Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds




Francis Engelmann*

Theodora Kontogianni*

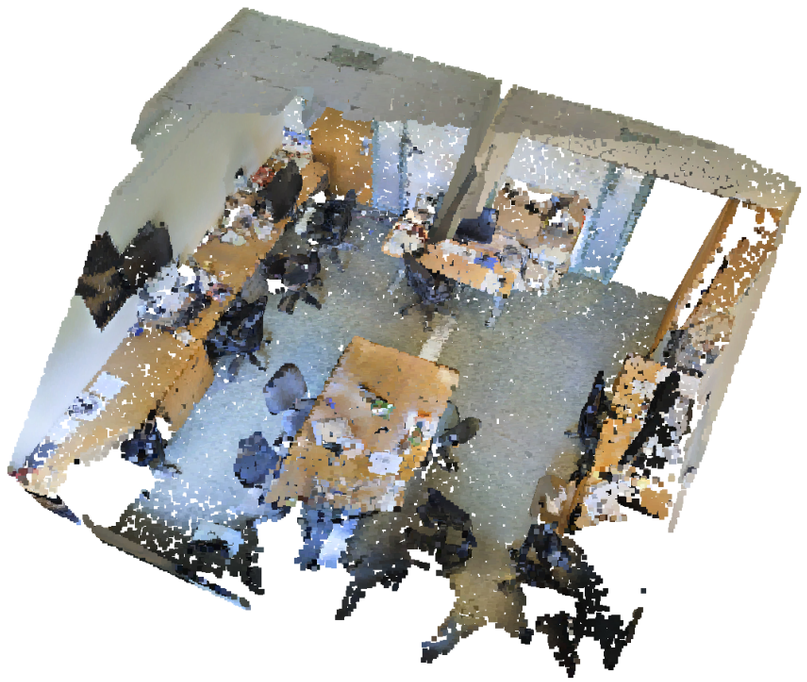
Alexander Hermans

Bastian Leibe

Visual Computing Institute | RWTH AACHEN UNIVERSITY

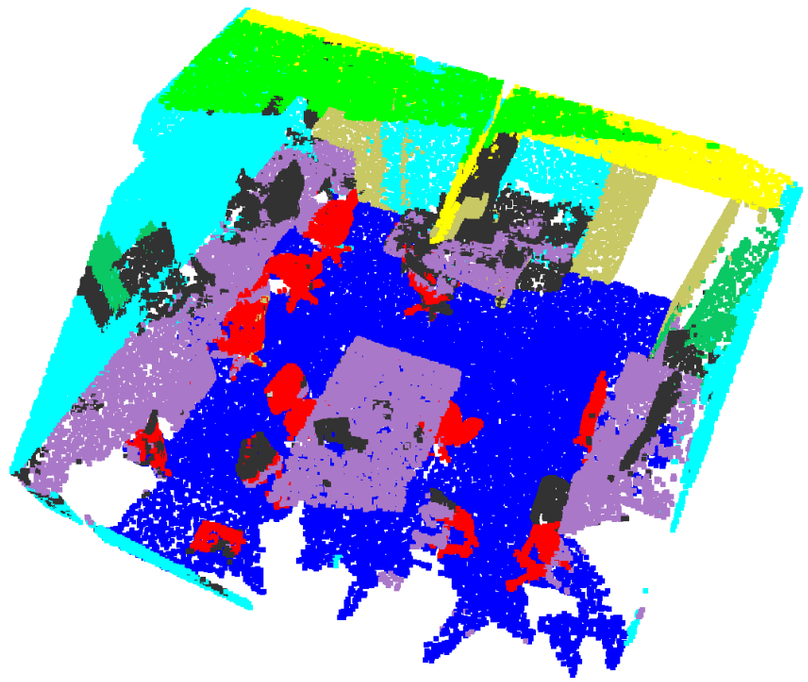
