NightAdapter: Learning a Frequency Adapter for Generalizable Night-time Scene Segmentation

Supplementary Material

In this supplementary material, first a per-category performance between the proposed NightAdapter and existing methods is compared (in Sec. 7). Then, additional hyperparameter analysis and their impact on the overall performance is discussed (in Sec. 8). Next, the effectiveness of NightAdapter is validated on various vision foundation models in Sec. 9. More ablation studies and more visual results are provided in Sec. 10 and Sec. 11. Finally, limitation and future work is discussed in Sec. 12.

7. Per-category Performance Comparison

Table 7 reports the results under the in-domain night-time segmentation setting. NightAdapter significantly outperforms the state-of-the-art DTP [70] on most of the 19 semantic categories. Besides, it also clearly outperforms the REIN baseline [71] on most of these semantic categories.

8. Hyperparameter Analysis

Impact of Randomization Threshold T. Table 8 studies the impact of randomization threshold T on unseen night-time domain performance. By default, T is set to be 0.3, and we test the situation when it varies from 0.1 to 0.9, under a range of 0.2. It is observed that the performance on unseen night-time domain is relatively stable when T is set a relatively small value. However, when T is relatively large, a high response on the illumination-sensitive bands may introduce too much noise and negatively impact the discriminative ability of the VFM feature.

Impact of Token Length m. Table 9 analyzes how the token length m impacts the performance on unseen night-time domains. By default we set m to be 100, and we also report the results when m is 50, 75, 125 and 150, respectively. The results indicate that 100 may be the optimal setting. A too-small or too-large m may under-fit and over-fit the representation, and lead to a slight performance decline.

Impact of Rank r. Following the baseline model REIN [71], in each adaptation module, the token sequence T is composed as two low-rank matrices to significantly reduce the number of parameter. Table 10 analyzes how the low-rank dimension r impacts the performance on unseen night-time domains. By default we set r to be 16, and we also report the results when r is 4, 8, 32 and 64, respectively. The results indicate that 16 may be the optimal setting. A too-small or too-large r may under-fit and over-fit the representation, and lead to a slight performance decline.

9. Feasibility on Different VFMs

In the main text, the proposed NightAdapter is validated on the pre-trained image encoder of DINOv2 [53]. We further validate its effectiveness on other pre-trained image encoders from various vision foundation models, namely, CLIP, MAE, SAM and EVA02. REIN [71] is used as the baseline. Only the pre-trained image encoder was changed, while the rest components and hyper-parameter settings keep the same. The results are shown in Table 11. The proposed NightAdapter shows a significant performance improvement (*i.e.*, at least 1% mIoU) than the REIN baseline, when using all types of the pre-trained image encoders. These outcomes indicate the effectiveness of the proposed NightAdapter and the discrete sine prior to understand unseen night-time scenes.

10. Ablation Studies on Frequency Band

We conduct a more detailed ablation on the impact of each of the eight frequency bands, compared with the REIN baseline. We follow the same band rejection analysis in the Preliminary section, but conduct the experiments on domain generalization in semantic segmentation. The results in Table 12 show that the contribution of the last two frequency bands $\mathcal{V}_i^{[768,896)}$ and $\mathcal{V}_i^{[896,1024)}$ plays an more important role to day-night and night-to-night generalization, which further indicates that both bands are illumination-insensitive.

The activation pattern of each frequency band is visualized in Fig. 8. We extract feature maps before the decoder to reveal the activation pattern on unseen night-time images. The high/middle frequency bands exhibit dispersed activations across the scene, with strong responses in background such as the sky, and are sensitive to illumination. Low frequency bands focus on key objects in the scene and are less affected by illumination.

To further validate the effectiveness of our design to leverage the frequency bands, Fig. 9 compares the activation patterns on two in-the-wild night images. Our NightAdapter shows more stable activation on key objects.

11. More Visual Results

In addition to Fig. 6 and Fig. 7 in the main text, in this supplementary material, more visual results are provided in Fig. 10 and Fig. 11. NightAdapter shows more complete and precise predictions than the state-of-the-art domain

