Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds

Francis Engelmann*, Theodora Kontogianni*, Alexander Hermans, Bastian Leibe

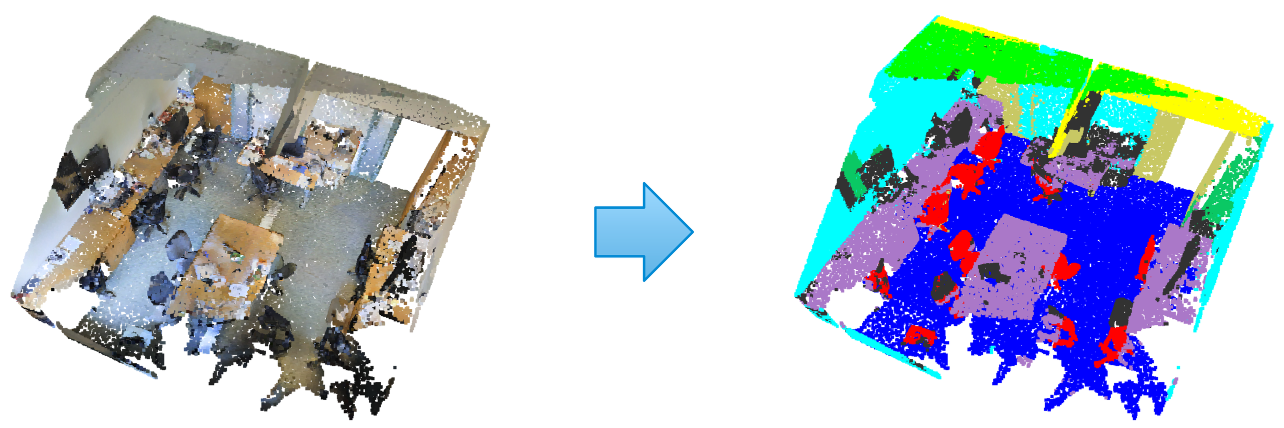
Computer Vision Group, Visual Computing Institute, RWTH Aachen University

Abstract

Deep learning approaches have made tremendous progress in the field of semantic segmentation over the past few years. However, most current approaches operate in the 2D image space. Direct semantic segmentation of unstructured 3D point clouds is still an open research problem. The recently proposed PointNet architecture presents an interesting step ahead in that it can operate on unstructured point clouds, achieving decent segmentation results. However, it subdivides the input points into a grid of blocks and processes each such block individually. In this paper, we investigate the question how such an architecture can be extended to incorporate larger-scale spatial context. We build upon PointNet and propose two extensions that enlarge the receptive field over the 3D scene. We evaluate the proposed strategies on challenging indoor and outdoor datasets and show improved results in both scenarios.

Problem Statement

Input **Output**

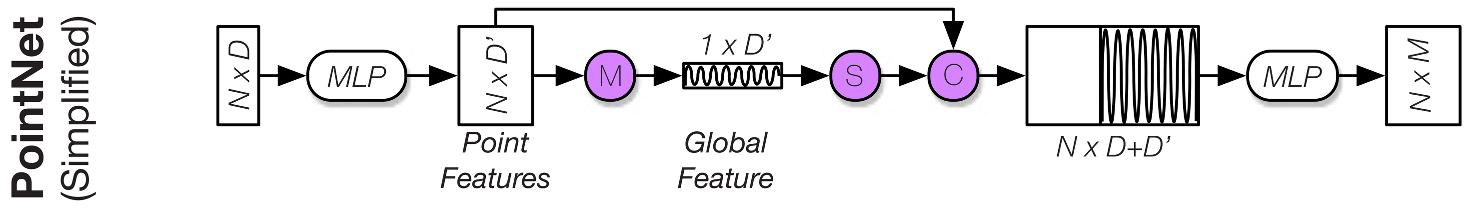


Unstructured 3D Point Cloud — Semantic Segmentation

Contributions

- We present two mechanisms that increase the spatial context for semantic segmentation of 3D point clouds: **input- and output-level context**.
- We verify experimentally that our proposed extensions achieve improved results on challenging **indoor and outdoor datasets**.
- We show competative results using **only geometric input features** (no color).

