Supplementary Materials
Category Differences Matter: A Broad Analysis of Inter-Category Error in Semantic Segmentation

Jingxing Zhou
Porsche Engineering Group GmbH
jingxing.zhou@porsche-engineering.de

Jürgen Beyerer
Fraunhofer IOSB & Karlsruhe Institute of Technology
juergen.beyerer@iosb.fraunhofer.de

A. Training Setup

Unless specified, we utilize SGD optimizer with a base learning rate of 1×10^{-2} for all the networks with ResNet backbones and AdamW optimizer with a base learning rate of 1×10^{-4} for other backbones. The training is scheduled with a linear warm-up policy with 1500 iterations. The backbones are pretrained with ImageNet-1k [4]. We use a batch size of 8 for the training across the datasets and train the network with 80k iterations with crop size of 1024×512. During training, we also apply common data augmentation methods like *random flip* and *photometric distortion* from [1] to increase the data variety. We do not apply methods like auxiliary head or stage-wise learning rate decay for simplification, although they may contribute to better network generalization or training stability. Since there exists higher uncertainty of the networks when the input data distribution drifts from the source, we repeat the training three times and report the average value of each class.

B. Class Taxonomies

In our work, four different class hierarchies are used for the evaluation. Detailed class taxonomy for Cityscapes [2] is provided in Fig. B.1. Fig. B.2 depicts a simplified class hierarchy for Mapillary [3], while Fig. B.3 and Fig. B.4 show the behavior-based class taxonomy and VRU-based class taxonomy for the ablation study. We discard the classes under *void* in Mapillary and relabel them as ignore.

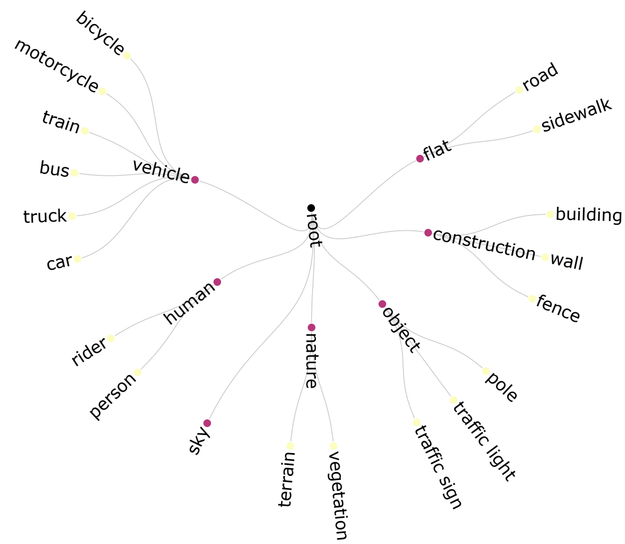

Figure B.1. Class taxonomy of Cityscapes, ACDC and BDD100k datasets.

C. Additional Evaluation Results

We provide additional evaluation results on ACDC [5] and BDD100k dataset [6] with visualization in Fig. C.1, which correspond to the quantitative results that we observe in the domain shift section.