Method	road	sidew.	build.	wall	fense	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
UPer-Swin [49]	92.7	57.0	85.6	60.0	56.0	39.1	46.6	64.7	65.1	26.0	89.5	63.7	42.8	86.8	67.8	78.1	62.0	40.7	52.5	61.1
+SOD [70]	93.2	59.1	86.8	54.0	56.9	41.0	46.4	64.1	65.9	28.4	90.6	66.8	45.5	87.4	77.0	79.6	62.2	44.9	51.0	63.7
DTP [70]	93.3	59.3	86.4	53.5	56.0	41.4	51.3	68.8	66.3	29.3	90.8	68.7	46.7	89.8	80.8	81.9	63.8	50.1	53.3	64.2
REIN [71]	93.7	60.1	87.7	55.5	58.5	42.3	50.7	70.8	66.8	29.1	91.4	65.7	46.6	89.5	80.8	82.7	61.2	45.0	49.1	64.6
NightAdapter	**95.2**	**62.8**	**89.4**	**57.9**	**60.0**	**43.9**	**52.2**	**71.3**	**68.4**	**32.9**	**92.5**	**67.1**	**46.6**	**90.7**	**83.9**	**84.6**	**62.9**	**45.4**	**49.9**	**66.2**

Table 7. Per-category performance under the in-domain night-time segmentation. Experiments are conducted on NightCity-fine [70]. Metric mIoU in percentage %. Top three results are highlighted as **best**, second and third, respectively.

Figure 8. Activation pattern of each frequency band. Zoom in for better view.

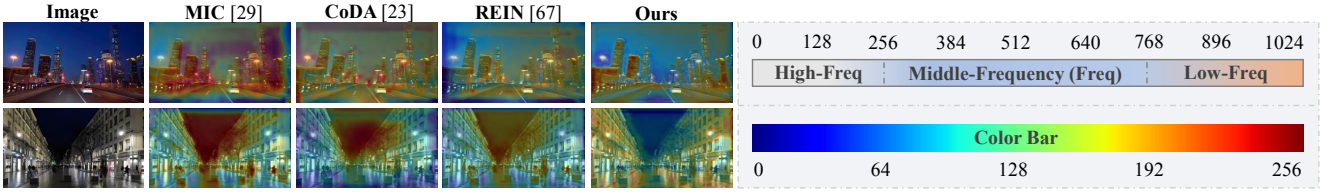

Figure 9. Activation pattern on in-the-wild images. Zoom in for better view.

T value	AN	DZ	ND
0.1	83.2	73.4	59.2
0.3	**84.0**	**74.1**	**60.3**
0.5	83.1	73.0	58.9
0.7	81.8	72.3	57.6
0.9	79.5	71.6	55.8

Table 8. Impact of the randomization threshold T on unseen night-time domain performance. mIoU in percentage (%).

m size	Trained on CityScapes		
	AN	DZ	ND
50	82.7	72.9	58.8
75	83.2	73.4	59.5
100	**84.0**	**74.1**	**60.3**
125	83.8	73.7	59.9
150	82.9	72.5	58.6

Table 9. Impact of the token length m on unseen night-time domain performance. mIoU in percentage (%).

Method	Trained on Cityscapes		
	AN	DZ	ND
4	82.9	73.2	59.4
8	83.4	73.6	59.9
16	**84.0**	**74.1**	**60.3**
32	84.2	**74.0**	60.1
64	83.7	73.8	59.5

Table 10. Impact of the rank dimension r on unseen night-time domain performance. Evaluation metric mIoU in %.

Method	Backbone	NC as Source				
		AN	DZ	ND	BN	AC
REIN Baseline [71]	CLIP	34.0	32.8	46.1	28.5	32.3
NightAdapter (Ours)		**41.9**	**38.6**	**53.4**	**34.5**	**38.7**
REIN Baseline [71]	MAE	33.6	33.1	46.5	29.0	31.8
NightAdapter (Ours)		**42.0**	**39.2**	**53.7**	**35.4**	**37.6**
REIN Baseline [71]	SAM	38.3	37.5	49.8	33.6	36.2
NightAdapter (Ours)		**45.9**	**42.6**	**54.4**	**38.0**	**41.3**
REIN Baseline [71]	EVA02	42.1	40.9	57.5	33.6	40.2
NightAdapter (Ours)		**48.7**	**45.3**	**60.8**	**39.1**	**45.4**
REIN Baseline [71]	DINOv2	43.6	42.4	59.7	34.8	41.1
NightAdapter (Ours)		**50.5**	**47.0**	**62.3**	**40.5**	**46.9**

Table 11. Night-to-night generalization performance on pre-trained image encoders from different vision foundation models (VFM).

adaptation methods and domain generalization methods, indicating its effectiveness on unseen night-time scenes.

12. Limitation Discussion & Future Work

The limitation of the proposed `NightAdapter` is twofold. Firstly, the training set of BDD-Night only has hundreds of images, which are small in amount and are more likely to cause the over-fit problem for a foundation model. Secondly, the property of each frequency band is only explored on the night images. In other adverse conditions such as fog and rain, different frequency bands may hold other properties.

