Visual Learning And Reasoning

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

GrowSP: Unsupervised Semantic Segmentation of 3D Point Clouds

Zihui Zhang, Bo Yang, Bing Wang, Bo Li

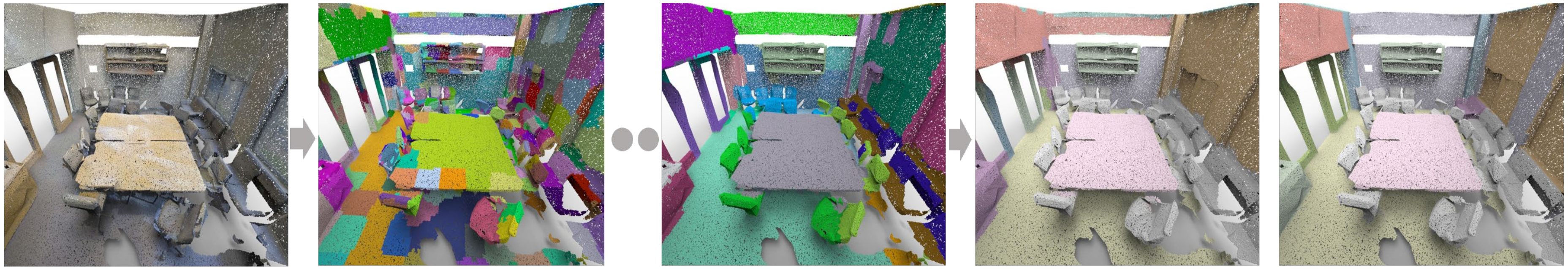
vLAR Group, The Hong Kong Polytechnic University

Paper Tag: THU-AM-109

Brief Introduction

Motivations: Annotation is costly

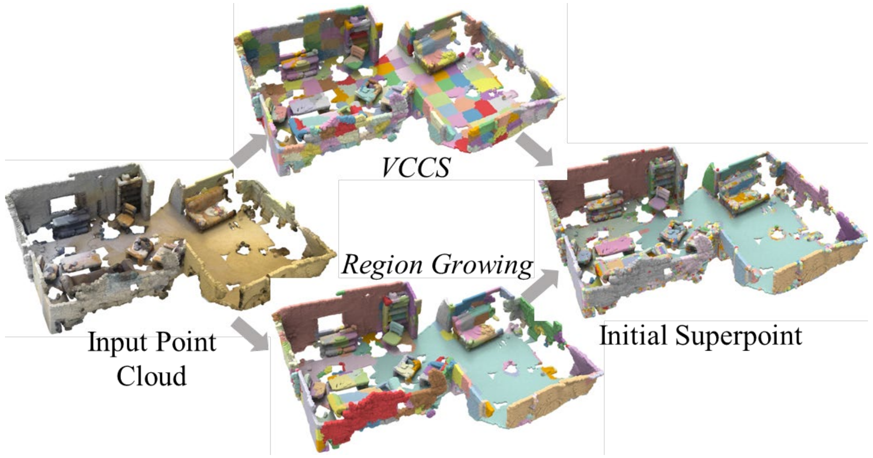
Overall Pipeline:



Input Point Cloud | Initial Superpoints | **Progressively Growing Superpoints** | Prediction by **GrowSP** | Ground Truth

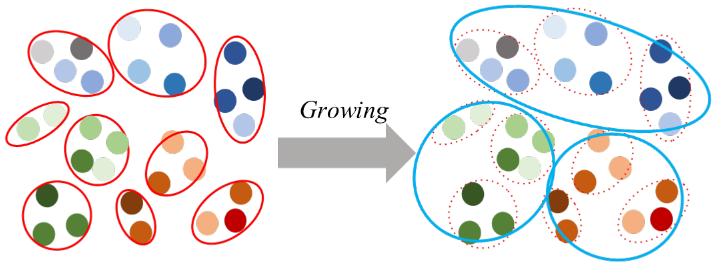
Superpoint Constructor → **Superpoint Growing** → **Semantic Primitives**

① Initial Superpoint Constructor:



② Superpoint Growing:

Merge similar superpoints.