Background

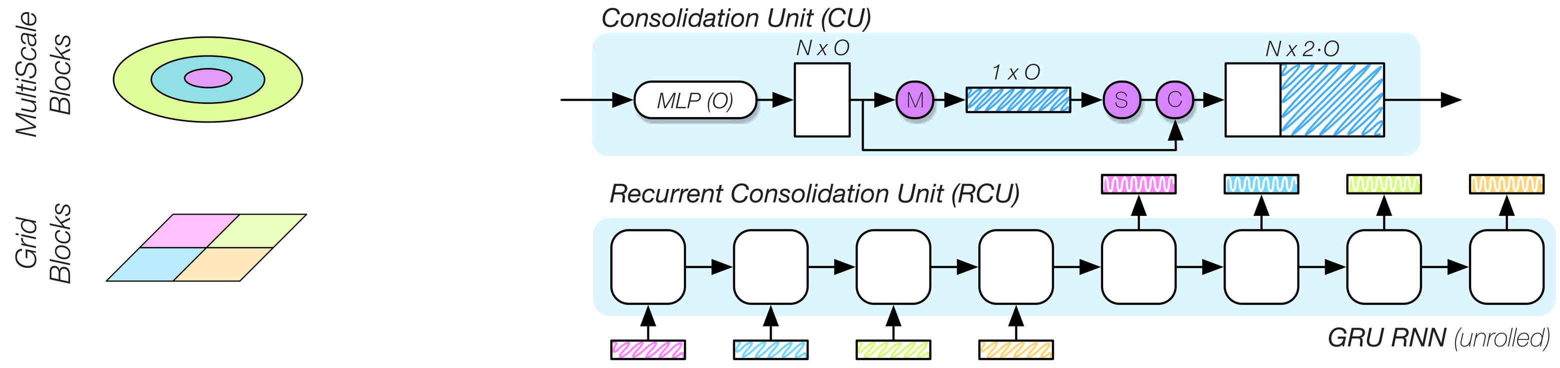


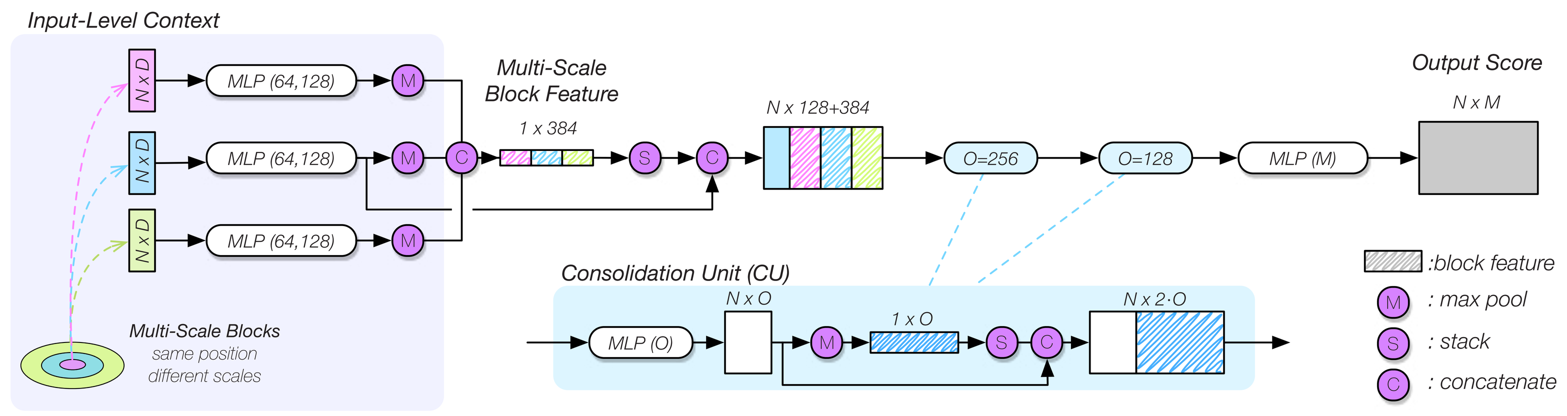
- Main Idea of PointNet: Compute a **global feature** summarizing a set of unordered point features using **max-pooling** (M).
- Prediction is based on point features representing local context concatenated (C) with the global feature representing neighboring context.
- Neigborhood is limited spatially up to a certain radius.

