Supplementary Material

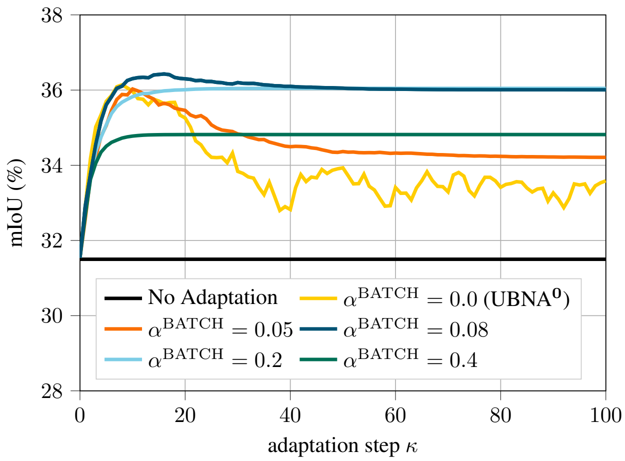

Figure 5. **Hyperparameter analysis**: Performance of **UBNA** on the **Cityscapes validation set** for the adaptation from **GTA-5** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) in **dependence of the batch decay factor** α^{BATCH}. Here, $\alpha^{\mathrm{LAYER}} = 0.0$ for all UBNA experiments.

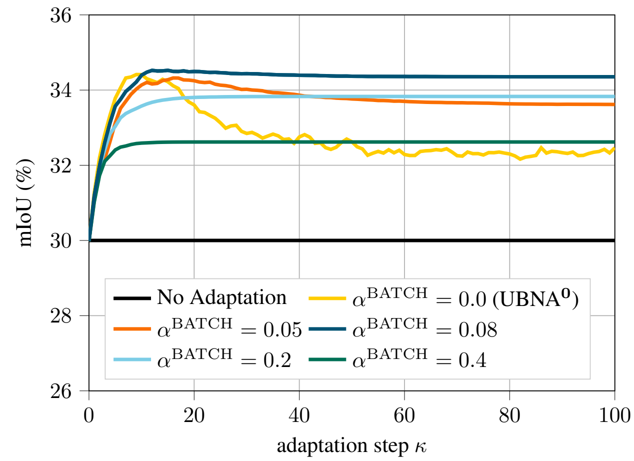

Figure 6. **Hyperparameter analysis**: Performance of **UBNA** on the **Cityscapes validation set** for the adaptation from **SYNTHIA** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) in **dependence of the batch decay factor** α^{BATCH}. Here, $\alpha^{\mathrm{LAYER}} = 0.0$ for all UBNA experiments.

A. Additional Experimental Evaluation

With this section our main aim is to provide a more complete overview on the results of our method. This in particular means that we provide a hyperparameter analysis, show our conducted experiments on other source/target domain combinations, provide results for more network architectures, and give a deeper analysis on the variance of different experiments with the same hyperparameter setting but different random seeds. Note that most subsection names are equal to the ones in Sec. 6 of the main paper, and thereby correspond to these respective sections.

A.1. Hyperparameter Analysis

To give a further insight into the hyperparameter selection for our UBNA method, we show for all three considered dataset settings in Figs. 5, 6, and 7 the performance (ordinate) in dependence of the number of adaptation steps (abscissa). Here, we can already see that the adaptation consistently converges after $\kappa = 50$ adaptation steps and a further adaptation does not improve the performance anymore. Note, that Figs. 5, 6, and 7 show the adaptation until $\kappa = 100$ adaptation steps, though we do not observe any change in performance after $\kappa = 50$ adaptation steps. For

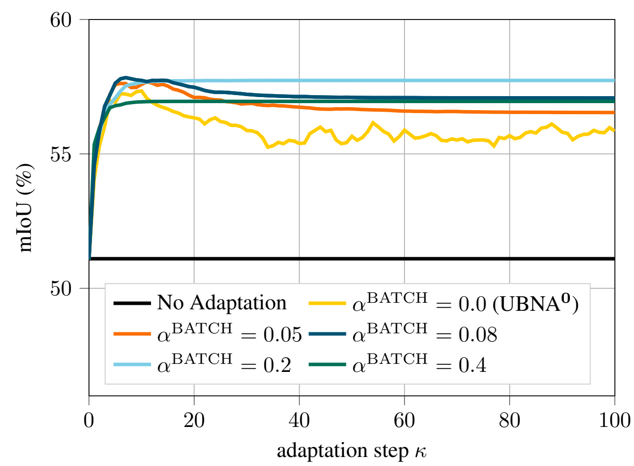

Figure 7. **Hyperparameter analysis**: Performance of **UBNA** on the **KITTI validation set** for the adaptation from **Cityscapes** ($\mathcal{D}^{\mathrm{S}}$) **to KITTI** ($\mathcal{D}^{\mathrm{T}}$) in **dependence of the batch decay factor** α^{BATCH}. Here, $\alpha^{\mathrm{LAYER}} = 0.0$ for all UBNA experiments.

$\alpha^{\mathrm{BATCH}} = 0.08$, using less adaptation steps could potentially still improve the results (cf. Fig. 5), however, this optimal point does not generalize well across different datasets (cf. Figs. 6 and 12), which is why we use $\kappa = 50$ adaptation steps in the main paper, where we observe a stable

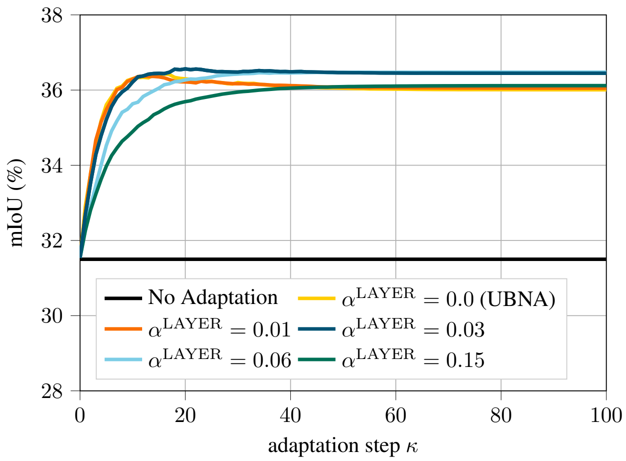

Figure 8. **Hyperparameter analysis**: Performance of **UBNA^{+}** on the **Cityscapes validation set** for the adaptation from **GTA-5 ($\mathcal{D}^{\text{S}}$) to Cityscapes ($\mathcal{D}^{\text{T}}$)** in **dependence of the layer decay factor** α^{LAYER}. Here, $\alpha^{\text{BATCH}} = 0.08$ for all **UBNA^{+}** experiments.