Initial Superpoints | Larger Superpoints

③ Semantic Primitives:

Cluster superpoints to primitives

$$primitives \xleftarrow{\textbf{Kmeans}} (f_1^1 \cdots f_m^H)$$

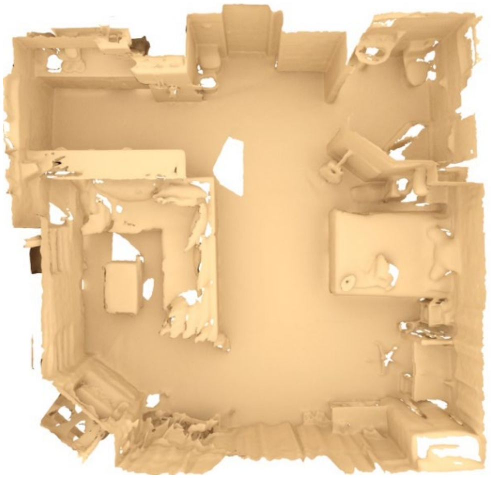
④ Semantic Classes:

Cluster primitives to categories

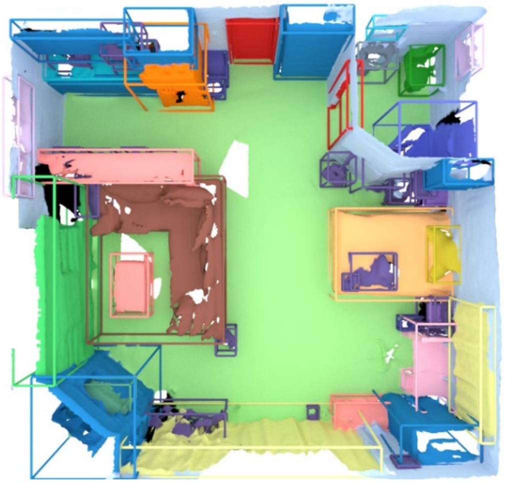
$$categories \xleftarrow{\textbf{Kmeans}} (pri_1 \cdots pri_S)$$

