	0.8±0.1	3.6±0	3.0±0	1.6±0	1.3±0.1	0±0	0.3±0.1	1.0±0	0.4±0	0.4±0	0.2±0	0±0	0.1±0	0±0	3.2±0.1
IIC-S-PFH [24]	18.9±0.3	6.3±0.2	3.0±0.1	18.0±0.1	15.9±0.3	3.4±0.2	0.9±0.1	7.1±1.4	0.6±0	0.8±0.2	4.3±0	1.6±1.6	3.5±0.1	0.4±0	0.1±0	0.3±0.1	0.3±0	0±0	0±0	0.1±0	0±0	0±0	2.7±1.6
PICIE [7]	20.4±0.5	16.5±0.3	7.6±0	14.7±0.5	24.5±1.8	6.3±0.2	5.2±1.8	18.0±3.3	8.4±1.3	33.2±1.2	6.7±0.5	4.8±0.3	9.3±4.0	2.1±0.7	0.1±0.1	2.7±1.1	8.0±1.3	1.1±0.2	2.1±1.8	0±0	0±0	0.5±0.5	5.0±0.3
PICIE-S [7]	35.6±1.1	13.7±1.5	8.1±0.5	38.4±0.8	53.9±0.8	4.3±0.3	2.7±0.3	10.2±1.0	6.3±0.8	14.1±1.4	5.2±0.6	4.0±0	6.0±0.2	0.2±0.3	1.3±0	2.1±0.7	1.5±0	0.2±0.2	0±0	2.6±0.1	3.1±0.7	1.3±0	4.3±0.2
PICIE-PFH [7]	33.1±1.4	14.0±0.1	8.1±0.3	34.7±1.1	54.8±3.0	3.9±0.3	5.4±1.9	13.3±4.1	6.5±1.6	11.7±2.4	4.2±0.4	3.8±0.2	6.5±1.1	0.5±0.1	1.0±0.6	2.6±1.5	5.0±0.4	**1.3**±0.3	1.0±0.9	0.6±0.8	0±0	0.7±0.1	4.4±0.6
PICIE-S-PFH [7]	23.6±0.4	15.1±0.6	7.4±0.2	18.1±0.8	39.1±1.5	5.4±0.2	4.9±0.4	13.4±0.9	6.9±0.4	20.3±5.8	5.8±0.1	4.5±0.3	7.7±0.5	1.2±1.0	3.0±1.9	5.8±0.7	4.7±0.8	0.6±0.5	1.2±1.0	0.4±0.3	0±0	1.1±0.3	4.5±0.2
GrowSP(Ours)	**57.3**±2.3	**44.2**±3.1	**25.4**±2.3	**40.7**±2.0	**89.8**±0.4	**24.0**±5.8	**47.2**±2.0	**45.5**±19.0	**43.0**±1.4	**39.4**±3.4	**14.1**±0.5	**20.0**±0.3	**53.5**±6.6	0.1±0.1	**5.4**±9.5	**13.3**±0.5	**8.4**±0.8	**2.1**±0.6	**11.3**±1.2	**20.6**±18.2	**19.4**±1.2	0±0	**9.8**±2.7

Table 16. Per-category quantitative results on the hidden test split of ScanNet dataset.

		mIoU(%)	wall.	floor.	cab.	bed.	chair.	sofa.	table	door.	wind.	books.	pic.	counter.	desk.	curtain.	fridge.	shower.	toilet.	sink.	bathtub.	otherf.
Supervised Methods	PointNet++ [42]	33.9	52.3	67.7	25.6	47.8	36	34.6	23.2	26.1	25.2	45.8	11.7	25.0	27.8	24.7	18.3	14.5	54.8	36.4	58.4	18.3
	DGCNN [61]	44.6	72.3	93.7	36.6	62.3	65.1	57.7	44.5	33.0	39.4	46.3	12.6	31.0	34.9	38.9	28.5	22.4	62.5	35.0	47.4	27.1
	PointCNN [27]	45.8	70.9	94.4	32.1	61.1	71.5	54.5	45.6	31.9	47.5	35.6	16.4	29.9	32.8	37.6	21.6	22.9	75.5	48.4	57.7	28.5
	SparseConv [12]	72.5	86.5	95.5	72.1	82.1	86.9	82.3	62.8	61.4	68.3	84.6	32.5	53.3	60.3	75.4	71.0	87.0	93.4	72.4	64.7	57.2
Unsupervised Methods	**GrowSP(Ours)**	26.9	32.8	89.6	15.2	62.9	55.3	38.9	32.0	14.4	23.0	59.9	0	12.5	11.4	6.1	1.2	9.3	43.9	14.0	0	16.5

PointNet++ [42]. Figure 9 shows more qualitative results.