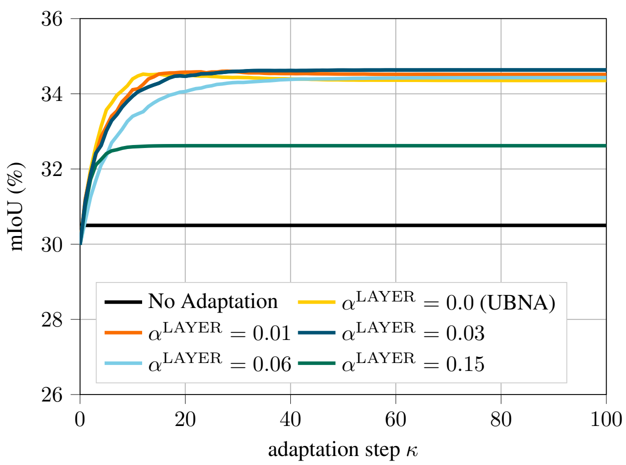

Figure 9. **Hyperparameter analysis**: Performance of **UBNA^{+}** on the **Cityscapes validation set** for the adaptation from **SYNTHIA ($\mathcal{D}^{\text{S}}$) to Cityscapes ($\mathcal{D}^{\text{T}}$)** in **dependence of the layer decay factor** α^{LAYER}. Here, $\alpha^{\text{BATCH}} = 0.08$ for all **UBNA^{+}** experiments.

convergence.

We also show the influence of using different batch-wise decay factors α^{BATCH}. Interestingly, we observe that adapting the statistics to the target domain is beneficial regardless of the used value for α^{BATCH}. However, the convergence is quite optimal for values around $\alpha^{\text{BATCH}} = 0.08$. For GTA-5 to Cityscapes (Fig. 5) and SYNTHIA to Cityscapes (Fig. 6), this yields the optimal performance, and for the Cityscapes to KITTI adaptation only a value of $\alpha^{\text{BATCH}} = 0.2$ yields a slightly better performance. Much smaller values of α^{BATCH} lead to an unstable convergence and a lower performance gain (as observed in UBNA^{0} with $\alpha^{\text{BATCH}} = 0$, yellow curves in Figs. 5, 6, and 7). On the other hand the more rapidly decreasing BN momentum for

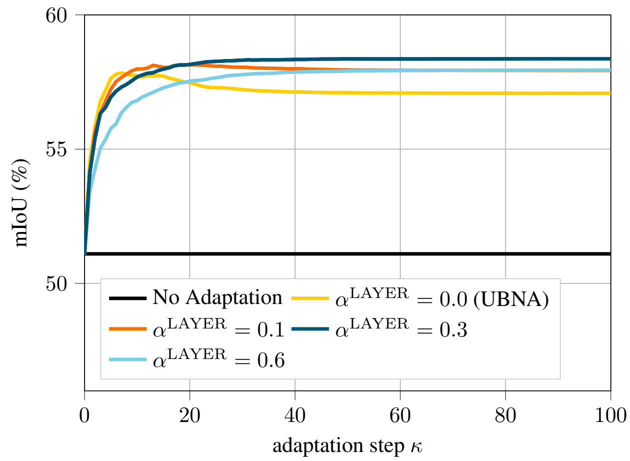

Figure 10. **Hyperparameter analysis**: Performance of **UBNA^{+}** on the **KITTI validation set** for the adaptation from **Cityscapes ($\mathcal{D}^{\text{S}}$) to KITTI ($\mathcal{D}^{\text{T}}$)** in **dependence of the layer decay factor** α^{LAYER}. Here, $\alpha^{\text{BATCH}} = 0.08$ for all **UBNA^{+}** experiments.

larger values of α^{BATCH} seem to stop the adaptation too rapidly and lead to a convergence much below the optimal point. Still, the approximate optimal hyperparameter values of $\alpha^{\text{BATCH}} = 0.08$ and $\kappa = 50$ generalize well across different dataset settings, which is essential for practical applications of our UBNA method.

Furthermore, we introduced a layer-wise weighting factor α^{LAYER}, which causes the statistics in the initial layers to be updated more rapidly than in the deeper layers. We show the performance for different values of α^{LAYER} in Figs. 8, 9, and 10. We observe that for a suitable value of α^{LAYER} this weighting again improves on top of the UBNA method ($\alpha^{\text{LAYER}} = 0$, yellow curves in Figs. 8, 9, and 10), resulting in the UBNA^{+} approach. This could hint at the fact that the domain gap can be compensated to a large extent in the initial layers, which extract domain-specific knowledge, while the deeper layers already learned more domain-invariant features that are rather task-specific. The optimal α^{LAYER} value of this method is, however, dataset-dependent. On the synthetic-to-real settings GTA-5 to Cityscapes (Fig. 8) and SYNTHIA to Cityscapes (Fig. 9) we found an optimal value of $\alpha^{\text{LAYER}} = 0.03$, while for the real-to-real setting Cityscapes to KITTI (Fig. 10) the optimal value was found to be $\alpha^{\text{LAYER}} = 0.3$. We suspect that due to the larger domain gap in the synthetic-to-real settings an adaptation only in the initial layers is not sufficient, thereby requiring a smaller value of α^{LAYER}. Conclusively, if one has access to a labelled validation set in the target domain, one can tune this hyperparameter for better performance.

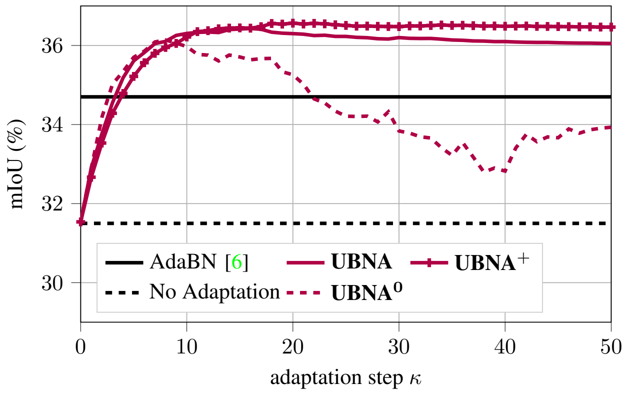

Figure 11. **Comparison to normalization approaches**: Performance on the **Cityscapes validation set** for the adaptation from **GTA-5** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) in **dependence of the adaptation step** κ (batch size $B = 6$). We also show results when using the statistics from the source domain (no adaptation) and from the target domain (AdaBN by Li *et al*. [6]).