The proposed method is attached on the pre-trained encoder of a vision foundation model, which can be attached to the decoders tailored for other tasks. We will explore its task adaptability on night-time object detection, classification and

#	Setting	Full	[0,128)	[128,256)	[256,384)	[384,512)	[512,640)	[640,768)	[768,896)	[896, 1024)
1	Day	**81.3**	81.1	80.8	81.0	80.5	79.3	78.6	77.3	76.6
2	Day-Night	68.9	69.1	**69.3**	68.3	67.6	67.7	66.9	59.6	58.5
3	Night	60.5	**61.0**	60.6	60.0	59.8	59.3	57.9	52.4	51.8
4	Night-Night	42.1	**42.2**	41.8	42.0	41.6	41.5	40.1	34.2	32.9
5	NightAdapter	**84.0**	83.8	83.5	83.7	83.4	83.2	82.9	79.2	78.4

Table 12. Band rejection test on day-to-night and night-to-night semantic segmentation. Metric mIoU (%).

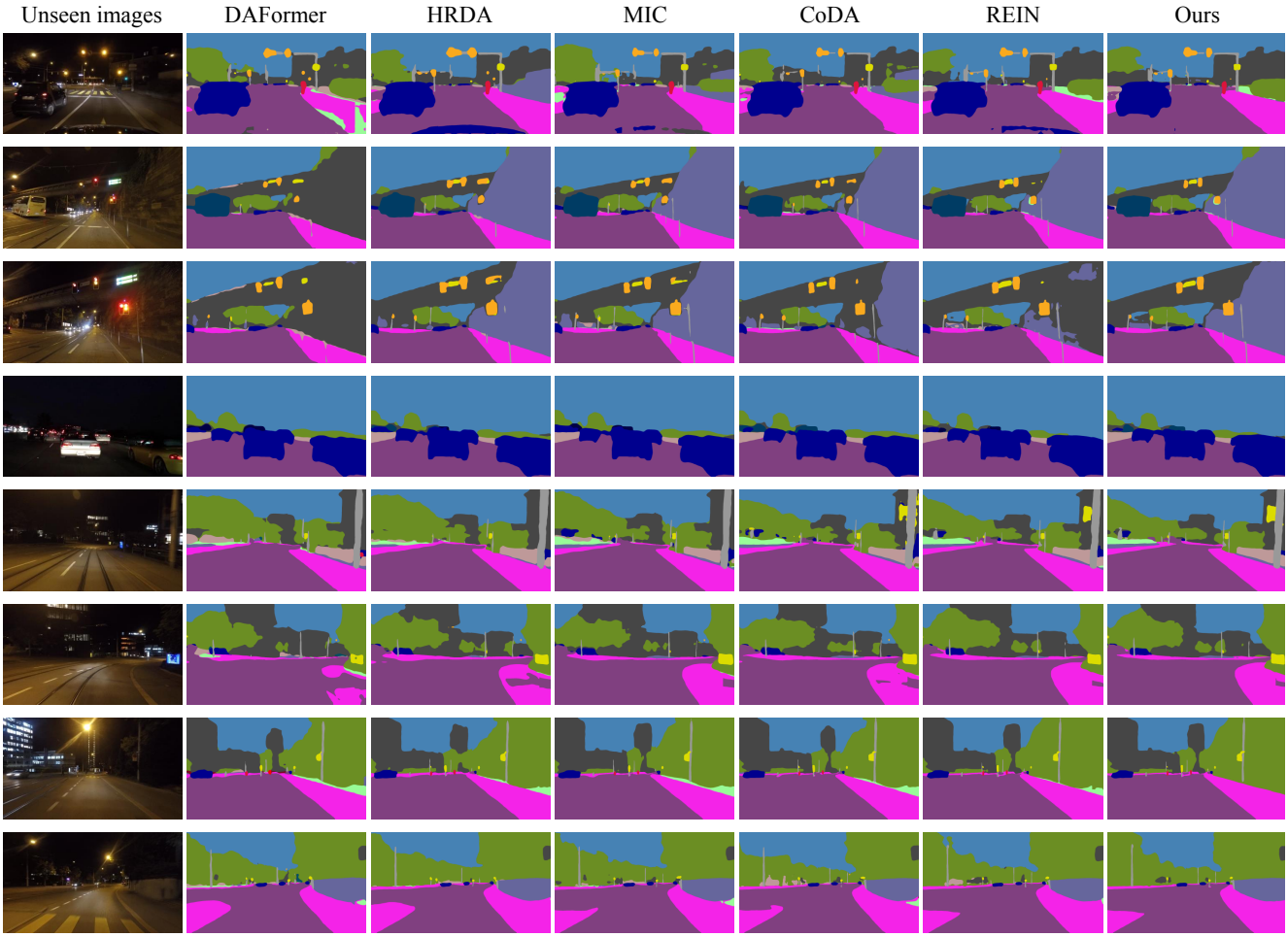

Figure 10. Visual segmentation results on unseen night-time scenes with CityScapes [15] as the source domain. The proposed NightAdapter is compared with DAFormer [28], HRDA [29], MIC [30], CoDA [24] and REIN [71].

etc. in the future. Besides, we will also explore techniques such as adversarial style augmentation to further improve generalization across various scenarios.

