A. Training Setup

Unless specified, we utilize SGD optimizer with a base learning rate of $1 \times 10^{-2}$ for all the networks with ResNet backbones and AdamW optimizer with a base learning rate of $1 \times 10^{-4}$ 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 $1024 \times 512$. 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.

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

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


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.