Problem Statement

Input



3D Point Cloud

Output

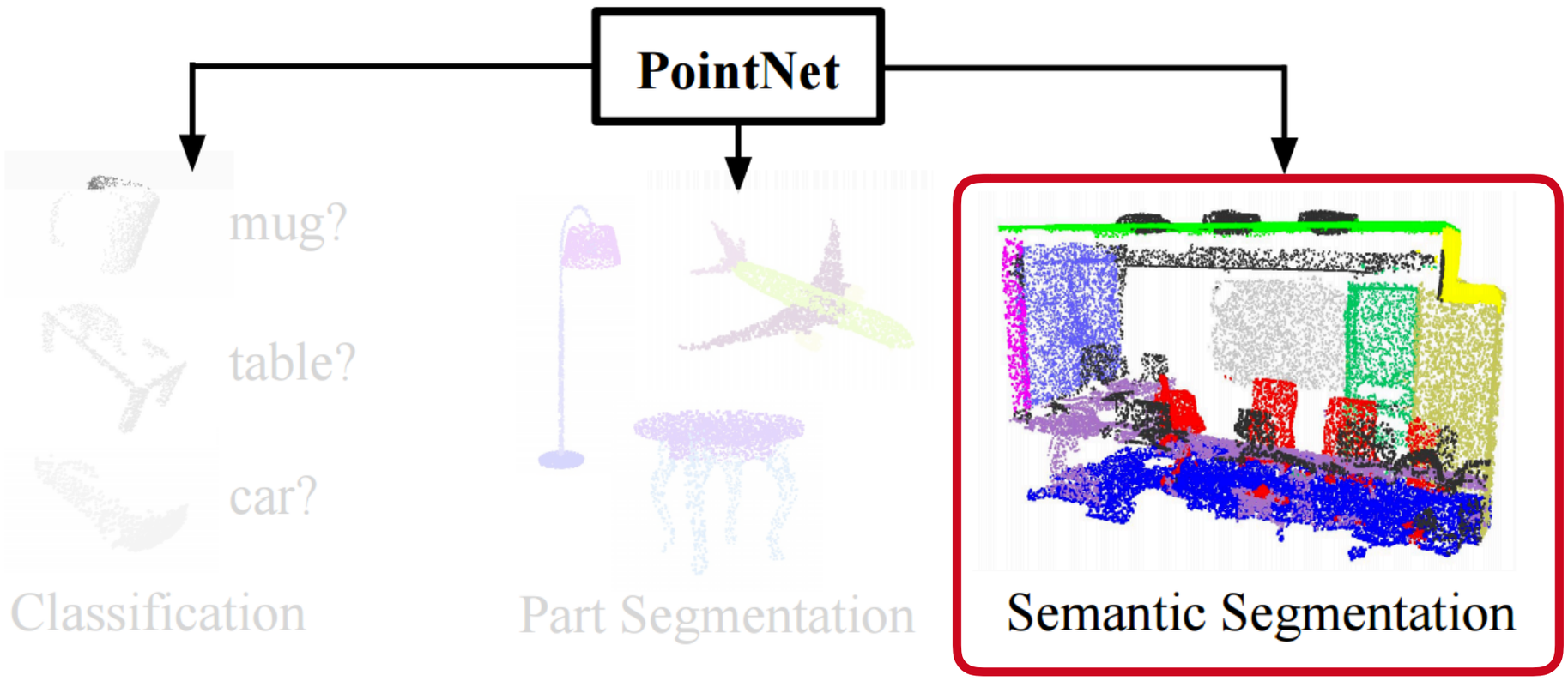


Semantic Segmentation

Previous Work

- Most existing approaches: first convert into another representation
  - Voxel-grid (3D CNN), Projection (2D CNN), …