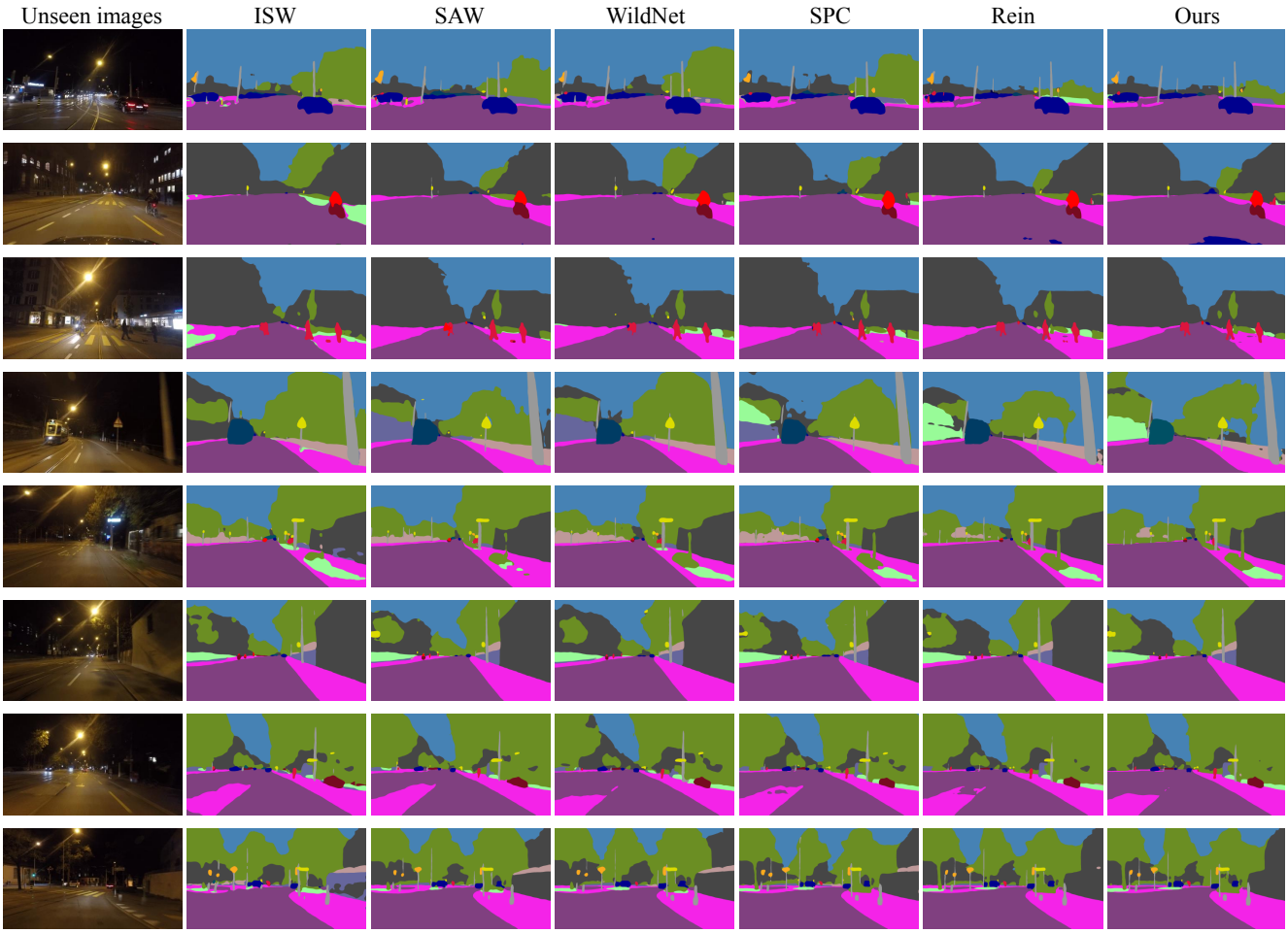

Figure 11. Visual segmentation results on unseen night-time scenes with NightCity [65] as the source domain. The proposed NightAdapter is compared with RobustNet [14], SAW [57],HGFormer [19], CMFormer [6] and REIN [71].