References

[1] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation, 2020. 1

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition(CVPR)*, 2016. 1

[3] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017. 1

[4] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015. 1

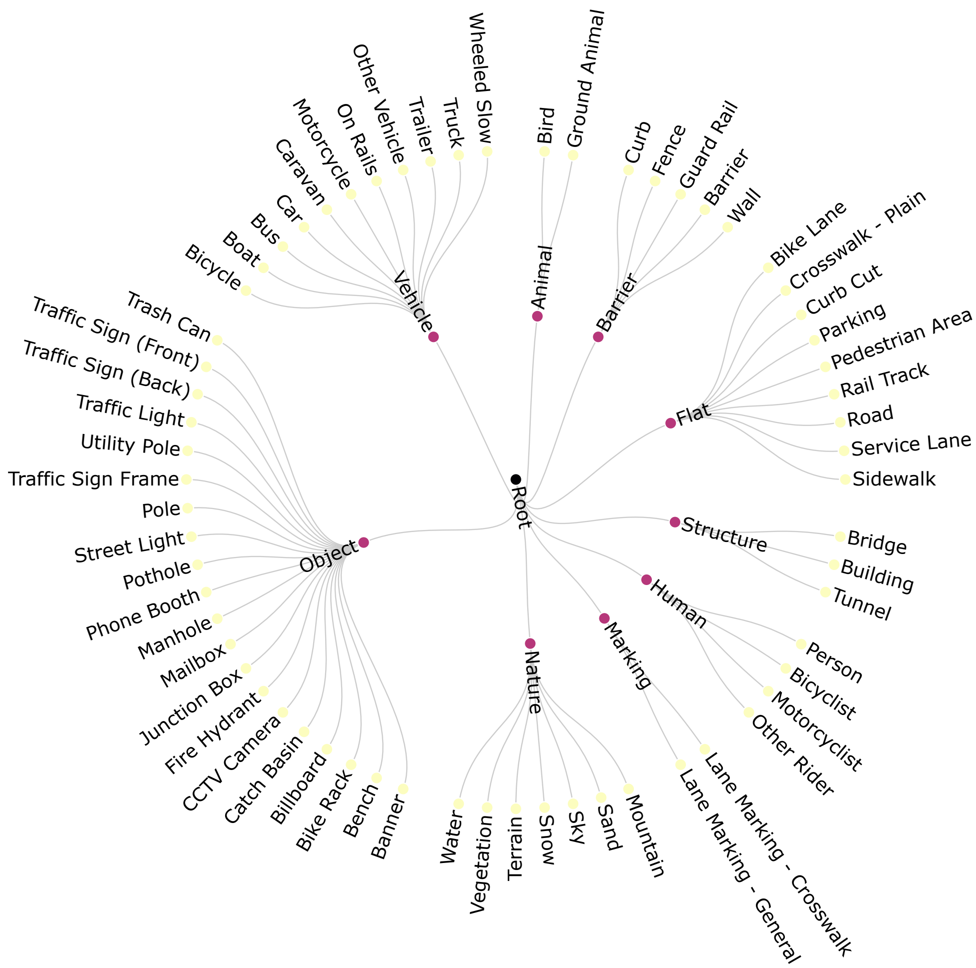

Figure B.2. Simplified class taxonomy of the Mapillary dataset (v1.2); void classes are remapped to ignore during training and validation.