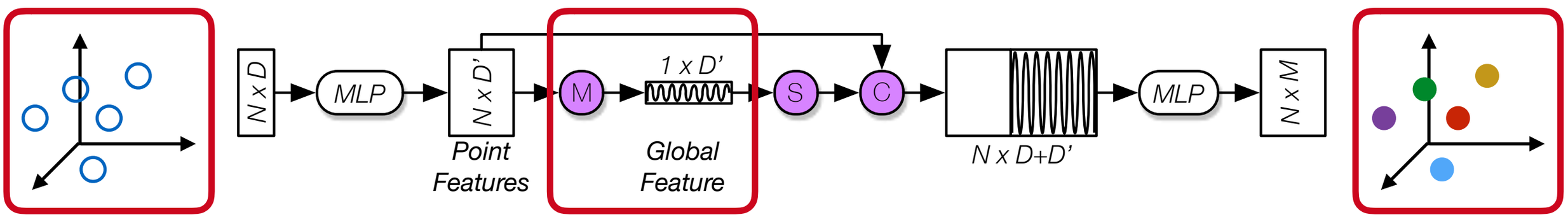
- Pioneering work: **PointNet** operates directly on point clouds [CVPR'17]



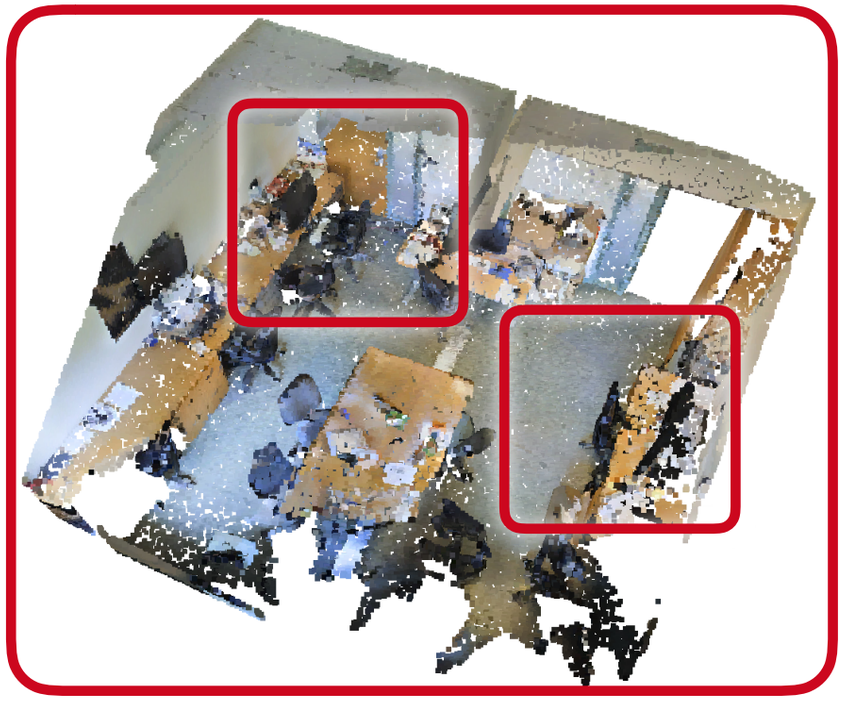
[Charles R. Qi et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR 2017]

Previous Work: PointNet

Idea: Given a point cloud, learn feature descriptor using max-pooling.



(simplified PointNet model)