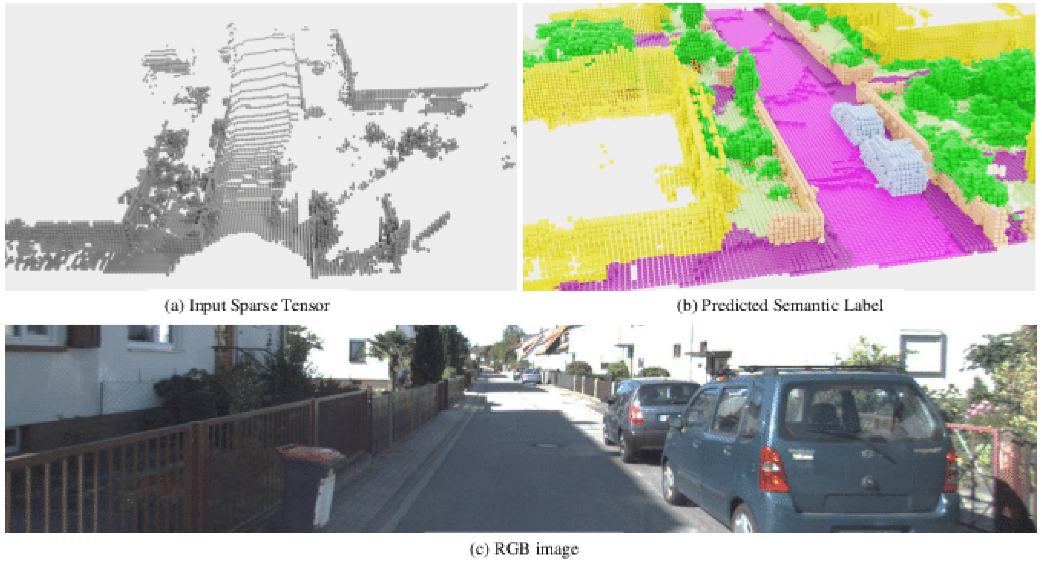
Background

Assign semantic labels to each point



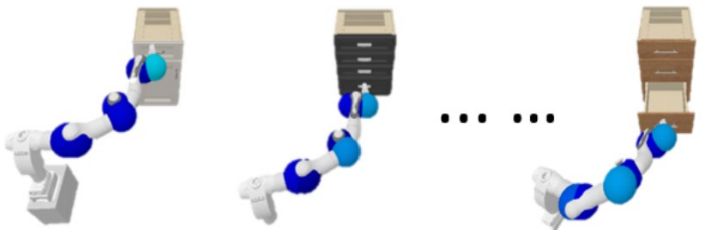
Semantic Segmentation





Autonomous Driving

VR



Robotic

3D scene semantic segmentation is crucial for practical application

Motivation

- **Fully supervised：**

ScanNet Benchmark

Method	Info	avg iou	bathtub	bed	bookshelf	cabinet	chair	counter	curtain
Mix3D	P	**0.781** 1	**0.964** 1	**0.855** 1	0.843 10	**0.781** 1	0.858 7	0.575 2	0.831 17
Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, Francis Engelmann: Mix3D: Out-of-Context Data Augmentation for 3D Scenes. 3DV 2021 (Oral)									
OccuSeg+Semantic		0.764 2	0.758 42	0.796 16	0.839 11	0.746 8	**0.907** 1	0.562 3	0.850 12
O-CNN	P	0.762 3	0.924 2	0.823 4	0.844 9	0.770 2	0.852 9	**0.577** 1	0.847 13
Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, Xin Tong: O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis. SIGGRAPH									
DMF-Net		0.752 4	0.906 4	0.793 19	0.802 25	0.689 22	0.825 25	0.556 4	0.867 8
PointTransformerV2		0.752 4	0.742 49	0.809 11	**0.872** 1	0.758 4	0.860 6	0.552 5	0.891 5

- **Weakly supervised：**

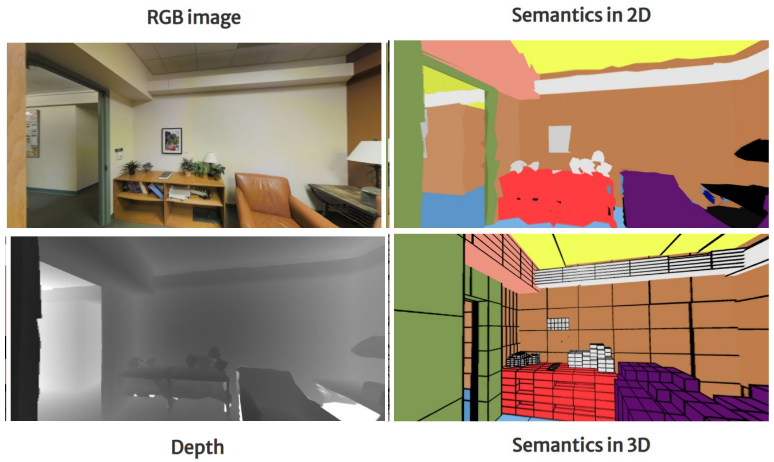
ScanNet Data Efficient   Limited Annotations ▾   Limited Rec

Trained points per scene: 20 points ▾

Method	Info	avg iou	bathtub	bed	bookshelf	cabinet	chair	counter	curtain	de
ActiveST		**0.703** 1	**0.977** 1	0.776 2	0.657 4	**0.707** 1	**0.874** 1	**0.541** 1	**0.744** 1	**0.60**
Gengxin Liu, Oliver van Kaick, Hui Huang, Ruizhen Hu: Active Self-Training for Weakly Supervised 3D Scene Semantic Segmentation.										
WeakLab-3D-Net(WS3D)	P	0.662 2	0.812 3	0.762 3	**0.742** 1	0.635 2	0.828 5	0.474 2	0.736 2	0.58
DE-3DLearner LA		0.639 3	0.839 2	0.723 5	0.681 2	0.629 3	0.839 4	0.424 3	0.728 3	0.53

