Tangent Convolutions for Dense Prediction in 3D: Supplementary Material

1. Robustness to noise

We evaluated the robustness of our approach to noise. Several instances of our network were trained on the S3DIS dataset perturbed with different amounts of additive Gaussian noise with standard deviation σ. The results are reported in Table 1. We selected small subsets of the data for training and testing (Area 1 for training and Area 5 for testing), which is why the final performance numbers are not compatible with those reported in the main paper.

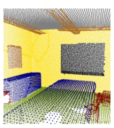
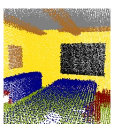
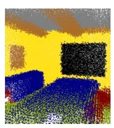
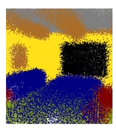
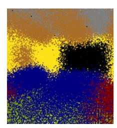
Training data					
σ, m	0.00	0.02	0.04	0.08	0.16
OA	0.59	0.63	0.63	0.68	0.17

Table 1. Performance evaluation with different levels of noise.

Surprisingly, reasonable amounts of noise improve overall accuracy. The method only suffers if the noise severely damages semantic structure in the point cloud. We did not tune any parameters in the pipeline for these experiments.

2. Signal interpolation

In this experiment we compare the effectiveness of two signal interpolation schemes: nearest neighbor and Gaussian mixture. Quantitative results on S3DIS using D and H as input signals are presented in Table 2. Both methods produce very similar results which is why the simpler nearest neighbor interpolation is used throughout the paper.

Signal	mIoU	mA	oA
NN	50.0	60.0	81.2
Gaussian	50.7	59.6	81.3

Table 2. Signal interpolation using the nearest neighbor scheme and the Gaussian mixture scheme produce similar results.

3. Comparison with SnapNet

We also compared our approach with the SnapNet by Boulch et al. [2]. They project a 3D scene onto a set of 2D images. Those images are then segmented with a regular 2D ConvNet. The main strength of this approach is the possibility to combine it with transfer learning and use the weights of a network pre-trained on ImageNet for initialization. Applying this strategy yields the mIoU score of 67.7 on the Semantic3D datset, compared to 66.4 produced by our approach. However, the non-trivial camera pose sampling procedure required by SnapNet did not allow us to apply it to indoor datasets.

4. Qualitative results

We provide more qualitative results of our method on different datasets in Figures 1-3.

References

[1] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3D semantic parsing of large-scale indoor spaces. In *CVPR*, 2016. 2

[2] A. Boulch, B. L. Saux, and N. Audebert. Unstructured point cloud semantic labeling using deep segmentation networks. In *Eurographics Workshop on 3D Object Retrieval*, 2017. 1

[3] A. Dai, A. X. Chang, M. Savva, M. Halber, T. A. Funkhouser, and M. Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *CVPR*, 2017. 2

[4] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys. Semantic3D.net: A new large-scale point cloud classification benchmark. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017. 3