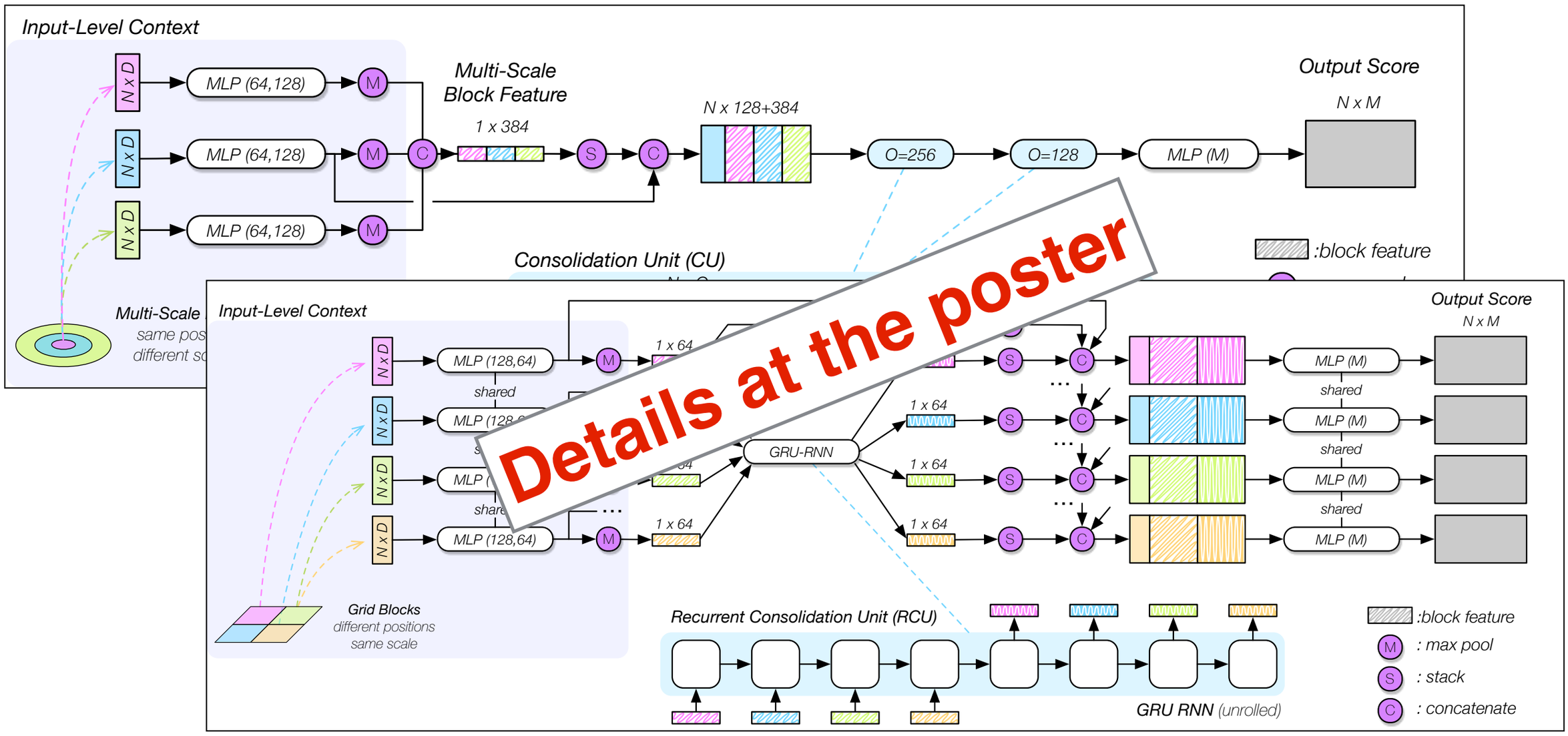
Local Context

Global Context

RWTH AACHEN UNIVERSITY

Our method

Two explorative models …

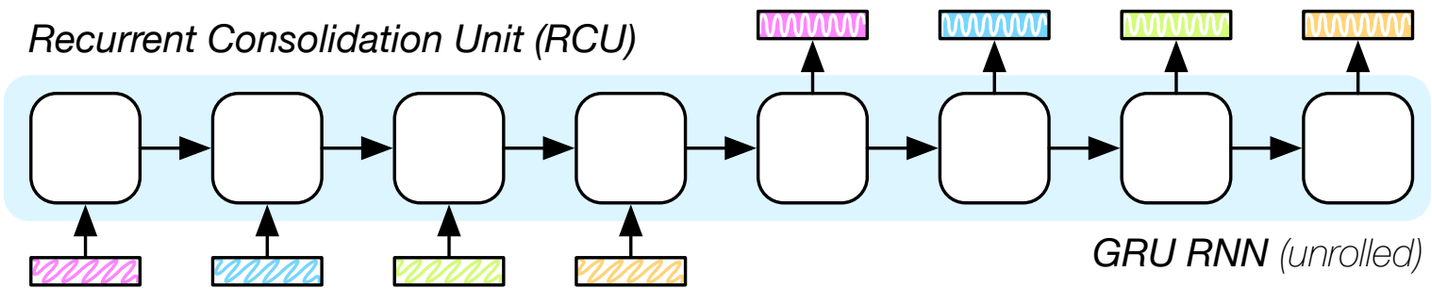


RWTH AACHEN UNIVERSITY

Our method: Consolidation Units