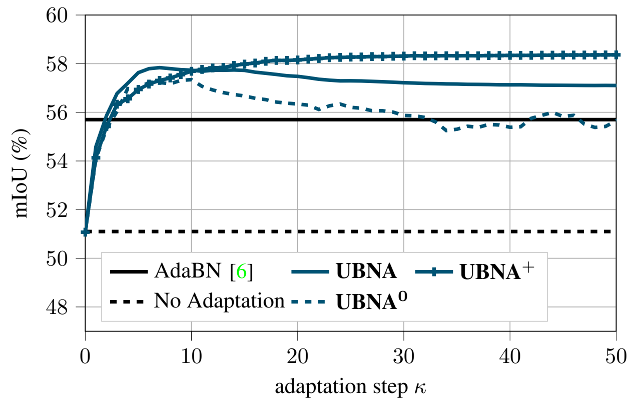

Figure 12. **Comparison to normalization approaches**: Performance on the **KITTI validation set** for the adaptation from **Cityscapes** ($\mathcal{D}^{\mathrm{S}}$) **to KITTI** ($\mathcal{D}^{\mathrm{T}}$) in **dependence of the adaptation step** κ (batch size $B = 6$). We also show results when using the statistics from the source domain (no adaptation) and from the target domain (AdaBN by Li *et al*. [6]).

A.2. Comparison to Normalization Approaches

While our main paper's comparison regarding baseline approaches was on the adaptation setting from SYNTHIA ($\mathcal{D}^{\mathrm{S}}$) to Cityscapes ($\mathcal{D}^{\mathrm{T}}$), in this part we want to show the same results for the adaptation from GTA-5 ($\mathcal{D}^{\mathrm{S}}$) to Cityscapes ($\mathcal{D}^{\mathrm{T}}$) and from Cityscapes ($\mathcal{D}^{\mathrm{S}}$) to KITTI ($\mathcal{D}^{\mathrm{T}}$) in Figures 11 and 12, respectively. Here, we also observe that the UBNA$^{\mathbf{0}}$ method has a peak after adapting with just a few batches, which outperforms the no adaptation and the AdaBN [6] baselines. Using our UBNA method allows convergence of the performance close to or even above this maximum performance, although there might still be a little potential to optimize the hyperparameter α^{BATCH} for a more stable convergence at the performance maximum. However, for the scope of this paper we rather wanted to show that our hyperparameter setting generalizes to a large degree over different dataset combinations. The more stable convergence of UBNA$^+$ also shows that with further hyperparameter tuning, the convergence behavior and maximum performance can even be improved. In conclusion, this shows that our results observed on other datasets are consistent with the results on the SYNTHIA to Cityscapes adaptation setting.

A.3. Comparison to UDA Approaches

To facilitate a more extensive comparison with respect to past and future works, we extend our comparison to UDA baseline approaches also to models using a ResNet-based network architecture, which we demonstrate in Tables 8 and 9. For our experiments we use a ResNet-50 backbone, which provides a good trade-off between computational complexity and performance. When analyzing the results we observe that UBNA improves the final model performance by 3.9% and 4.7/5.6% absolute for the adaptations from GTA-5 to Cityscapes and SYNTHIA to Cityscapes, respectively. Additionally, the performance on the single classes improves for the large majority of classes. This is in consistency with the results obtained using a VGG-16-based model. Interestingly, for ResNet-50, the UBNA$^+$ method does not give an improvement over UBNA, as the hyperparameters were optimized for the VGG-16 backbone. Accordingly, we can see that UBNA has a higher degree of generalization over different network architectures, while UBNA$^+$ is a bit more hyperparameter-sensitive.

A.4. Few-Shot Capability

To give further insight into the number of images necessary to obtain a stable adaptation using just a few images as shown in Tab. 4, in Figures 13 and 14 we show the mean value as well as the standard deviation over the results from ten experiments. We observe that the standard deviation is quite high for only one or two images per batch. The performance slightly increases for three or more images per adaptation batch. To reduce the dependency on a single image, at least three images should be chosen for adaptation, as from this number on, the standard deviation of different experiments is significantly lower and the final performance significantly higher as can be seen in Figures 13 and 14.

A.5. Ablation Studies

While the layer-wise weighting (UBNA$^+$) worked particularly well in the real-to-real adaptation setting (cf. Tab. 7), the success on settings with a larger domain gap is a little ambiguous, as can be seen in Tables 10 and 11. While the UBNA$^+$ method always improves on top of the non-adapted baseline, we observe that the performance

Table 8. **Comparison to UDA approaches**: Performance of **UBNA** and **UBNA^{+}** in **comparison to UDA methods** on the **Cityscapes validation set** for the adaptation from **GTA-5** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$). We evaluate different models (first column), compare different methods (second column, values from the respective papers), show if they use labeled source data during unsupervised domain adaptation (third column), evaluate the IoU performance (%) on the single semantic classes, and finally give an mIoU (%) over 19 classes. For ResNet results we use the **ResNet-50** topology, as this yielded better results at lower computational complexity compared to the ResNet-101 often used in UDA approaches. Best results for UDA and for UDA without source data are printed in boldface.