G. Generalization to Unseen Datasets

In this section, we further evaluate whether the learned features of our unsupervised method are indeed general to unseen scenes. In particular, having the well-trained models of both S3DIS and ScanNet in Sections 4.1&4.2, we conduct the following two groups of experiments.

- *Group 1: Generalization from ScanNet to S3DIS*. We directly use the well-trained model of ScanNet to test on 6 areas of S3DIS. The final classifier uses the 20 centroids estimated by K-means on the training split of ScanNet. This means that, the neural network is completely unaware of any information of S3DIS dataset.
- *Group 2: Generalization from S3DIS to ScanNet*. We directly use the well-trained 6 models of S3DIS to test on the validation split of ScanNet.

Analysis: Table 17 shows results of Group 1. We can see that our method trained on ScanNet dataset has achieved superior generalization capability to the unseen S3DIS dataset, with mIoU scores about 30% in several areas, which are clearly better than all baselines. Notably, these scores are slightly higher than the results in Tables 3&4. We speculate it is because ScanNet dataset itself consists of very diverse geometries and semantic elements which are well discovered by our model, whereas S3DIS dataset consists of fewer semantic elements and is relatively easier. Table 18 shows results of Group 2. It can be seen that the generalization performance of our method from S3DIS to ScanNet is clearly better than baselines, although the overall capability is weaker than that of Group 1 because of the relatively simple rooms in S3DIS dataset for training.

Table 17. Generalization ability of models trained on ScanNet [10] to the unseen 6 areas of S3DIS [2]. The mIoU scores with standard deviations of 12 categories are reported.

test on →	Area-1	Area-2	Area-3	Area-4	Area-5	Area-6
IIC [24]	3.7±0.5	3.8±0.4	3.8±0.2	4.0±0.5	3.8±0.2	3.7±0.4
IIC-S [24]	6.7±0.1	5.7±0	6.4±0.2	5.8±0	5.9±0	6.5±0.1
PICIE [7]	13.5±0.1	12.7±0.2	13.4±0.1	12.8±0.1	11.3±0.4	13.1±0.1
PICIE-S [7]	14.7±0.9	13.9±0.8	15.1±0.7	14.7±0.4	14.2±0.3	15.8±0.2
GrowSP (Ours)	**24.2**±1.9	**21.9**±1.7	**26.1**±2.8	**25.0**±2.8	**23.7**±2.3	**27.9**±2.5

Table 18. Generalization ability of models trained on different areas of S3DIS [2] to the unseen val split of ScanNet [10]. The mIoU scores with standard deviations of 20 categories are reported.

model trained on →	Areas 2/3/4/5/6	Areas 1/3/4/5/6	Areas 1/2/4/5/6	Areas 1/2/3/5/6	Areas 1/2/3/4/6	Areas 1/2/3/4/5
IIC [24]	3.5±0	3.4±0	3.7±0.1	3.5±0.1	3.5±0	3.6±0
IIC-S [24]	3.9±0.1	3.9±0.1	4.0±0.1	3.9±0	3.9±0.1	3.9±0
PICIE [7]	5.6±0.2	5.1±0.1	5.0±0.1	5.9±0.3	6.0±0.3	5.5±0.2
PICIE-S [7]	6.9±0.3	6.9±0.7	6.9±0.8	8.1±0.4	8.4±0.3	6.7±0.9
GrowSP (Ours)	**16.9**±0.6	**17.8**±0.6	**16.4**±0.5	**16.1**±0.6	**17.1**±0.8	**15.3**±0.3

H. Evaluation on SemanticKITTI

Considering that outdoor point clouds are usually dominated by *road* and the point density is significantly different from that of indoor datasets, we opt for (RANSAC + Euclidean Clustering) as an alternative to (VCCS + region growing) to construct initial superpoints,

RANSAC: In our experiment, we choose to fit a plane by RANSAC and take points with a distance smaller than 0.2m as a huge superpoint.

Euclidean Clustering: After fitting the largest plane which normally corresponds to *road*, we construct initial superpoints for remaining points by Euclidean clustering. Specifically, if the Euclidean distance of two points is smaller than 0.2m, they are assigned into the same superpoint; otherwise not.

Training/Testing Details: In training, we voxelize point clouds by a grid size of 15cm without any other preprocessing. Following Mix3D [37], we use an AdamW optimizer with a batchsize of 16 to train 400 epochs. The

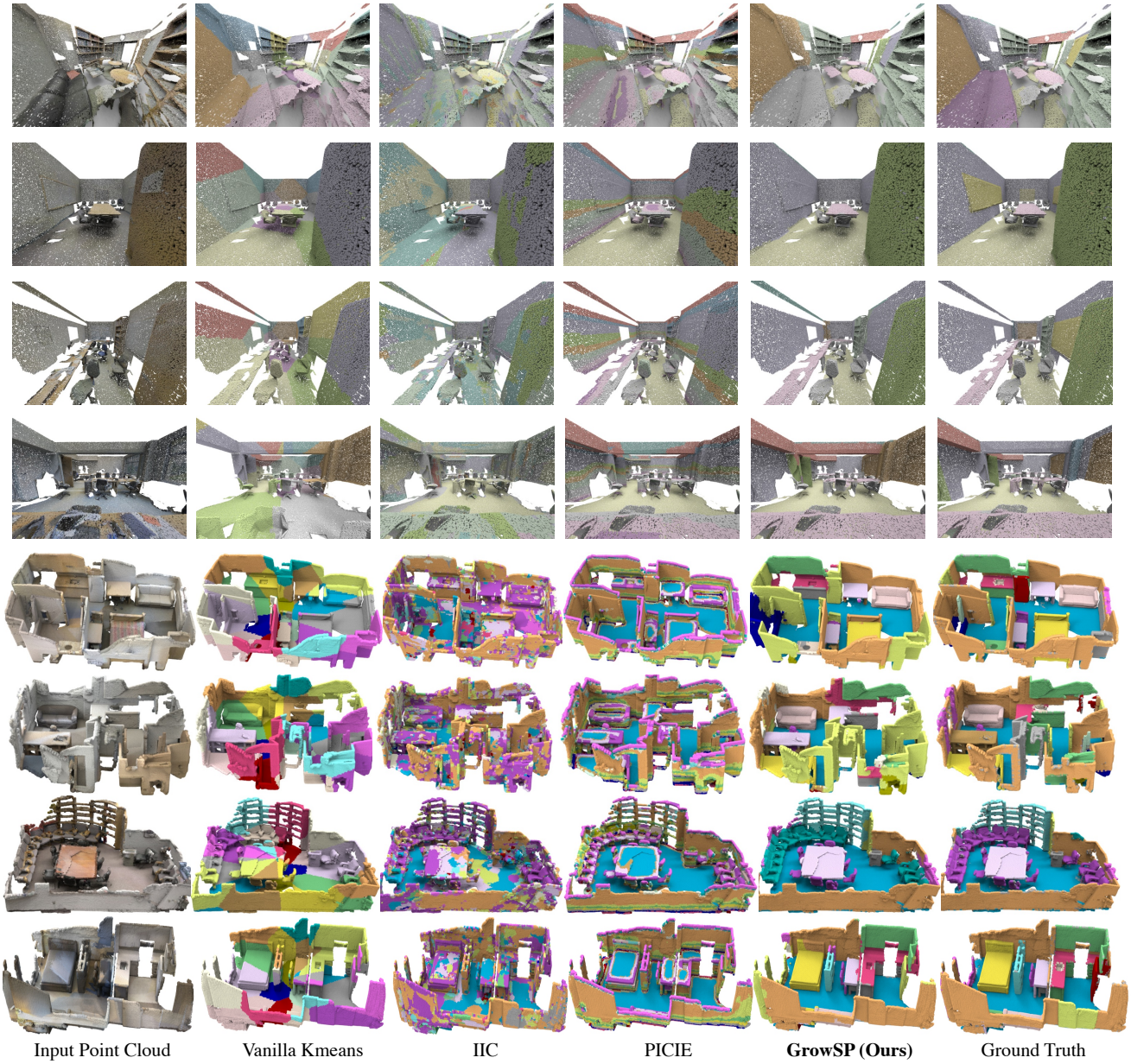

Figure 9. The top four rows show qualitative results for S3DIS, the bottom four rows for ScanNet.

learning rate is decreased by OneCycleLR with a max learning rate of 0.01. The number of primitive S is set as 500, M^1 as 80, and M^T as 30. The semantic primitive clustering, superpoint growing, and pseudo labels updating are conducted every 10 epochs. Since SemanticKITTI has nearly 20000 training scans and each covers an extremely spacious range, we only randomly select 1500 scenes in each round (10 epochs) and crop each point cloud by a 50m radius sphere to train the network for efficiency. Similarly, the *undefined background* points do not have consistent geometry, so we exclude(masked) these points during training. We feed these points into the network but neither compute their losses nor apply K-means on them. During testing, all raw points are fed into the network for evaluation.

Tables 19&20 show the per-category results on both validation and hidden test sets, and our methods achieve SOTA compared with other unsupervised approaches. Figure 10 shows qualitative results.

More qualitative results and video demo can be found at https://github.com/vLAR-group/GrowSP

Table 19. Per-category quantitative results on the **offline validation split** of SemanticKITTI dataset.

	OA(%)	mAcc(%)	mIoU(%)	car.	bike.	mbike.	truck.	vehicle.	person.	cyclist.	mcyclist.	road.	parking.	sidewalk.	other-gr.	building.	fence.	veget.	trunk.	terrain.	pole.	sign.
RandCNN	25.4±3.3	6.0±0.2	3.3±0.1	2.5±0.4	0±0	0±0	0±0	0.2±0.1	0±0	0±0	0±0	8.5±2.1	0.8±0.5	4.9±1.8	0.3±0.3	6.2±1.3	1.3±0.3	29.0±3.1	1.0±0.2	8.1±1.6	0.4±0.1	0.1±0
van Kmeans	8.1±0	8.2±0.1	2.4±0	5.6±0.2	0.1±0	0.1±0	0.2±0	0.5±0.1	0.1±0	0±0	0±0	12.3±0.1	1.1±0.1	4.4±0.1	0.3±0	5.8±0.2	2.0±0	5.7±0.1	1.4±0	5.0±0.1	0.5±0	0.1±0
van Kmeans-S	10.3±0.3	7.7±0.1	2.6±0	5.6±0.4	0.1±0.1	0.1±0.1	0.1±0.1	0.3±0	0.1±0	0±0	0±0	13.5±0.6	1.0±0.4	5.0±0.2	0.3±0	7.1±0.6	1.5±0.2	7.5±0.7	1.5±0.1	6.0±0.1	0.4±0.1	0.1±0
van Kmeans-PFH	11.2±0.6	7.5±0.7	2.7±0.1	4.5±0.8	0.1±0	0.1±0.1	0.2±0.1	0.1±0.1	0.2±0	0.2±0.2	0±0	9.1±0.8	1.6±0.6	4.9±0.4	0.2±0.1	8.2±0.8	1.6±0.2	9.6±0.5	1.4±0	7.5±0.7	0.3±0	0.3±0.2
van Kmeans-S-PFH	13.2±1.8	8.1±0.4	3.0±0.2	4.8±0.5	0.1±0.1	0.1±0.1	0.3±0.2	0.5±0.2	0.1±0	0.2±0.2	0±0	11.3±2.5	1.5±0.5	5.3±0.8	0.3±0	8.4±0.7	1.5±0.2	11.3±3.2	1.5±0.1	8.4±0.5	0.4±0.1	0.3±0.1
IIC [24]	26.2±1.5	5.8±0.4	3.1±0.3	1.6±0.9	0±0	0±0	0±0	0±0	0±0	0±0	0±0	8.9±2.0	0.1±0.1	2.6±1.8	0±0	7.1±4.2	0.2±0.1	26.5±2.5	0.3±0.4	11.5±1.5	0.1±0.1	0.1±0.1
IIC-S [24]	23.9±1.1	6.1±0.3	3.2±0.2	1.6±0.8	0±0	0±0	0.1±0.1	0.1±0.1	0±0	0.1±0.1	**9.7**±1.9	0.6±0.5	**4.3**±2.8	0.1±0.1	**8.8**±3.2	0.5±0.6	**24.3**±2.3	0.6±0.5	9.7±2.6	0.3±0.3	0.1±0.1	0±0.1
IIC-PFH [24]	20.1±0.1	7.2±0.1	3.6±0	5.8±0	0.1±0.1	0.2±0	0.2±0	0.5±0	0.2±0	0.1±0	0±0	14.5±0.2	1.1±0.3	6.6±0.2	0.1±0	6.8±0.1	1.6±0.2	19.7±0	2.1±0	8.4±0.1	0.6±0.1	0.1±0
IIC-S-PFH [24]	23.4±0	9.0±0	4.6±0	10.0±0.1	0.1±0	0±0	0.3±0	0.4±0	**0.3**±0	**0.3**±0	0±0	21.7±0.2	2.4±0	10.0±0.1	0±0	8.7±0	1.6±0	19.7±0.2	1.1±0	9.7±0.1	0.4±0	0.2±0
PICIE [7]	22.3±0.4	14.6±0.3	5.9±0.1	**7.4**±0.2	0.3±0.2	0±0	0.1±0	0.6±0.1	0.3±0.1	0.1±0.1	0±0	26.5±0.3	1.6±0.1	14.8±1.4	0.6±0.3	20.5±0.4	4.8±0.1	16.3±1.0	2.1±0.9	14.2±0.9	1.4±0.3	0.4±0.2
PICIE-S [7]	18.4±0.5	13.2±0.2	5.1±0.1	6.1±1.4	0.1±0	0±0	0.1±0.1	0.4±0.1	0.3±0.1	0.1±0.1	0±0	21.3±1.4	1.7±0.1	12.9±2.3	0.4±0.2	21.2±0.9	2.6±0.3	13.4±0.4	2.4±0.3	11.5±2.9	**2.6**±0.2	0.4±0
PICIE-PFH [7]	46.6±0.2	10.1±0	4.7±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	39.7±0.8	0±0	0±0	0±0	0±0	0±0	50.2±0.3	0±0	0±0	0±0	0±0
PICIE-S-PFH [7]	42.7±2.1	11.5±0.2	6.8±0.6	4.8±2.6	0±0	0±0	0±0	0±0	0±0	0±0	0±0	32.0±6.3	0.6±1.0	12.5±8.7	0±0	25.5±0.6	0.8±1.0	43.6±1.2	0.5±0.4	9.2±9.3	0±0	0±0
GrowSP(Ours)	**38.3**±1.0	**19.7**±0.6	**13.2**±0.1	**76.0**±0.4	0±0	**0.4**±0.2	**0.9**±0.7	**1.0**±0.1	0.1±0.2	0.1±0.2	0±0	**26.8**±3.5	1.0±0.4	**13.8**±4.5	0.4±0.3	**39.2**±2.1	1.3±0.4	**26.7**±1.5	**25.1**±0.7	**35.5**±1.9	0.2±0.1	**2.1**±0.1

Table 20. Per-category quantitative results on the **hidden test split** of SemanticKITTI dataset.

		mIoU(%)	car.	bike.	mbike.	truck.	vehicle.	person.	cyclist.	mcyclist.	road.	parking.	sidewalk.	other-gr.	building.	fence.	veget.	trunk.	terrain.	pole.	sign.
Supervised Methods	PointNet [41]	14.6	46.3	1.3	0.3	4.6	0.8	0.2	0.2	0	61.6	15.8	35.7	1.4	41.4	12.9	31.0	4.6	17.6	2.4	3.7
	PointNet++ [42]	20.1	53.7	0.9	0.2	0.9	0.2	0.9	1.0	0	72.0	18.7	41.8	5.6	62.3	16.9	46.5	13.8	30.0	6.0	8.9
	SparseConv [8]	53.2	94.0	26.4	24.5	27.5	18.4	40.5	46.7	13.5	88.4	57.1	71.4	22.6	90.4	62.5	83.5	65.3	65.8	54.0	59.1
Unsupervised Methods	**GrowSP(Ours)**	14.3	81.9	0.1	0.5	0.2	1.0	0.3	0	0	25.0	0	17.4	0.5	64.6	1.4	29.4	26.6	22.4	0.3	0.5

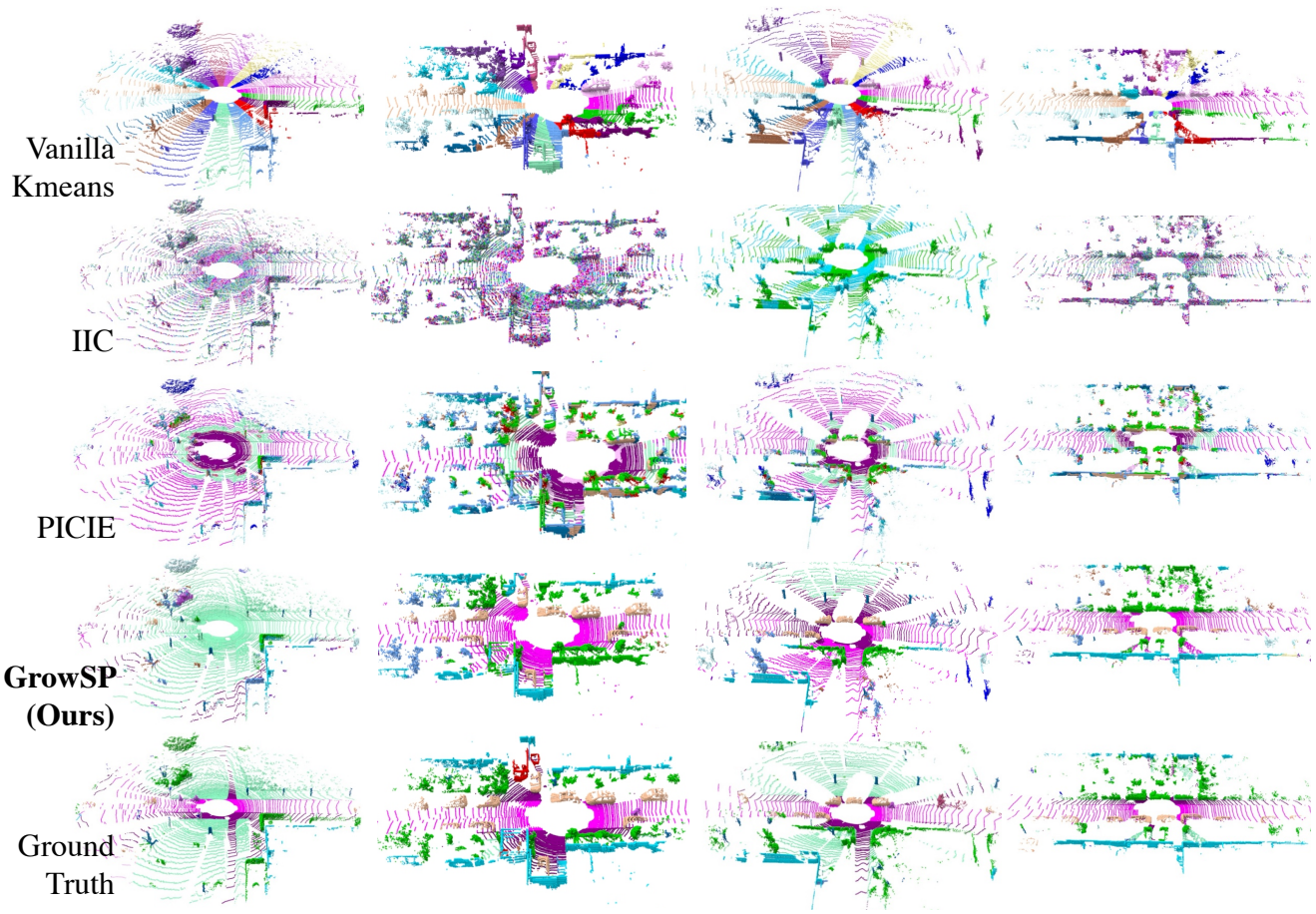

Figure 10. Qualitative results on SemanticKITTI dataset.