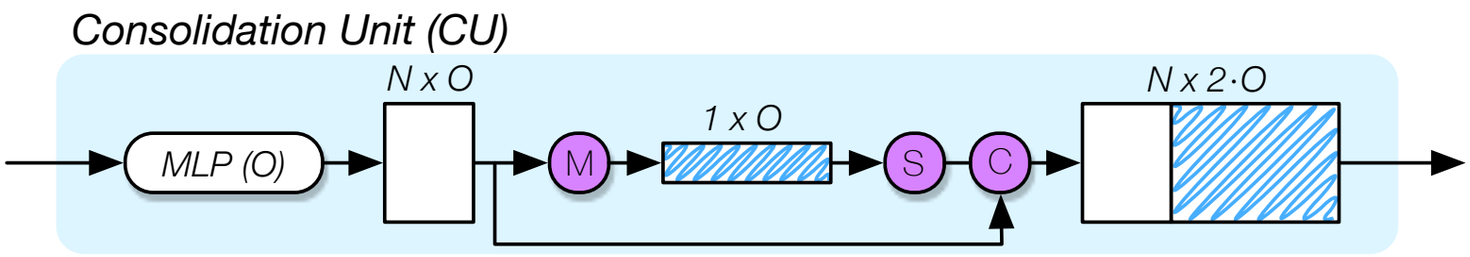
Recurrent Consolidation Units:

Share context between neighboring subsets of points.



Consolidation Units:

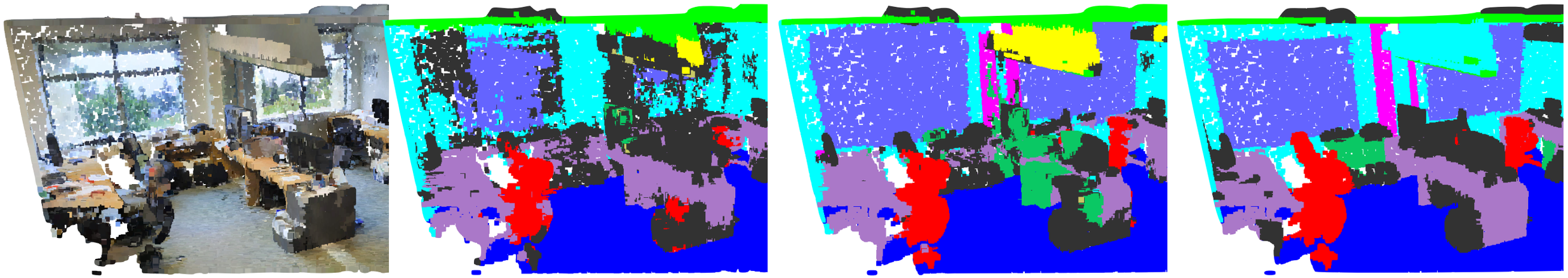
Share and reinforce context between points within the same subset.