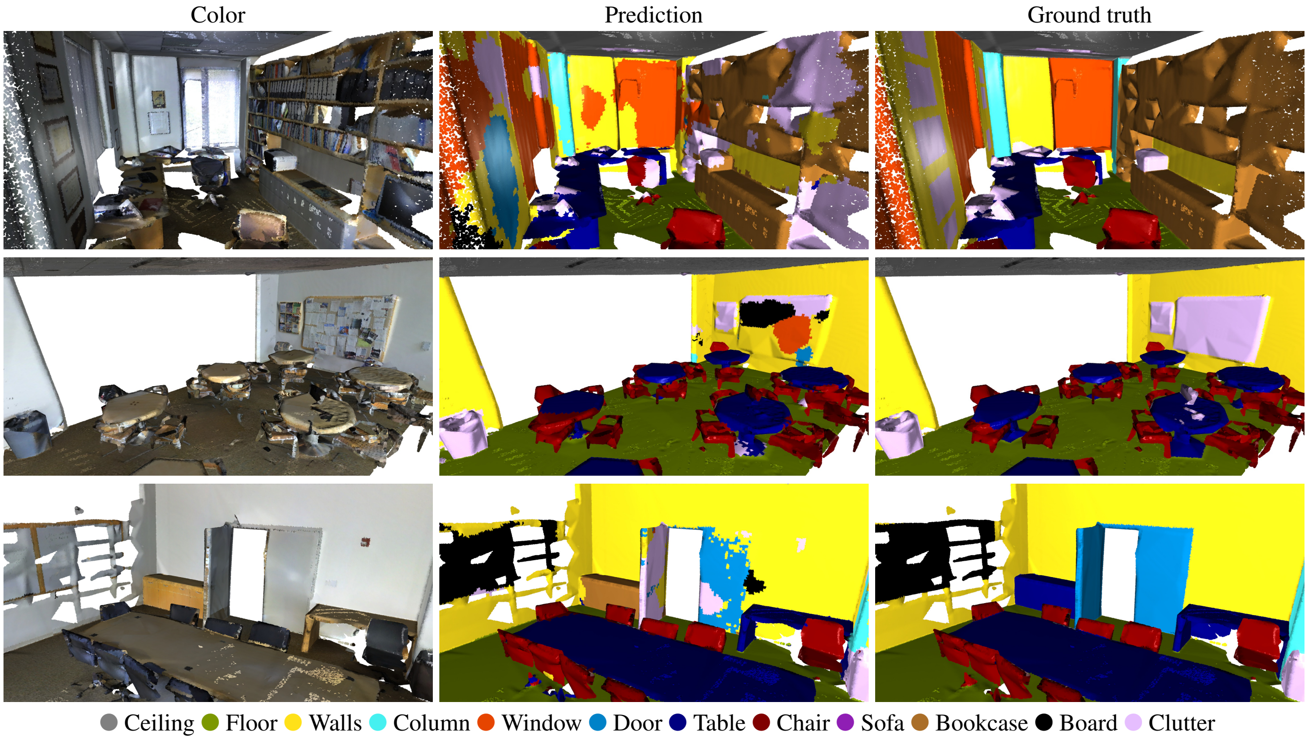


Figure 1. Qualitative results on S3DIS [1].