[5] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10765–10775, 2021. 1

[6] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 1

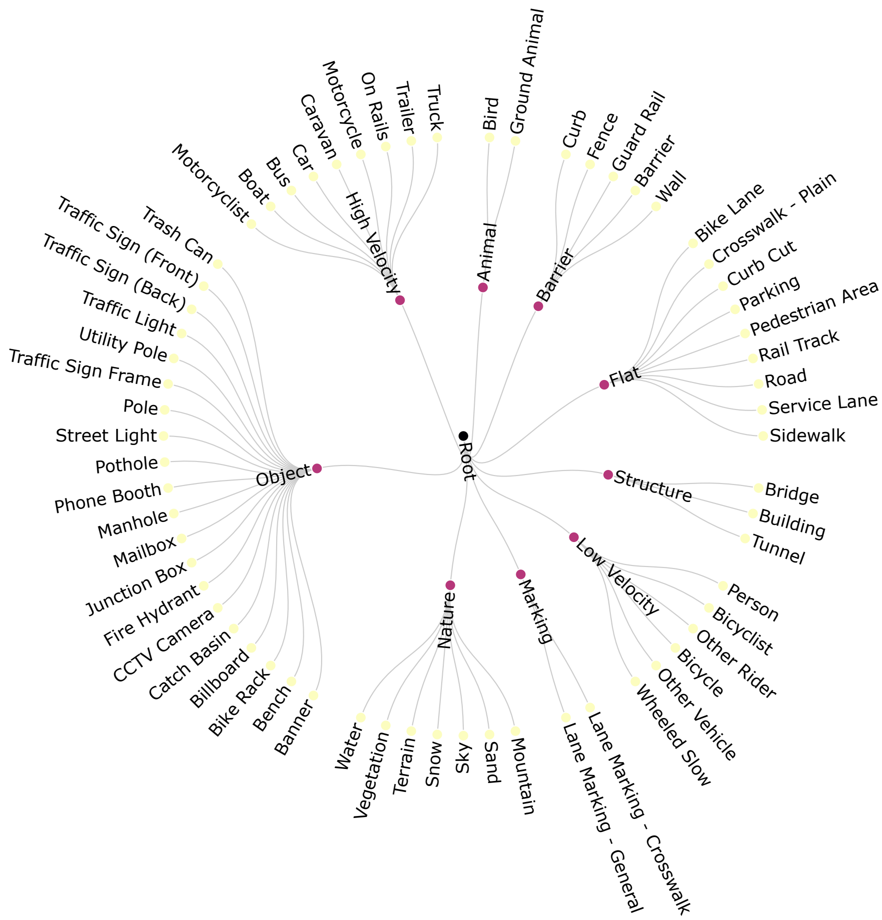

Figure B.3. Behavior-based class taxonomy used for the ablation study from the Mapillary dataset (v1.2); void classes are remapped to ignore during training and validation.

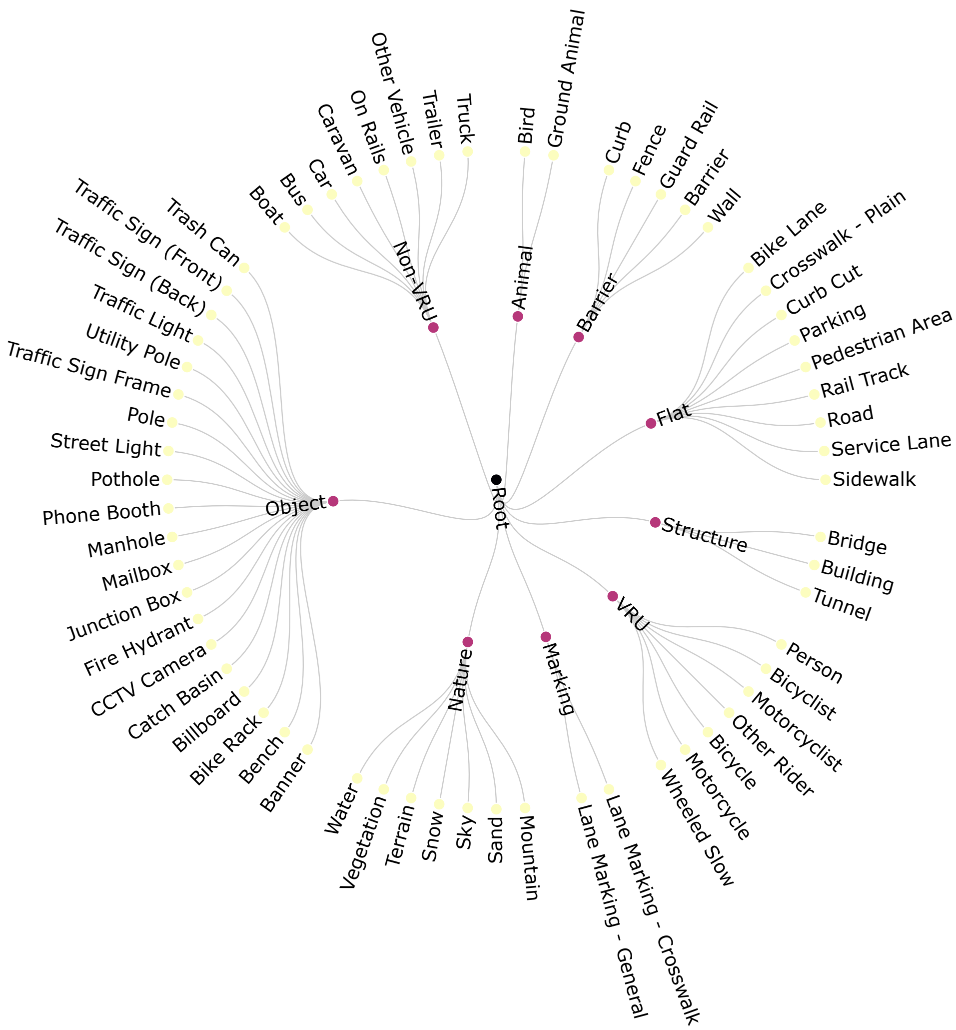

Figure B.4. Class taxonomy that addresses VRUs used for ablation study from the Mapillary dataset (v1.2); void classes are remapped to ignore during training and validation.