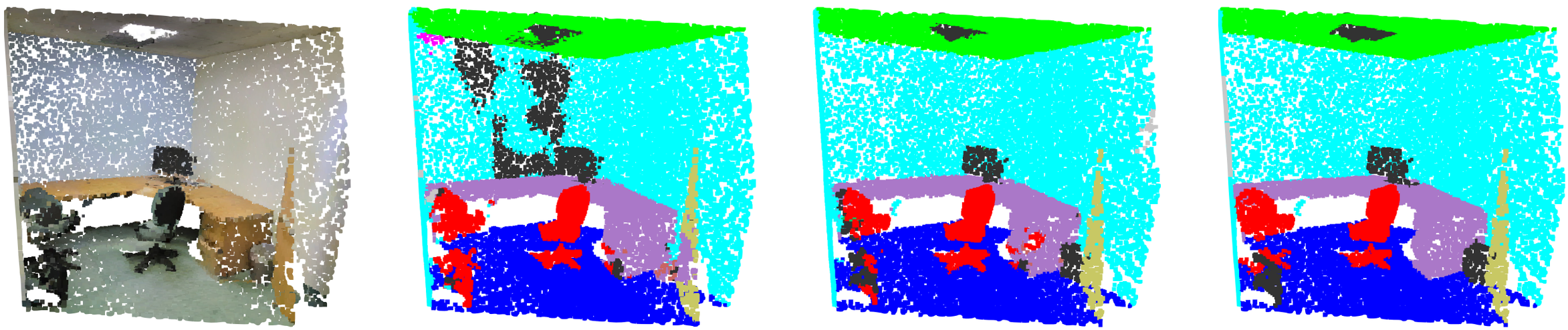
Qualitative Results [S3DIS dataset, Armeni et al. CVPR'16]

Input **PointNet** **Ours** **Ground Truth**

Example 1



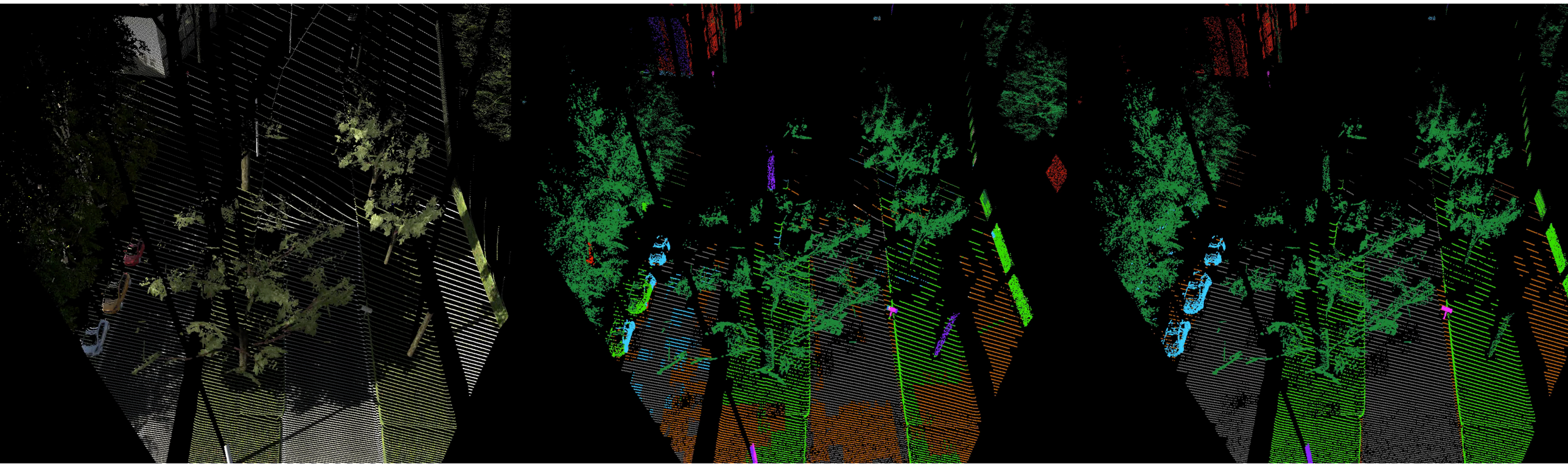
Example 2



RWTH AACHEN UNIVERSITY

Qualitative Results [virtual KITTI dataset, Gaidon et al. CVPR16]