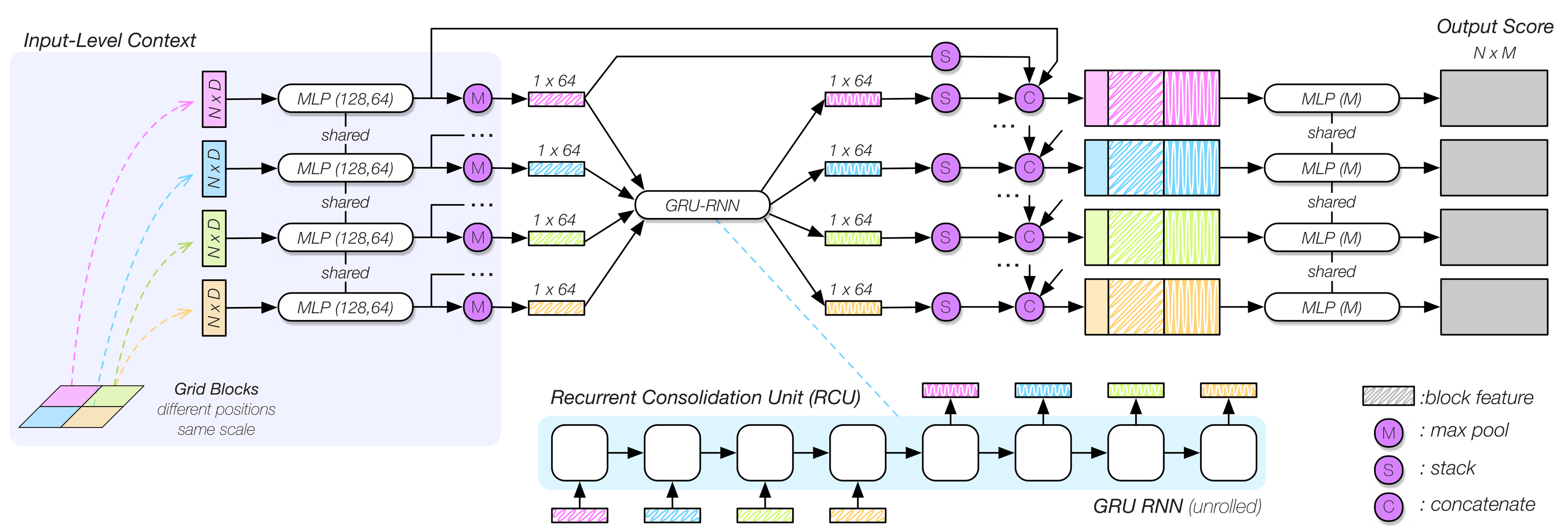
Our Method

We present different mechanisms to **increase** and **consolidate** spatial context.

Input-Level Context: **Output-Level Context:**







Evaluation

Datasets and evaluation metrics: We evaluate our method on S3DIS [2] and virtual KITTI [3] in terms of mean IoU, overall pixel accuracy and average class accuracy. Both datasets include per point semantic class annotations.

Quantitative Results

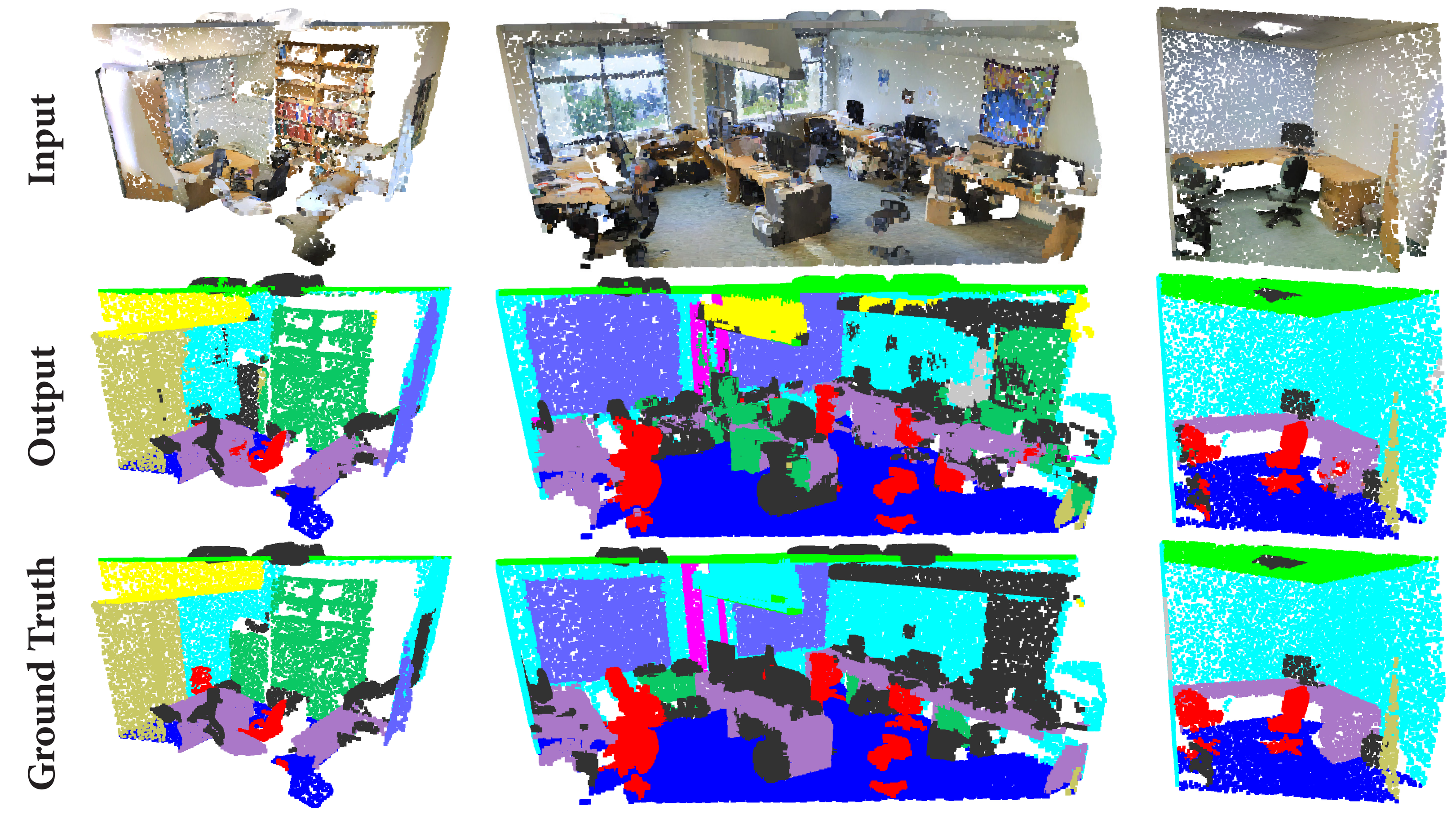
Geometry + Appearance

S3DIS Dataset [2] XYZ-RGB	mean IoU	overall accuracy	avg. class accuracy
*PointNet [1]	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 [1]	47.6	78.5	66.2
G + RCU	**49.7**	**81.1**	**66.4**

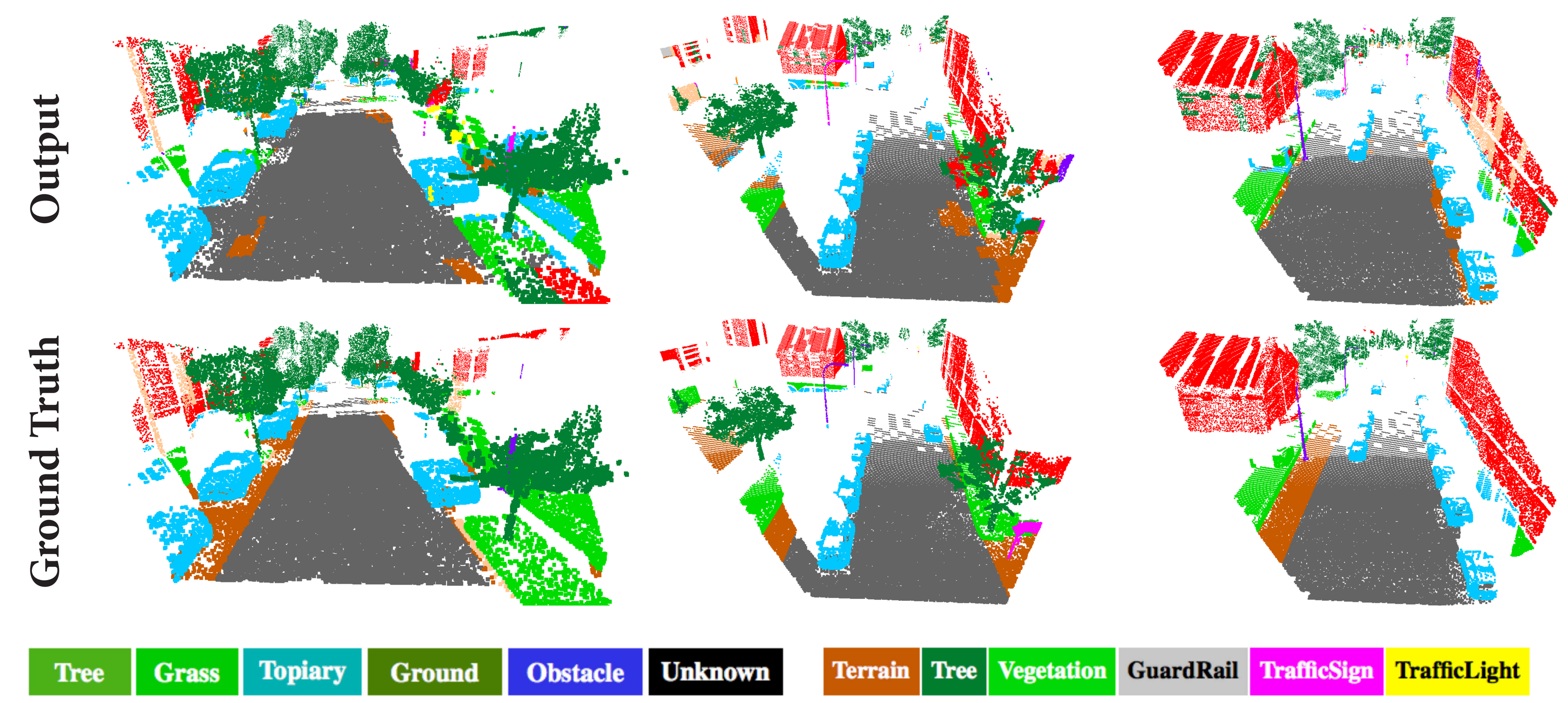
Geometry Only

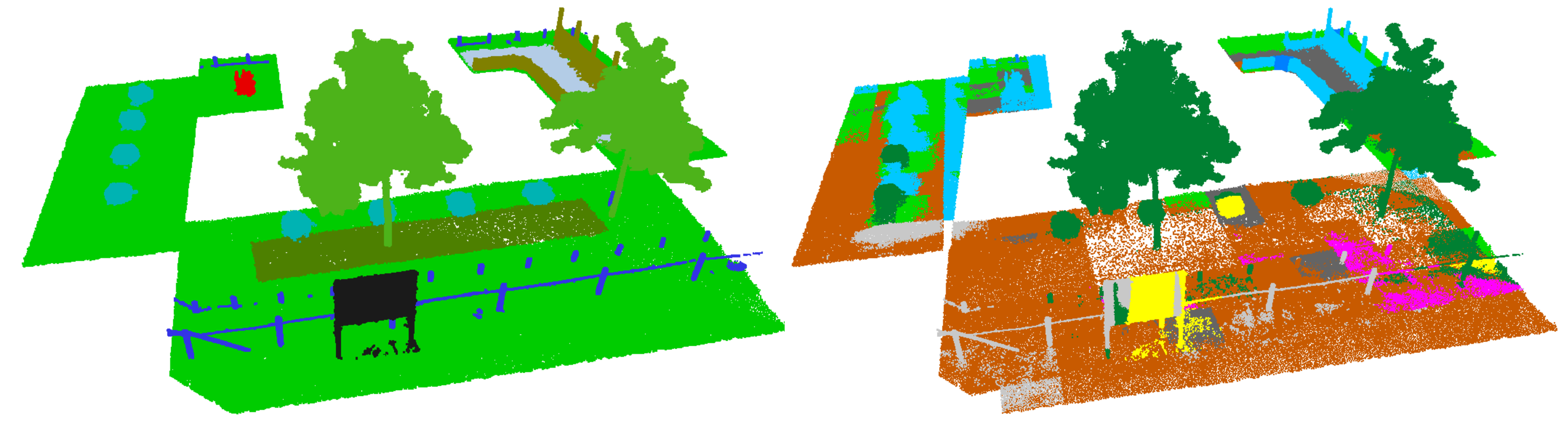
	mean IoU	overall accuracy	avg. class accuracy
S3DIS Dataset [2] - no RGB			
*PointNet [1]	40.0	72.1	52.9
*MS + CU(2)	**43.0**	**75.4**	**55.2**
vKITTI Dataset [3] - no RGB			
*PointNet [1]	17.9	63.3	29.9
*MS + CU(2)	**26.4**	**73.2**	**40.9**

Qualitative Results



Geometry-Only Results





References

[1] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[2] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3D Semantic Parsing of LargeScale Indoor Spaces. Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[3] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig. Virtual Worlds as Proxy for Multi-Object Tracking Analysis. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

M.Sc. Francis Engelmann, M.Sc. Theodora Kontogianni
Phone: +49 241 80 20 {760, 759}
E-Mail: {engelmann, kontogianni}@vision.rwth-aachen.de
Website: http://www.vision.rwth-aachen.de/publication/00146/

This project has received funding from the European Research Council (ERC) CV-SUPER (ERC-2012-StG-307432).

Computer Vision Group, Visual Computing Institute
RWTH Aachen University
Mies-van-der-Rohe Str. 15, D-52074 Aachen, Germany
http://www.vision.rwth-aachen.de

Visual Computing Institute

RWTH AACHEN UNIVERSITY