Manually annotating real-word 3D point cloud is costly

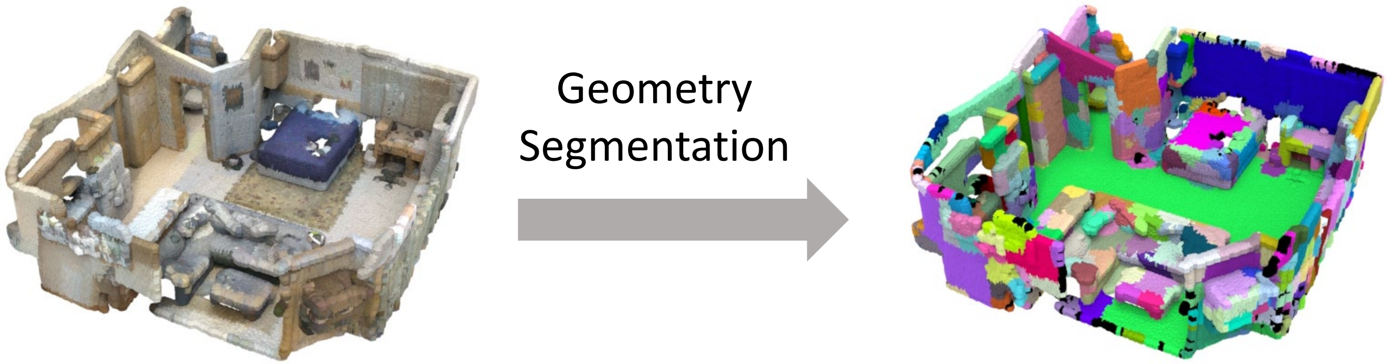
Semanic labeling on 3D points:



Semanic labeling on RGBD

GrowSP:

- Semantic Significance of Point Regions vs Individual Points

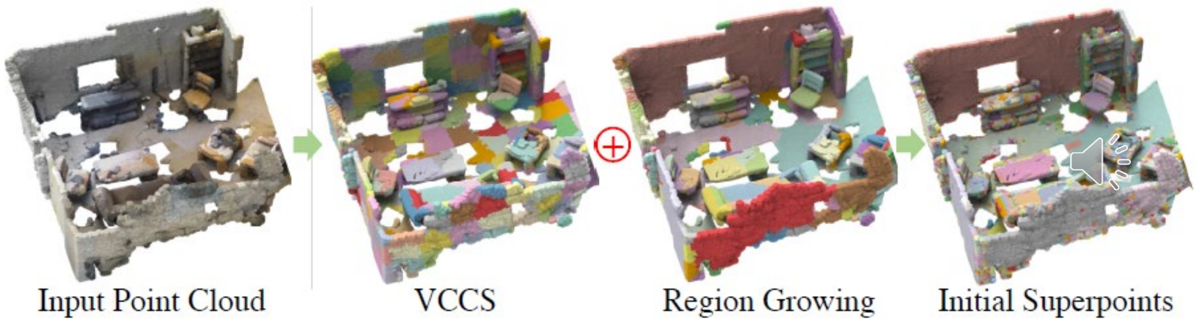


Input Point Cloud　　　　Initial Superpoints

Superpoints Constructors:

Region Growing: normal similarity and connectivity

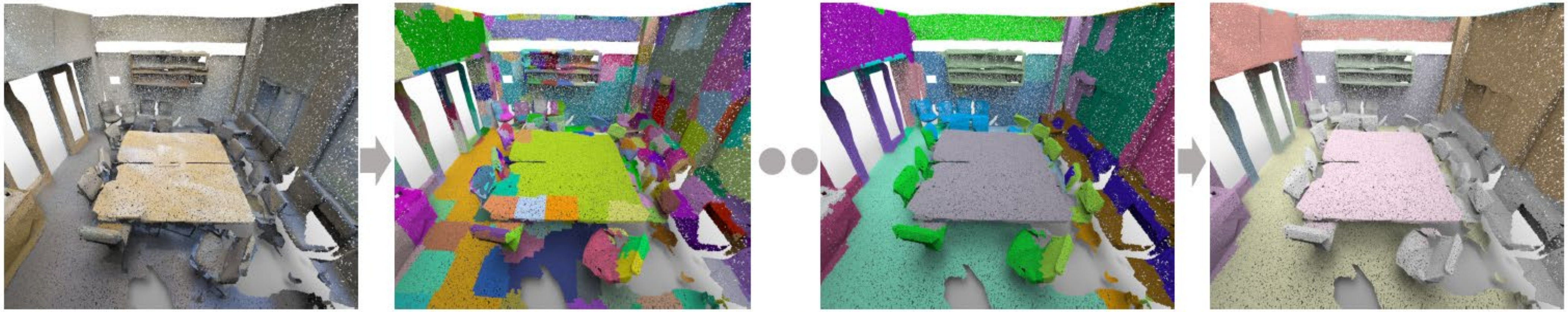
VCCS: $$\boldsymbol{D} = \sqrt{w_c D_c^2 + \frac{w_s D_s}{3R_{seed}^2} + w_n D_n}$$