Input XYZ-RGB **Our prediction** **Ground Truth**



Quantitative Results

Geometry & Appearance

XYZ-RGB input features

S3DIS Dataset [1] XYZ-RGB	mean IoU	overall accuracy	avg. class accuracy
*PointNet [26]	43.5	75.0	55.5
*MS	44.4	75.5	57.6
*MS + RCU	45.5	77.2	57.2
*SS + CU(1)	45.9	77.8	57.7
*MS + CU(2)	**47.8**	**79.2**	**59.7**
PointNet [26]	47.6	78.5	66.2
G + RCU	**49.7**	**81.1**	**66.4**

Geometry Only

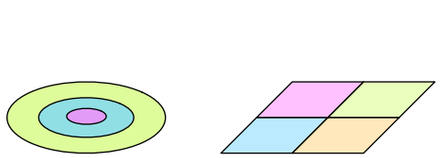
XYZ input features

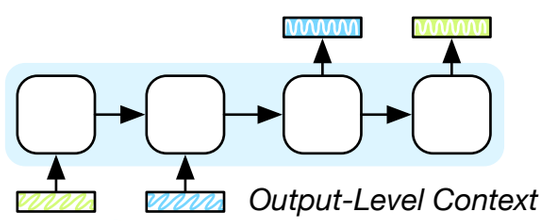
	mean IoU	overall accuracy	avg. class accuracy
S3DIS Dataset [1] – no RGB			
*PointNet [26]	40.0	72.1	52.9
*MS + CU(2)	**43.0**	**75.4**	**55.2**
vKITTI Dataset [8] – no RGB			
*PointNet [26]	17.9	63.3	29.9
*MS + CU(2)	**26.4**	**73.2**	**40.9**

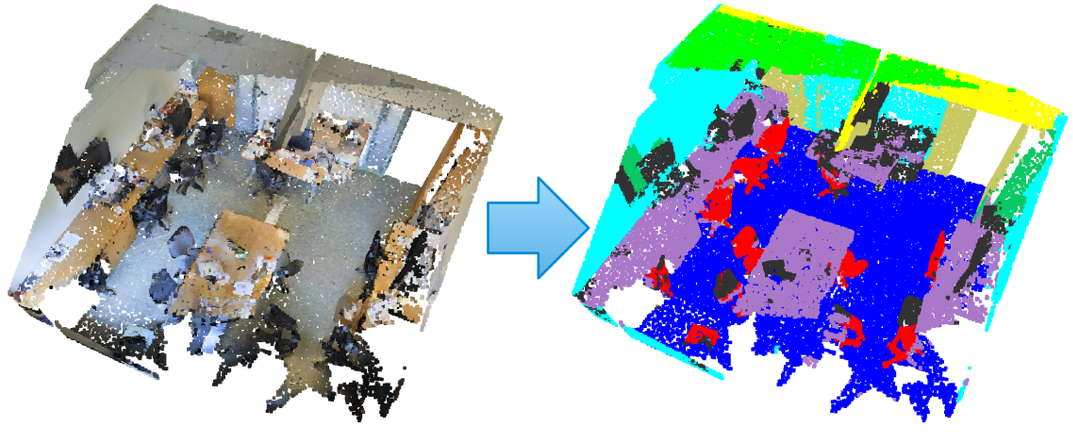
Conclusion

We present novel mechanisms ***(Consolidation Units)*** to:

- share local context globally across the scene
- reinforce/consolidate local context







See you at our poster!

Project page: https://www.vision.rwth-aachen.de/page/3dsemseg

RWTH AACHEN UNIVERSITY