Network	Method	Source data during adaptation	road	sidewalk	building	wall**	fence**	pole**	traffic light	traffic sign	vegetation	terrain*	sky	person	rider	car	truck*	bus	on rails*	motorbike	bike	mIoU (%) (19 classes)
ResNet-50	Du *et al.* [3]	yes	90.3	38.9	81.7	24.8	22.9	30.5	37.0	21.2	84.8	38.8	76.9	58.8	30.7	85.7	30.6	38.1	5.9	28.3	36.9	45.4
	Vu *et al.* [9]	yes	89.4	33.1	81.0	26.6	26.8	27.2	33.5	24.7	83.9	36.7	78.8	58.7	30.5	84.8	38.5	44.5	1.7	31.6	32.4	45.5
	Li *et al.* [7]	yes	91.0	44.7	84.2	34.6	27.6	30.2	36.0	36.0	85.0	**43.6**	83.0	58.6	31.6	83.3	35.3	49.7	3.3	28.8	35.6	48.5
	Dong *et al.* [2]	yes	89.6	50.4	83.0	35.6	26.9	31.1	37.3	35.1	83.5	40.6	84.0	60.6	34.3	80.9	35.1	47.3	0.5	34.5	33.7	48.6
	Wang *et al.* [10]	yes	90.6	44.7	84.8	34.3	28.7	31.6	35.0	37.6	84.7	43.3	85.3	57.0	31.5	83.8	42.6	48.5	1.9	30.4	39.0	49.2
	Yang *et al.* [11]	yes	90.8	41.4	84.7	35.1	27.5	31.2	38.0	32.8	**85.6**	42.1	84.9	59.6	34.4	85.0	**42.8**	52.7	3.4	30.9	38.1	49.5
	Zhang *et al.* [13]	yes	90.4	51.6	83.8	34.2	27.8	**38.4**	25.3	**48.4**	85.4	38.2	78.1	58.6	**34.6**	84.7	21.9	42.7	**41.1**	29.3	37.2	50.2
	Kim *et al.* [4]	yes	92.9	55.0	85.3	34.2	**31.1**	34.9	40.7	34.0	85.2	40.1	87.1	61.0	31.1	82.5	32.3	42.9	0.3	**36.4**	46.1	50.2
	Yang *et al.* [12]	yes	92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	**53.1**	16.9	27.7	**46.4**	50.5
	Mei *et al.* [8]	yes	**94.1**	**58.8**	**85.4**	**39.7**	29.2	25.1	**43.1**	34.2	84.8	34.6	**88.7**	**62.7**	30.3	**87.6**	42.3	50.3	24.7	35.2	40.2	**52.2**
	No adaptation	-	58.1	23.8	70.5	14.8	19.2	30.5	29.0	17.7	**79.1**	**21.8**	**83.1**	56.4	**14.8**	72.3	**19.5**	4.5	0.9	16.5	5.6	33.6
	UBNA	no	**81.8**	**32.3**	**79.5**	**18.2**	**23.8**	**34.9**	**29.5**	**19.8**	74.2	17.9	82.4	**57.5**	11.1	**81.6**	16.1	**19.0**	**2.5**	**21.3**	**9.8**	**37.5**
	UBNA^{+}	no	70.6	25.8	78.5	17.7	23.7	34.2	28.9	19.0	77.8	19.6	82.6	57.4	11.5	81.5	16.9	17.1	1.0	20.6	9.1	36.5
VGG-16	Vu *et al.* [9]	yes	86.9	28.7	78.7	28.5	25.2	17.1	20.3	10.9	80.0	26.4	70.2	47.1	8.4	81.5	26.0	17.2	18.9	11.7	1.6	36.1
	Du *et al.* [3]	yes	88.7	32.1	79.5	29.9	22.0	23.8	21.7	10.7	80.8	29.8	72.5	49.5	16.1	82.1	23.2	18.1	3.5	24.4	8.1	37.7
	Li *et al.* [7]	yes	89.2	40.9	81.2	29.1	19.2	14.2	29.0	19.6	83.7	35.9	80.7	54.7	23.3	82.7	25.8	28.0	2.3	25.7	19.9	41.3
	Dong *et al.* [2]	yes	89.8	46.1	75.2	30.1	**27.9**	15.0	20.4	18.9	82.6	39.1	77.6	47.8	17.4	76.2	28.5	33.4	0.5	29.4	**30.8**	41.4
	Yang *et al.* [12]	yes	86.1	35.1	80.6	30.8	20.4	27.5	30.0	26.0	82.1	30.3	73.6	52.5	21.7	81.7	24.0	30.5	**29.9**	14.6	24.0	42.2
	Kim *et al.* [4]	yes	**92.5**	**54.5**	**83.9**	**34.5**	25.5	31.0	30.4	18.0	**84.1**	**39.6**	**83.9**	53.6	19.3	81.7	21.1	13.6	17.7	12.3	6.5	42.3
	Wang *et al.* [10]	yes	88.1	35.8	83.1	25.8	23.9	29.2	28.8	28.6	83.0	36.7	82.3	53.7	22.8	82.3	26.4	**38.6**	0.0	19.6	17.1	42.4
	Choi *et al.* [1]	yes	90.2	51.5	81.1	15.0	10.7	**37.5**	**35.2**	**28.9**	**84.1**	32.7	75.9	**62.7**	19.9	82.6	22.9	28.3	0.0	23.0	25.4	42.5
	Yang *et al.* [11]	yes	90.1	41.2	82.2	30.3	21.3	18.3	33.5	23.0	**84.1**	37.5	81.4	54.2	**24.3**	**83.0**	**27.6**	32.0	8.1	**29.7**	26.9	**43.6**
	No adaptation	-	55.8	21.9	65.9	15.2	14.7	27.5	**31.0**	17.9	**77.8**	**19.5**	74.4	55.2	12.1	71.7	11.9	3.3	0.5	13.2	9.6	31.5
	UBNA	no	**80.8**	29.4	77.6	19.8	**17.1**	**33.9**	29.3	20.5	73.9	16.8	76.7	58.3	**15.2**	79.1	13.6	12.5	5.7	14.1	**10.8**	36.1
	UBNA^{+}	no	79.9	**29.9**	**78.1**	**21.1**	16.5	33.8	29.7	**20.6**	75.6	18.4	**78.0**	**58.4**	14.6	**79.4**	**14.8**	**13.0**	**5.8**	**14.6**	10.6	**36.5**

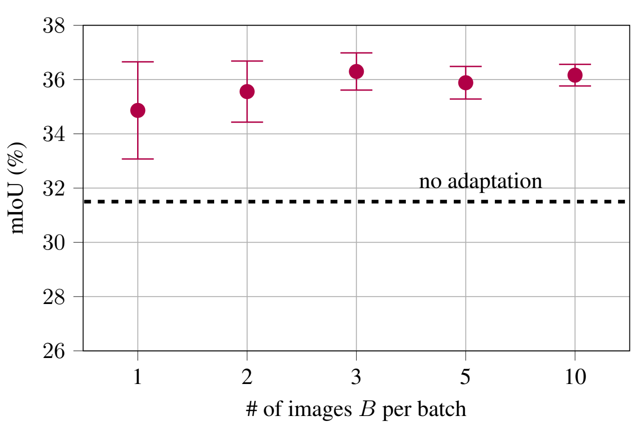

Figure 13. **Few-shot adaptation**: Performance of **UBNA^{+}** on the **Cityscapes validation set** for the adaptation from **GTA-5** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) for **different numbers of total images used for adaptation**. We show the mean and the standard deviation over 10 different experiments after the $\kappa = 50$th step, whereby in each experiment all batches contain the same image(s).