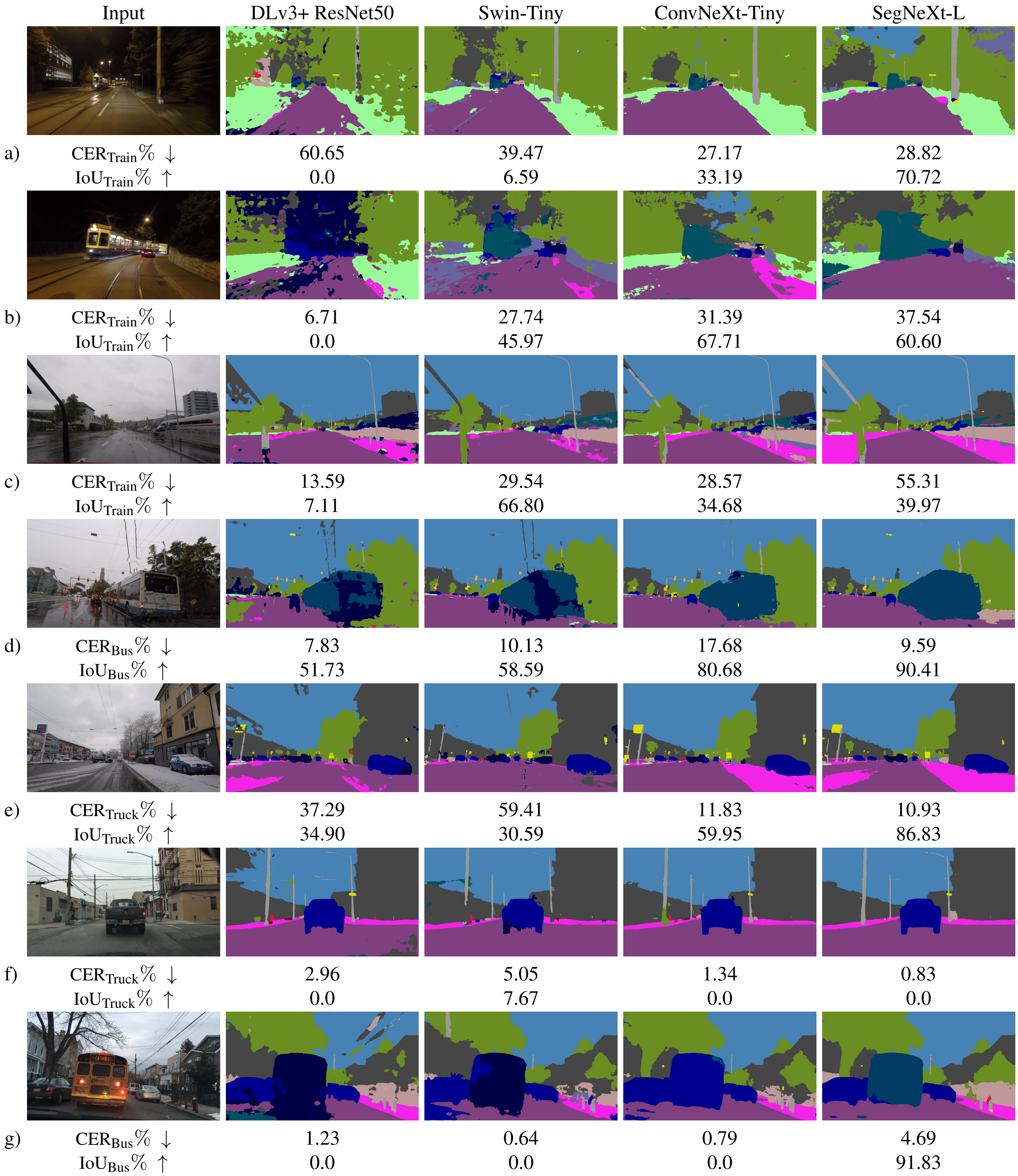

Figure C.1. Evaluation results on ACDC and BDD100k dataset. The neural networks are trained on Cityscapes dataset and evaluate in a domain shift setup. We observe severe impact from the varying label policy that affects the evaluation on BDD100k dataset based on IoU metric in comparison to ACDC dataset. Best viewed in color.