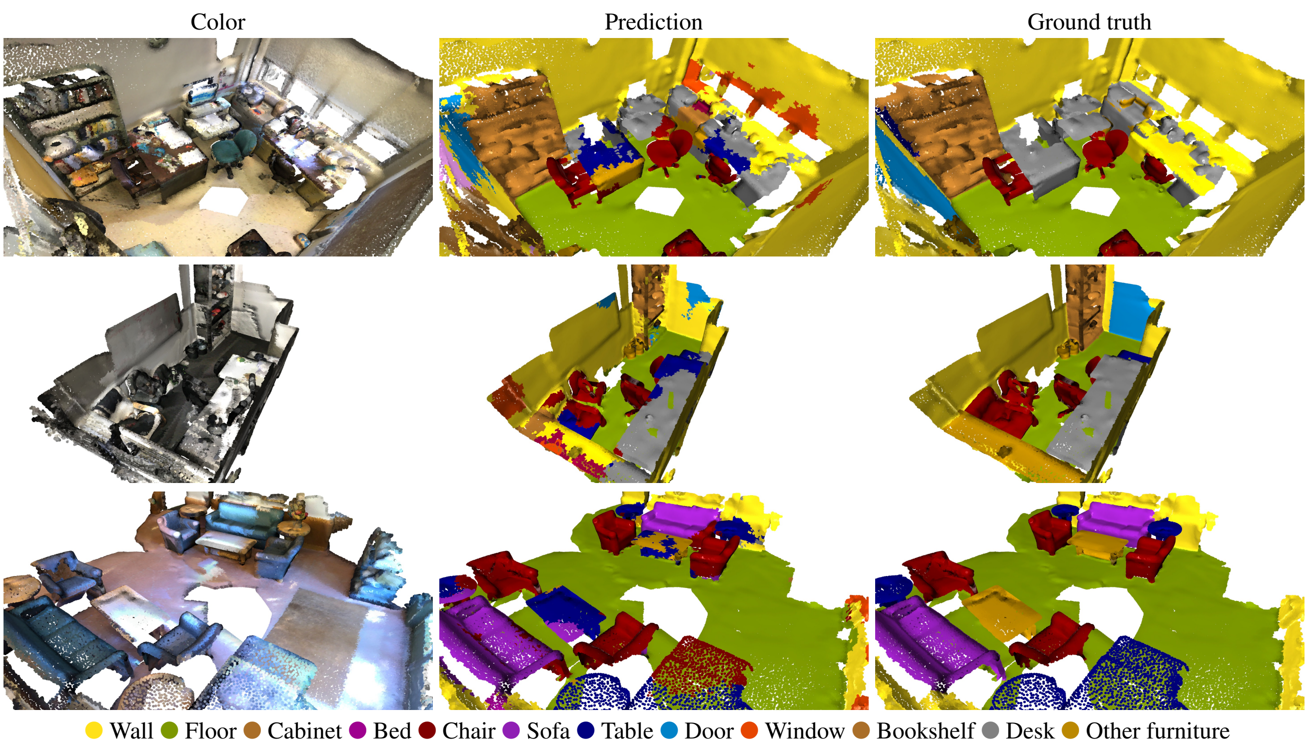


Figure 2. Qualitative results on ScanNet [3].

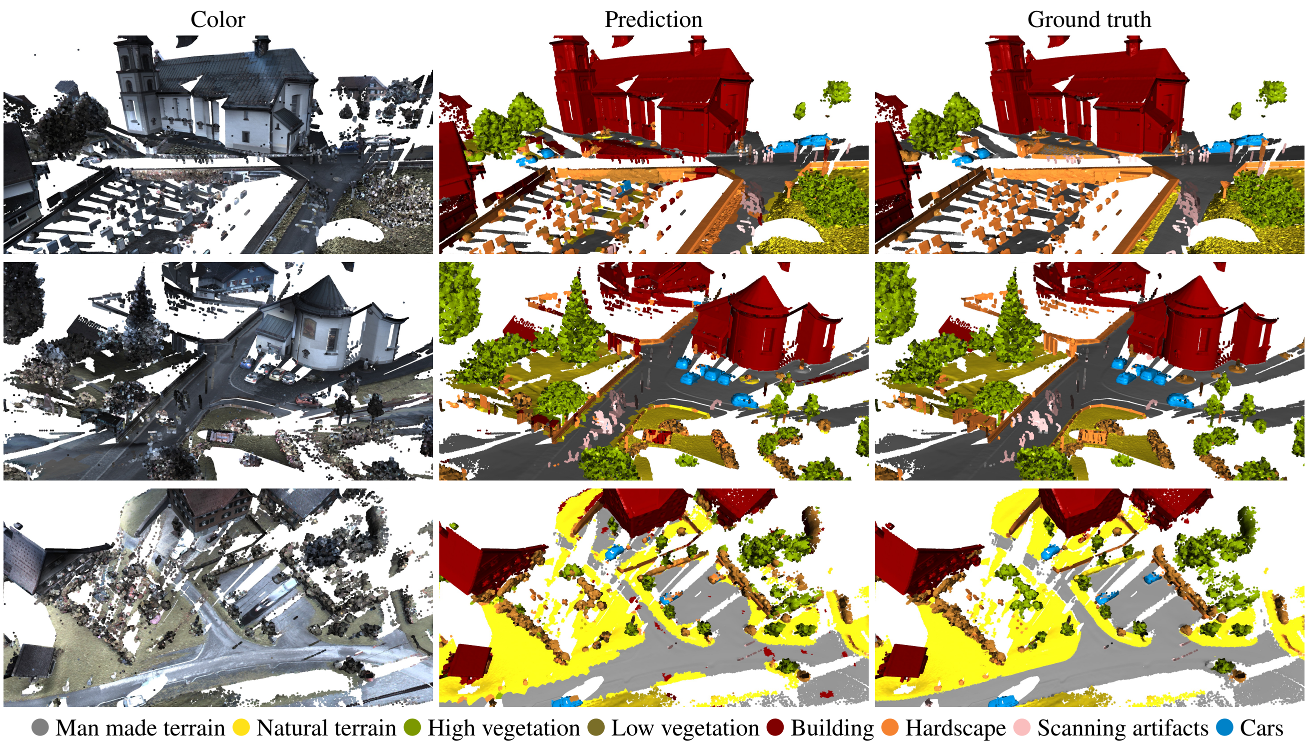


Figure 3. Qualitative results on Semantic3D [4].