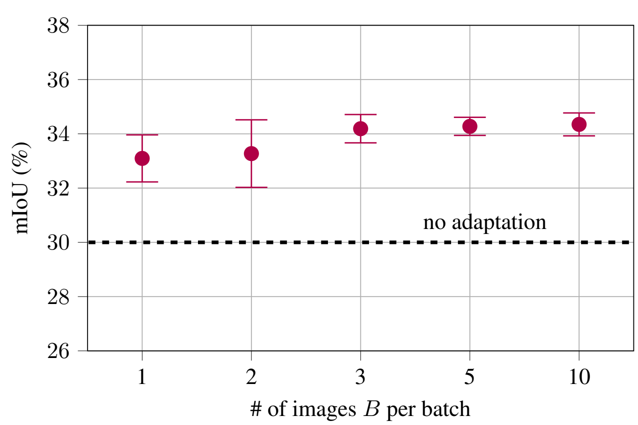

Figure 14. **Few-shot adaptation**: Performance of **UBNA^{+}** on the **Cityscapes validation set** for the adaptation from **SYNTHIA** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) for **different numbers of total images used for adaptation**. We show the mean and the standard deviation over 10 different experiments after the $\kappa = 50$th step, whereby in each experiment all batches contain the same image(s).

gains over the UBNA method are rather small and in some cases, we even observe a better performance with UBNA (*e.g.*, the last three rows in Tab. 10). Also, we had to reduce the layer-wise weighting factor by a factor of 10 to a value

Table 9. **Comparison to UDA approaches**: Performance of **UBNA** and **UBNA⁺** in **comparison to UDA methods** on the **Cityscapes validation set** for the adaptation from **SYNTHIA** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$). We evaluate different models (first column), compare different methods (second columns, values from the respective papers), show if they use labeled source data during unsupervised domain adaptation (third column), evaluate the IoU performance (%) on the single semantic classes, and finally give an mIoU (%) over 13 classes and over 16 classes (the latter including also wall, fence and pole) as in [5, 9]. For ResNet results we use the **ResNet-50** topology, as this yielded better results at lower computational complexity compared to the ResNet-101 commonly used in other approaches. Best UDA results and best UDA without source data results are printed in boldface.

Network	Method	Source data during adaptation	road	sidewalk	building	wall**	fence**	pole**	traffic light	traffic sign	vegetation	terrain*	sky	person	rider	car	truck*	bus	on rails*	motorbike	bike	mIoU (%) (16 classes)	mIoU (%) (13 classes)
ResNet-50	Vu *et al.* [9]	yes	85.6	42.2	79.7	8.7	0.4	25.9	5.4	8.1	80.4	-	84.1	57.9	23.8	73.3	-	36.4	-	14.2	33.0	41.2	48.0
	Du *et al.* [3]	yes	84.6	41.7	80.8	-	-	-	11.5	14.7	80.8	-	**85.3**	57.5	21.6	82.0	-	36.0	-	19.3	34.5	-	50.0
	Li *et al.* [7]	yes	**86.0**	**46.7**	80.3	-	-	-	14.1	11.6	79.2	-	81.3	54.1	27.9	73.7	-	**42.2**	-	25.7	45.3	-	51.4
	Wang *et al.* [10]	yes	83.0	44.0	80.3	-	-	-	17.1	15.8	80.5	-	81.8	59.9	**33.1**	70.2	-	37.3	-	28.5	45.8	-	52.1
	Yang *et al.* [11]	yes	85.1	44.5	81.0	-	-	-	16.4	15.2	80.1	-	84.8	59.4	31.9	73.2	-	41.0	-	32.6	44.7	-	53.1
	Dong *et al.* [2]	yes	80.2	41.1	78.9	**23.6**	0.6	31.0	27.1	**29.5**	82.5	-	83.2	62.1	26.8	81.5	-	37.2	-	27.3	42.9	47.2	-
	Mei *et al.* [8]	yes	81.9	41.5	**83.3**	17.7	**4.6**	**32.3**	**30.9**	28.8	**83.4**	-	85.0	**65.5**	30.8	**86.5**	-	38.2	-	**33.1**	**52.7**	**49.8**	**57.0**
	No adaptation	-	36.5	18.6	68.3	2.0	0.2	30.3	6.0	10.2	74.5	-	81.6	**51.9**	10.6	41.3	-	9.5	-	2.2	22.6	29.1	34.1
	UBNA	no	**62.5**	**22.8**	**75.6**	3.1	**0.5**	32.5	**8.6**	**11.3**	73.0	-	**82.7**	42.5	12.5	**67.1**	-	12.5	-	**5.7**	27.8	**33.8**	**39.7**
	UBNA⁺	no	57.4	22.3	75.0	**3.6**	0.4	**33.8**	8.5	11.1	**76.2**	-	82.6	46.8	**12.8**	61.2	-	**12.8**	-	4.8	**28.5**	33.6	39.4
VGG-16	Vu *et al.* [9]	yes	67.9	29.4	71.9	6.3	0.3	19.9	0.6	2.6	-	74.9	74.9	35.4	9.6	67.8	-	21.4	-	4.1	15.5	31.4	36.6
	Lee *et al.* [5]	yes	71.1	29.8	71.4	3.7	0.3	**33.2**	6.4	15.6	**81.2**	-	78.9	52.7	13.1	75.9	-	25.5	-	10.0	20.5	36.8	**42.4**
	Dong *et al.* [2]	yes	70.9	**30.5**	**77.8**	**9.0**	**0.6**	27.3	8.8	12.9	74.8	-	81.1	43.0	**25.1**	73.4	-	**34.5**	-	**19.5**	38.2	39.2	-
	Yang *et al.* [11]	yes	**73.7**	29.6	77.6	1.0	0.4	26.0	**14.7**	**26.6**	80.6	-	**81.8**	**57.2**	24.5	**76.1**	-	27.6	-	13.6	**46.6**	**41.1**	-
	No adaptation	-	49.4	20.8	61.5	**3.6**	0.1	30.5	**13.6**	14.1	74.4	-	75.5	**53.5**	10.6	47.2	-	4.8	-	3.0	17.1	30.0	35.2
	UBNA	no	**72.3**	26.6	**73.0**	2.3	**0.3**	31.5	12.1	16.6	72.1	-	**75.6**	45.4	**13.6**	61.2	-	**8.5**	-	**8.5**	**30.1**	34.4	40.7
	UBNA⁺	no	71.5	**27.3**	72.9	2.5	**0.3**	**32.0**	12.7	**16.7**	**74.6**	-	75.4	47.1	**13.6**	**61.4**	-	**8.5**	-	8.3	29.2	**34.6**	**41.0**

Table 10. **Network topology ablation**: Performance of **UBNA⁰**, **UBNA**, and **UBNA⁺** on the **Cityscapes validation set** for the adaptation from **GTA-5** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) for **various network topologies**; best results for each network topology in boldface; mIoU values in %; batch size $B = 6$; $\alpha^{\mathrm{BATCH}} = 0.08$ (**UBNA**, **UBNA⁺**); $\alpha^{\mathrm{LAYER}} = 0.03$ (**UBNA⁺**).

Network	No adaptation	UBNA⁰	UBNA	UBNA⁺
VGG-16	31.5	33.9	36.1	**36.5**
VGG-19	31.2	35.3	37.4	**37.5**
ResNet-18	34.1	29.4	34.8	**35.5**
ResNet-34	33.6	34.3	**36.0**	35.5
ResNet-50	33.6	35.6	**37.5**	36.5
ResNet-101	30.0	33.4	**34.6**	32.6

Table 11. **Network topology ablation**: Performance of **UBNA⁰**, **UBNA**, and **UBNA⁺** on the **Cityscapes validation set** for the adaptation from **SYNTHIA** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$) for **various network topologies**; best results for each network topology in boldface; mIoU values in %; batch size $B = 6$; $\alpha^{\mathrm{BATCH}} = 0.08$ (**UBNA**, **UBNA⁺**); $\alpha^{\mathrm{LAYER}} = 0.03$ (**UBNA⁺**).

Network	No adaptation	UBNA⁰	UBNA	UBNA⁺
VGG-16	30.0	32.6	34.4	**34.6**
VGG-19	30.9	29.1	30.8	**31.7**
ResNet-18	28.6	31.3	**31.9**	31.7
ResNet-34	27.4	30.1	30.0	**30.3**
ResNet-50	29.1	32.9	**33.8**	33.6
ResNet-101	30.8	30.0	31.8	**34.3**

$\alpha^{\mathrm{LAYER}} = 0.03$ to obtain decent results, which shows the higher hyperparameter dependency compared to the plain UBNA method. Therefore, on a new setting we would recommend to start with the UBNA method and (if possible) tune afterwards with the layer-wise weighting characterizing our UBNA⁺ method.

B. Additional Qualitative Results

To emphasize also the qualitative effects our method has on the results, we show images of the adaptations from GTA-5 ($\mathcal{D}^{\mathrm{S}}$) to Cityscapes ($\mathcal{D}^{\mathrm{T}}$), SYNTHIA ($\mathcal{D}^{\mathrm{S}}$) to Cityscapes ($\mathcal{D}^{\mathrm{T}}$), and Cityscapes ($\mathcal{D}^{\mathrm{S}}$) to KITTI ($\mathcal{D}^{\mathrm{T}}$) in Figures 15, 16, and 17, respectively. In these figures, we show the input images, the ground truth segmentation mask, the predicted segmentation mask when using the model without adaptation, and the results from the adapted model using the UBNA method, from left to right. Interestingly, the qualitative effects are quite similar for all three dataset setups as already observed in the class-wise improvements in Sec. A.3. First of all, the overall artifacts present in different parts of the image are significantly reduced, which is especially well observable on the synthetic-to-real adaptation (*e.g.*, Fig. 15, row 1, or Fig. 16, row 4). However, even in the real-to-real adaptation setting, where the baseline approach already is significantly better, we observe that artifacts are removed by UBNA⁺ from the image, as visible in the last three rows of Fig. 17, where the human-class artifact in the center of the image is removed. Also, in all three dataset setups, the detection of the road class is significantly improved (*e.g*. Fig. 15, rows 4 and 5). Also, cars are better distinguished from their surrounding and less wrongfully predicted cars are present in the semantic segmentation prediction (*e.g*. Fig. 16, rows 6 and 7). Nevertheless, there are also some classes that are not detected as good as before (cf. Tables 8 and 9) which is also visible qualitatively, *e.g*., the vegetation class contains slightly more artifacts than before the adaptation (cf. Fig. 16, row 2).

References

[1] Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-Ensembling With GAN-Based Data Augmentation for Domain Adaptation in Semantic Segmentation. In *Proc. of ICCV*, pages 6830–6840, Seoul, Korea, Oct. 2019. 4

[2] Jiahua Dong, Yang Cong, Gan Sun, Yuyang Liu, and Xiaowei Xu. CSCL: Critical Semantic-Consistent Learning for Unsupervised Domain Adaptation. In *Proc. of ECCV*, pages 745–762, Glasgow, UK, Aug. 2020. 4, 5

[3] Liang Du, Jingang Tan, Hongye Yang, Jianfeng Feng, Xiangyang Xue, Qibao Zheng, Xiaoqing Ye, and Xiaolin Zhang. SSF-DAN: Separated Semantic Feature Based Domain Adaptation Network for Semantic Segmentation. In *Proc. of ICCV*, pages 982–991, Seoul, Korea, Oct. 2019. 4, 5

[4] Myeongjin Kim and Hyeran Byun. Learning Texture Invariant Representation for Domain Adaptation of Semantic Segmentation. In *Proc. of CVPR*, pages 12975–12984, Seattle, WA, USA, June 2020. 4

[5] Kuan-Hui Lee, German Ros, Jie Li, and Adrien Gaidon. SPI-GAN: Privileged Adversarial Learning from Simulation. In *Proc. of ICLR*, pages 1–14, New Orleans, LA, USA, Apr. 2019. 5

[6] Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. Adaptive Batch Normalization for Practical Domain Adaptation. *Pattern Recognition*, 80:109–117, Aug. 2018. 3

[7] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional Learning for Domain Adaptation of Semantic Segmentation. In *Proc. of CVPR*, pages 6936–6945, Long Beach, CA, USA, June 2019. 4, 5

[8] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance Adaptive Self-Training for Unsupervised Domain Adaptation. In *Proc. of ECCV*, pages 415–430, Glasgow, UK, Aug. 2020. 4, 5

[9] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Perez. ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. In *Proc. of CVPR*, pages 2517–2526, Long Beach, CA, USA, June 2019. 4, 5

[10] Zhonghao Wang, Mo Yu, Yunchao Wei, Rogerio Feris, Jinjun Xiong, Wen-Mei Hwu, Thomas S. Huang, and Honghui Shi. Differential Treatment for Stuff and Things: A Simple Unsupervised Domain Adaptation Method for Semantic Segmentation. In *Proc. of CVPR*, pages 12635–12644, Seattle, WA, USA, June 2020. 4, 5

[11] Jinyu Yang, Weizhi An, Sheng Wang, Xinliang Zhu, Chaochao Yan, and Junzhou Huang. Label-Driven Reconstruction for Domain Adaptation in Semantic Segmentation. In *Proc. of ECCV*, pages 480–498, Glasgow, UK, Aug. 2020. 4, 5

[12] Yanchao Yang and Stefano Soatto. FDA: Fourier Domain Adaptation for Semantic Segmentation. In *Proc. of CVPR*, pages 4085–4095, Seattle, WA, USA, June 2020. 4

[13] Qiming Zhang, Jing Zhang, Wei Liu, and Dacheng Tao. Category Anchor-Guided Unsupervised Domain Adaptation for Semantic Segmentation. In *Proc. of NeurIPS*, pages 433–443, Vancouver, Canada, Dec. 2019. 4

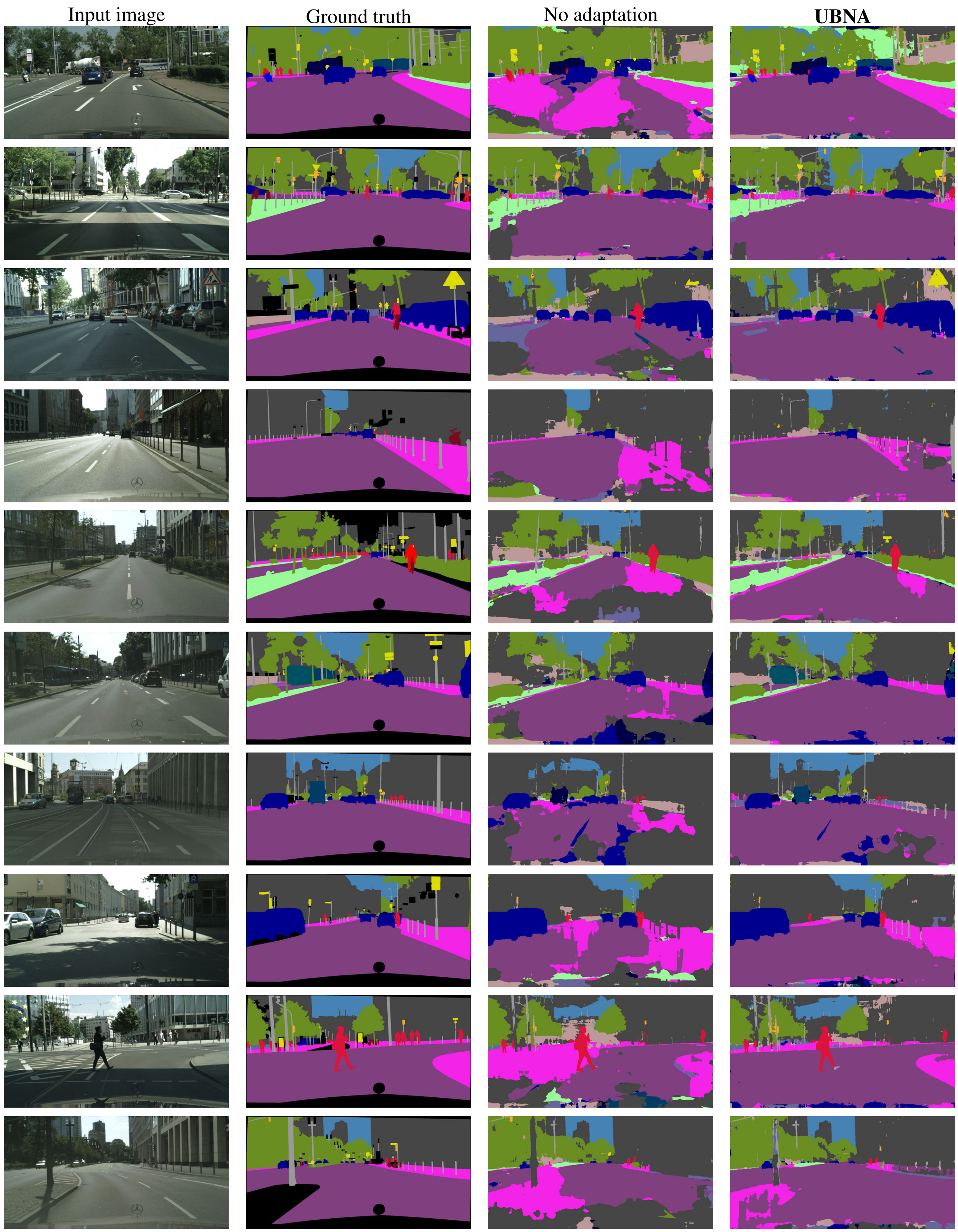

Figure 15. **Qualitative comparison**: Qualitative results of **UBNA** in **comparison to no adaptation and the ground truth** on the **Cityscapes validation set** for the adaptation from **GTA-5** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$). The figure is best viewed on screen and in color.

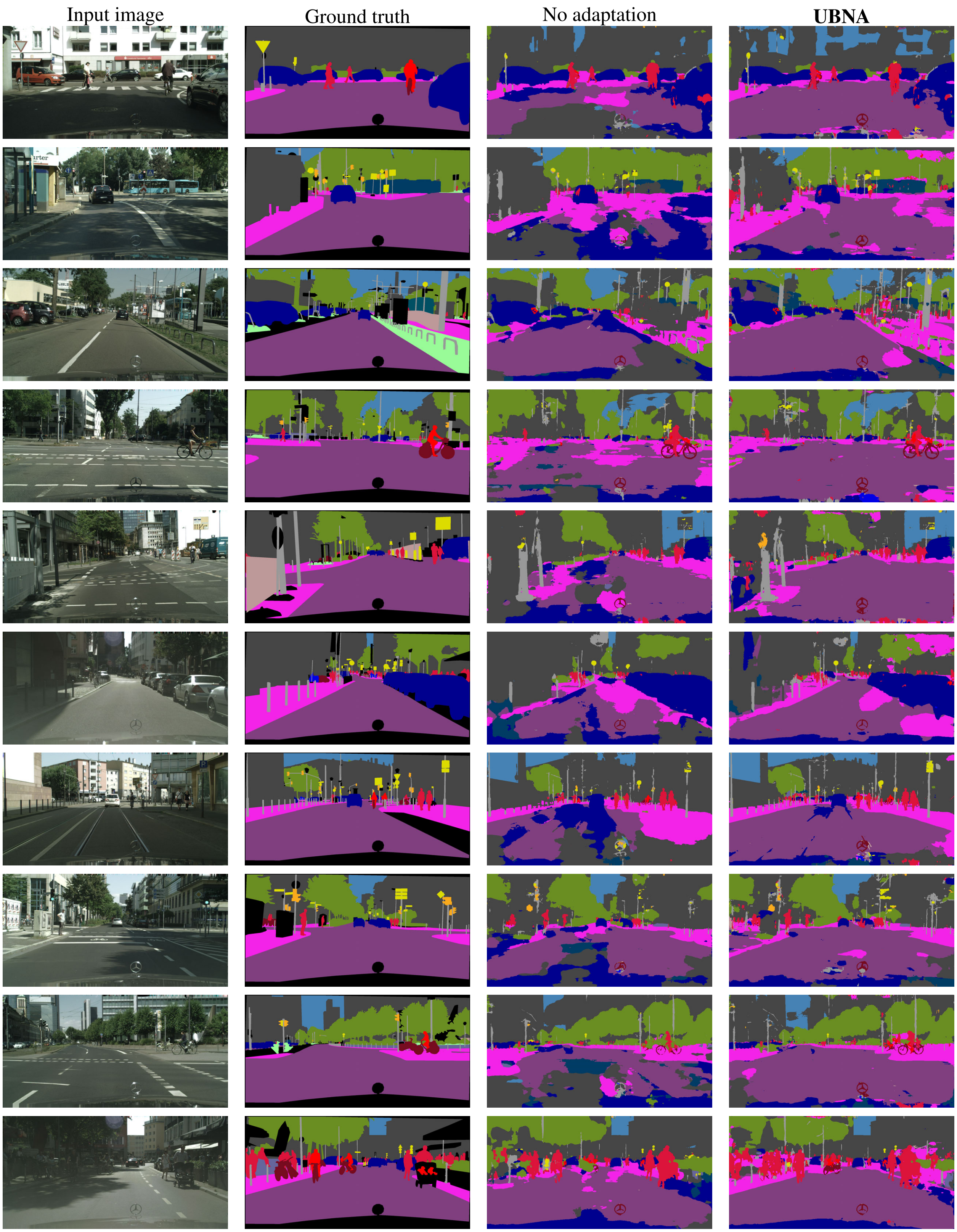

Figure 16. **Qualitative comparison**: Qualitative results of **UBNA** in **comparison to no adaptation and the ground truth** on the **Cityscapes validation set** for the adaptation from **SYNTHIA** ($\mathcal{D}^{\mathrm{S}}$) **to Cityscapes** ($\mathcal{D}^{\mathrm{T}}$). The figure is best viewed on screen and in color.

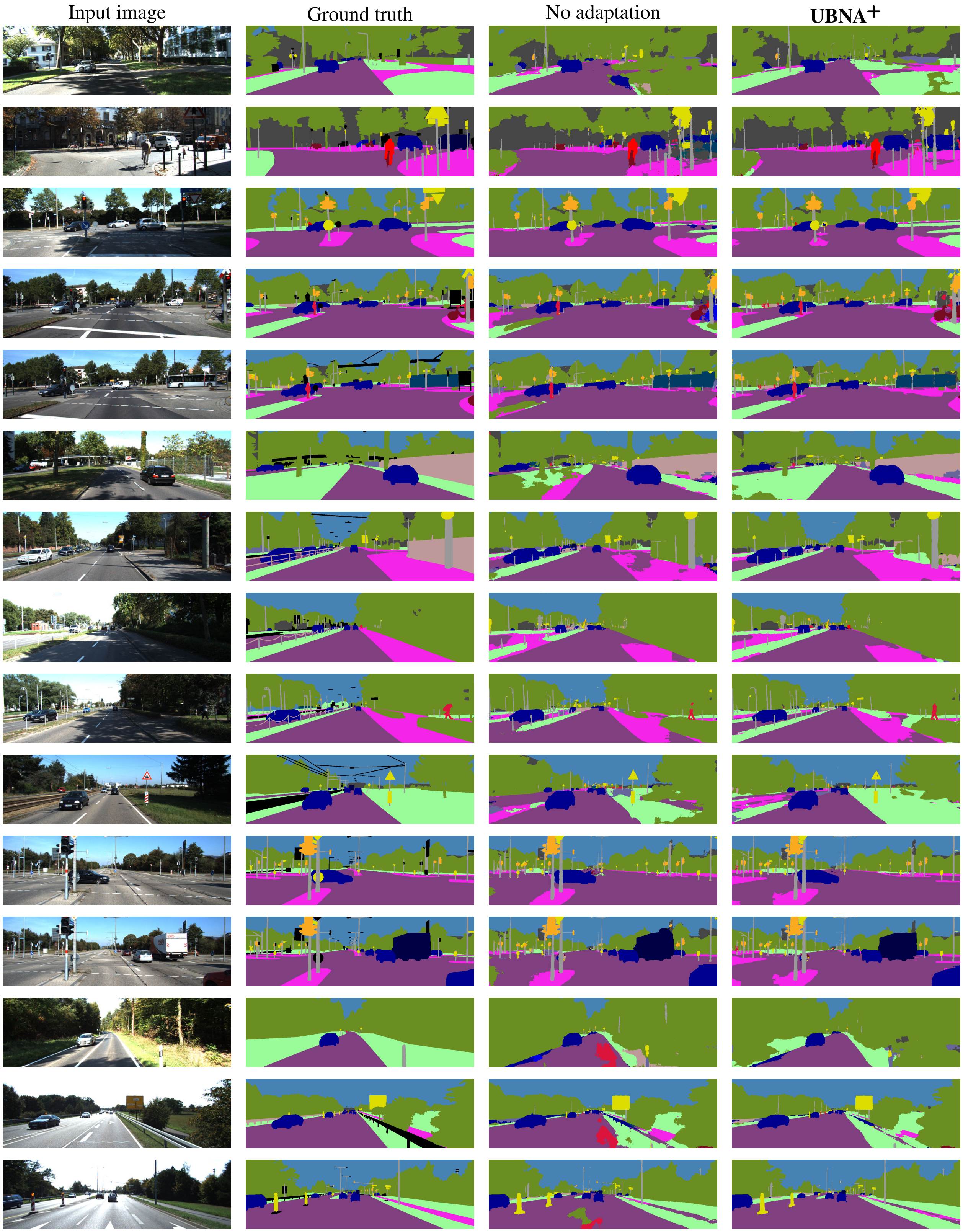

Figure 17. **Qualitative comparison**: Qualitative results of **UBNA**$^{+}$ ($\alpha^{\mathrm{LAYER}} = 0.3$) in **comparison to no adaptation and the ground truth** on the **KITTI validation set** for the adaptation from **Cityscapes** ($\mathcal{D}^{\mathrm{S}}$) **to KITTI** ($\mathcal{D}^{\mathrm{T}}$). The figure is best viewed on screen and in color.