GrowSP:

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

- Growing Superpoints to contain more semantics:

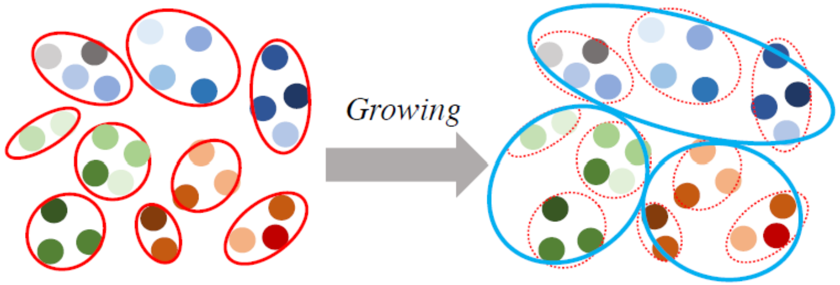


Input Point Cloud | Initial Superpoints | **Progressively Growing Superpoints** | Prediction by **GrowSP**

In feature space:

$$\tilde{\boldsymbol{f}}^h_{m^0} = \frac{1}{Q}\sum_{q=1}^{Q} \boldsymbol{f}^h_q,$$

$$\{\tilde{\boldsymbol{p}}^h_1 \cdots \tilde{\boldsymbol{p}}^h_{m^1} \cdots \tilde{\boldsymbol{p}}^h_{M^1}\} \xleftarrow{\text{Kmeans}} \{\tilde{\boldsymbol{f}}^h_1 \cdots \tilde{\boldsymbol{f}}^h_{m^0} \cdots \tilde{\boldsymbol{f}}^h_{M^0}\}$$



Initial Superpoints | Larger Superpoints

GrowSP:

- Semantic Primitives(sub-class) and Auxiliary features:

1). Cluster superpoints into semantic categories is aggressive.

- We choose to constantly group superpoints into semantic primitives rather than semantic categories.

$$S\ primitives \xleftarrow{\text{Kmeans}} (\{\hat{\boldsymbol{f}}_1^1 \cdots \hat{\boldsymbol{f}}_{m^0}^1 \cdots\} \cdots \{\hat{\boldsymbol{f}}_1^H \cdots \hat{\boldsymbol{f}}_{m^0}^H \cdots\})$$

2). Auxiliary features.

