

# Supplementary Material for 2D Feature Distillation for Weakly- and Semi-Supervised 3D Semantic Segmentation

Ozan Unal<sup>1</sup> Dengxin Dai<sup>2</sup> Lukas Hoyer<sup>1</sup> Yigit Baran Can<sup>1</sup> Luc Van Gool<sup>1,3,4</sup>  
<sup>1</sup>ETH Zurich, <sup>2</sup>Huawei Technologies, <sup>3</sup>KU Leuven, <sup>4</sup>INSAIT  
{ozan.unal, dai, lukas.hoyer, cany, vangool}@vision.ee.ethz.ch

## 1. Implementation Details

We use Cylinder3D [6] as a baseline 3D model. For the mean teacher, we follow convention and set the update hyperparameter  $\alpha = 0.999$  [4]. For the domain adaptive 2D pipeline we follow DAFormer [3]. We heuristically balance the losses by setting  $\lambda = 0.001$  and  $\lambda_p = 10$ . For semi-supervised, we restrict set  $A$  in FOVMix to labeled frames to ensure we have direct supervision in all samples and do additional rotation augmentation before the FOVMix operations to increase variability.

## 2. Datasets

We run our experiments on the ScribbleKITTI [5] dataset that provides realistic weak labels for LiDAR semantic segmentation in the form of scribbles. ScribbleKITTI is built on SemanticKITTI [1, 2], the most popular large-scale outdoor-scene dataset for LiDAR semantic segmentation, shares the same *valid*-set. The weak labels only provide annotations to 8% of the point count and completely forgo class boundaries. Thus, compared to dense annotations, labeling times are reduced by 10 fold.

For the 2D syntetic training, we use the GTA-V dataset which contains 24966 synthetic images with pixel level semantic annotation. The images are generated using a modded version of the open-world video game Grand Theft Auto 5.

## References

- [1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. SemanticKITTI: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9297–9307, 2019. 1
- [2] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 1
- [3] Lukas Hoyer, Dengxin Dai, and Luc Van Gool. DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1
- [4] Ozan Unal, Dengxin Dai, Ali Tamer Unal, and Luc Van Gool. Discwise active learning for lidar semantic segmentation. *IEEE Robotics and Automation Letters*, 2023. 1
- [5] Ozan Unal, Dengxin Dai, and Luc Van Gool. Scribble-supervised lidar semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2697–2707, June 2022. 1
- [6] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. *arXiv preprint arXiv:2011.10033*, 2020. 1