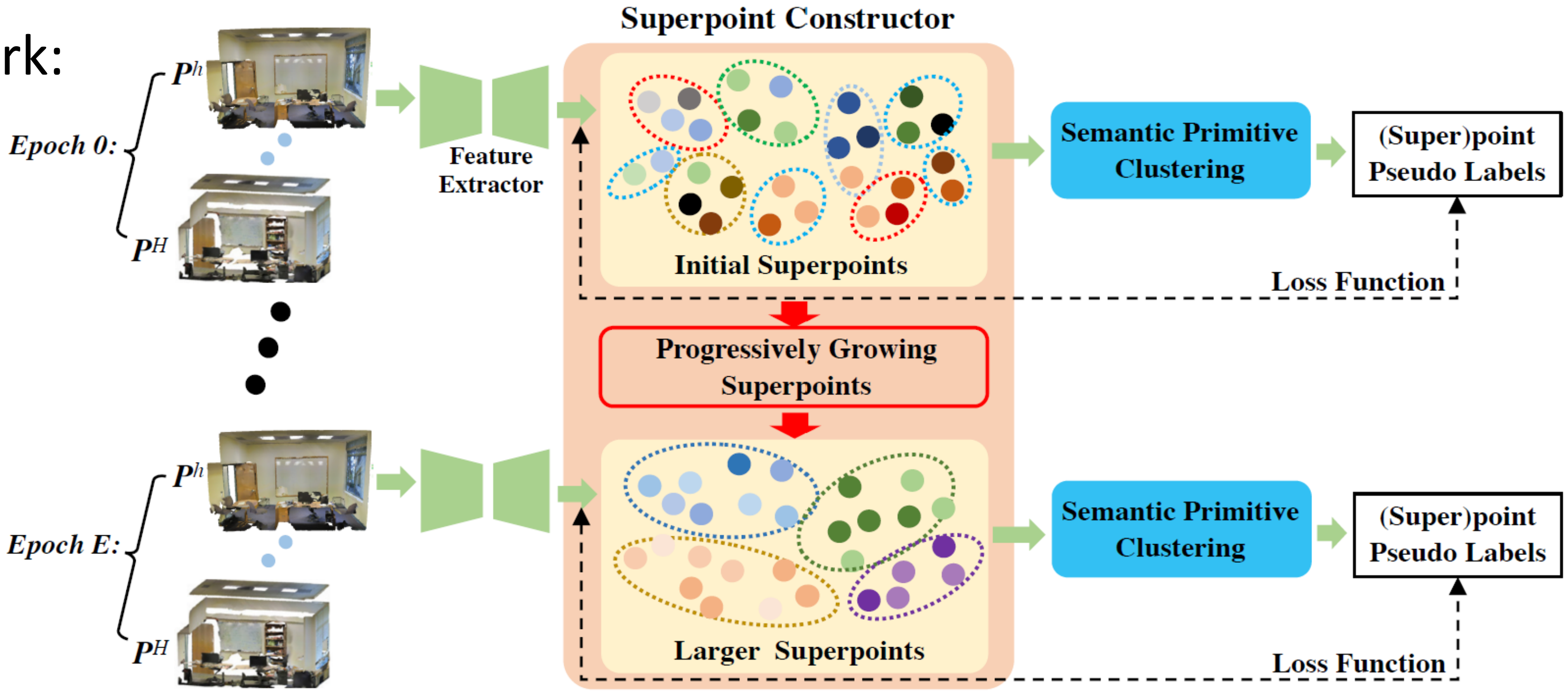
- The network output features are semantically meaningless, especially at the early training stages.

For each superpoints having k points:

- Compute k normal vectors and their cosine distance of any two points.
- Count the distribution of cosine distance to form a histogram within the range [-1, 1]

GrowSP:

- Framework:



1. Get per-point features.
2. Get superpoint-wise features.
3. Do clustering on superpoint-wise features.
4. Training backbone by pseudo labels.

Experiments:

- Compare with other unsup methods:

S3DIS:

		OA(%)	mAcc(%)	mIoU(%)
Unsupervised Methods	RandCNN	23.1	18.4	9.3
	van Kmeans	20.0	21.5	8.8
	van Kmeans-S	20.0	22.3	8.8
	van Kmeans-PFH	23.9	24.7	10.9
	van Kmeans-S-PFH	23.4	20.8	9.5
	IIC [24]	32.8	14.7	8.5
	IIC-S [24]	29.4	15.1	7.7
	IIC-PFH [24]	29.5	13.2	6.7
	IIC-S-PFH [24]	26.3	13.6	7.2
	PICIE [7]	46.4	28.1	17.8
	PICIE-S [7]	50.7	30.8	21.6
	PICIE-PFH [7]	55.0	38.8	26.6
	PICIE-S-PFH [7]	49.1	40.5	26.7
	GrowSP (Ours)	**76.0**	**59.4**	**44.6**

ScanNet:

		OA(%)	mAcc(%)	mIoU(%)
Unsupervised Methods	RandCNN	11.9±0.4	8.4±0.1	3.2±0
	van Kmeans	10.1±0.1	10.0±0.1	3.4±0
	van Kmeans-S	10.2±0.1	9.8±0.3	3.4±0.1
	van Kmeans-PFH	10.4±0.2	10.3±0.7	3.5±0.2
	van Kmeans-S-PFH	12.2±0.6	9.3±0.5	3.6±0.1
	IIC [24]	27.7±2.7	6.1±1.2	2.9±0.8
	IIC-S [24]	18.3±2.6	6.7±0.6	3.4±0.1
	IIC-PFH [24]	25.4±0.1	6.3±0	3.4±0
	IIC-S-PFH [24]	18.9±0.3	6.3±0.2	3.0±0.1
	PICIE [7]	20.4±0.5	16.5±0.3	7.6±0
	PICIE-S [7]	35.6±1.1	13.7±1.5	8.1±0.5
	PICIE-PFH [7]	23.1±1.4	14.0±0.1	8.1±0.3
	PICIE-S-PFH [7]	23.6±0.4	15.1±0.6	7.4±0.2
	GrowSP (Ours)	**57.3**±2.3	**44.2**±3.1	**25.4**±2.3

SemanticKITTI:

	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
IIC [23]	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 [23]	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
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	4.826.5±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
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

Experiments:

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

- Cross-datasets(Novel class discovery):

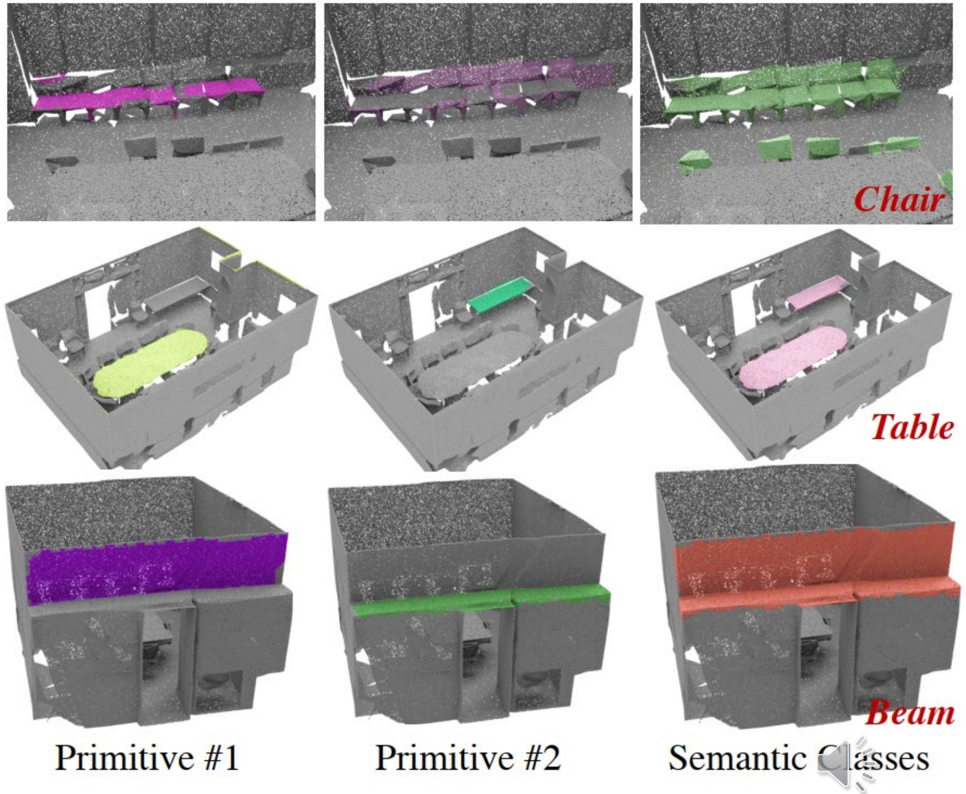
S3DIS →ScanNet:

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 [23]	3.5±0	3.4±0	3.7±0.1	3.5±0.1	3.5±0	3.6±0
IIC-S [23]	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

ScanNet → S3DIS:

test on →	Area-1	Area-2	Area-3	Area-4	Area-5	Area-6
IIC [23]	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 [23]	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)	**28.2**±1.4	**22.9**±2.5	**31.4**±1.5	**25.2**±1.0	**28.6**±2.5	**30.6**±2.0

- Learned Semantic Primitives



Experiments:

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

- Visualizations: