Sun Off, Lights On: Photorealistic Monocular Nighttime Simulation for Robust Semantic Perception — Supplementary Material —

Konstantinos Tzevelekakis[1], Shutong Zhang[2], Luc Van Gool[1,3,4], and Christos Sakaridis[1]

[1]ETH Zürich, [2]University of Toronto, [3]KU Leuven, [4]INSAIT

A. ACDC Light Sources Dataset Labels

As described in Sec. 3.2, to create the ACDC Light Sources dataset we first manually annotated a small subset. A semantic segmentation network was then trained on the manually annotated data to predict the semantic masks for the rest of the images in the dataset. The labels used initially were constrained by the ones the network could accurately predict. For example, "parked_front" and "moving_front" that correspond to the front lights of a vehicle when it is parked or moving respectively, are merged into one label due to the inability of the network to tell them apart. A complete list of the labels used both for annotation and prediction is displayed in Table 1. Additionally, in Figure 1 we provide samples of the panoptic light source annotations.

B. Qualitative Results

Some additional visual results for the qualitative comparison of day-to-night translation methods presented in Sec. 4.3 are illustrated in Fig. 2.

C. Nighttime Illuminants Dataset Samples

As described in Sec. 3.2, for the collection of the Nighttime Illuminants dataset, raw images of a gray card positioned under the illuminant of interest were captured. A gray card is a diffuse surface of known and fixed spectral reflectance across the visible spectrum. In our setting, a 20x25cm Kodak Gray Card reflecting 18% of the light across the visible spectrum was used. These images are then processed by an image processing pipeline to acquire the corresponding chromaticity coordinates. First, the area of the collected images to be processed was manually annotated. Second, the metadata of the raw images stored in the Adobe DNG format [1] were parsed, including active sensor's area, Bayer filter pattern, black and white levels and the camera color matrices. The pipeline continues by filtering the inactive sensor's pixels and applying normalization and black level subtraction. After that, the image is white-balanced using the CIE's [2] standard illuminant E. The demosaicing process follows, along with a transformation from the color space of the camera to the XYZ color space. The chromaticity coordinates (x, y) are finally computed from the average (X, Y, Z) coordinates of the annotated image region. In Figure 3 some samples of the nighttime illuminants dataset are provided.

D. Probabilistic Instantiation Uniform Bounds

In Sec. 3.2, the light source instantiation module has been described. In Table 2 we additionally provide the empirically selected parameters (i.e. bounds) for the discrete uniform distribution from which the Bernoulli parameters that control the activation of the light source are sampled from. The values of those parameters are also conditioned to the light source group. This group is selected based on the panoptic information of the reference images of the ACDC dataset.

References

[1] Yanchao Yang and Stefano Soatto. FDA: Fourier domain adaptation for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 3

[2] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 3

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 3

[1]https://helpx.adobe.com/camera-raw/digital-negative.html

[2]https://cie.co.at/publications/colorimetric-illuminants

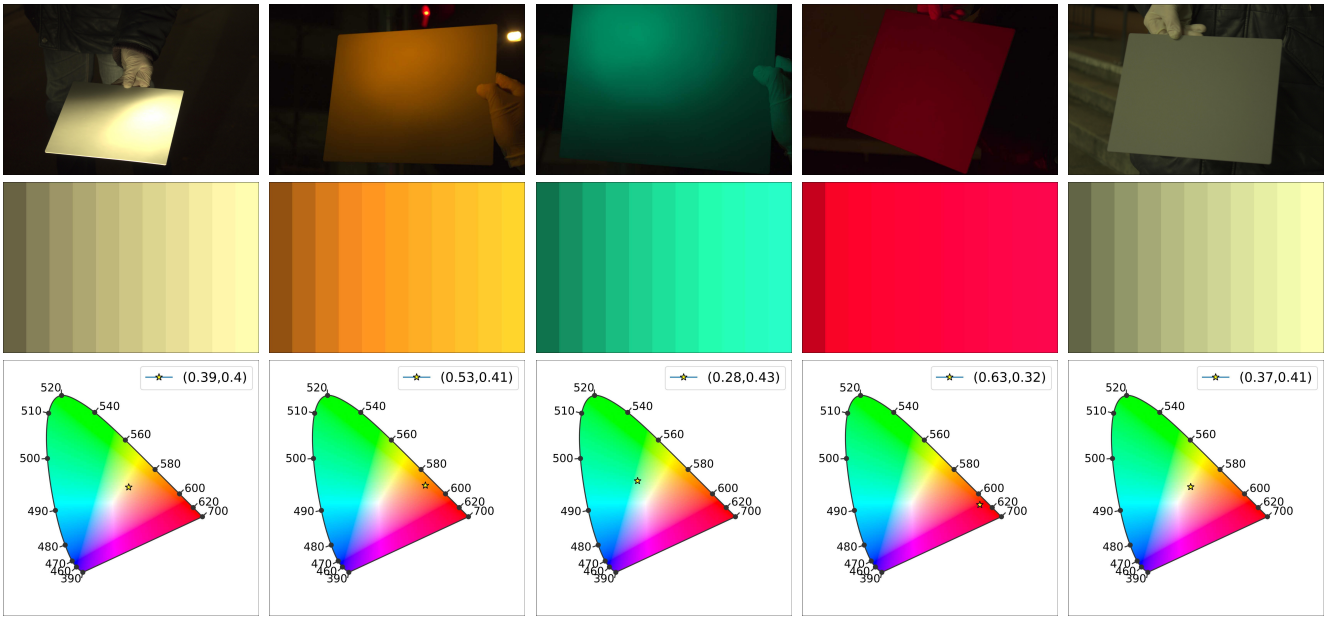
Description	Annotation labels	Prediction labels
Windows of buildings	window_building	window_building
Windows of parked vehicles	window_parked	window_parked
Windows of public transport vehicles (e.g. bus, tram, etc.)	window_transport	window_transport
Traffic lights	traffic_light	traffic_light
Street lights of high correlated color temperature	street_light_HT	street_light
Street lights of low correlated color temperature	street_light_LT	
Front lights of parked vehicles	parked_front	front_light
Front lights of moving vehicles	moving_front	
Rear lights of parked vehicles	parked_rear	rear_light
Rear lights of moving vehicles	moving_rear	
Emitting advertisement panels	advertisement	advertisement
Clocks (e.g. such as those near bus stops)	clock	inferred
Lights whose light color can be inferred from the daytime image	inferred	

Table 1. **ACDC Light Sources dataset labels.** The annotation and prediction labels are described here.

Figure 1. **Panoptic light source annotations samples of the ACDC Light Sources Dataset.** Both manually and automatically – predicted by the light source segmentation model F_L – created annotations are displayed here. The labels used in the annotation process are presented in Table 1. The annotations are overlaid on the corresponding RGB image to improve interpretability.

(a) Input image (b) CycleGAN (c) ControlNet (d) FDA (e) SOLO

Figure 2. **Additional qualitative comparisons of day-to-night translation methods.** From left to right: daytime input images, and synthesized nighttime results of CycleGAN [3], ControlNet [2], FDA [1], and SOLO (ours).



(a) vehicle's front lights (b) traffic light orange (c) traffic light green (d) vehicle's rear lights (e) building's window light

Figure 3. **Nighttime Illuminants dataset samples.** In the first row, the gray card images from which the average chromaticity coordinates (x, y) are calculated are illustrated. The generated color palette (varying luminance) of the sampled chromaticity coordinates is displayed in the second row. Lastly, the (x, y) coordinates are plotted on the CIE 1931 2° Standard Observer chromaticity diagram in the third row.

light source group	light source	uniform bounds
N/A	inferred	(1, 1)
	traffic_light_G	(1, 1)
	traffic_light_R	(1, 1)
	traffic_light_O	(1, 1)
	street_light_HT	(1, 1)
	street_light_LT	(1, 1)
	advertisement	(.6, .8)
	clock	(.8, 1)
	window_building	(.3, .6)
building floor	window_building	(.3, .6)
car	window_parked	(.1, .4)
	moving_front	(.95, 1)
	moving_rear	(.95, 1)
	parked_front	(.1, .3)
	parked_rear	(.1, .3)
bus	window_parked	(.1, .4)
	moving_front	(.95, 1)
	moving_rear	(.95, 1)
	parked_front	(.1, .3)
	parked_rear	(.1, .3)
	window_transport	(.9, 1)
	inferred	(1, 1)
tram	moving_front	(.95, 1)
	moving_rear	(.95, 1)
	parked_front	(.1, .3)
	parked_rear	(.1, .3)
	window_transport	(.9, 1)
	inferred	(1, 1)
truck	window_parked	(.1, .4)
	moving_front	(.95, 1)
	moving_rear	(.95, 1)
	parked_front	(.1, .3)
	parked_rear	(.1, .3)
motorcycle	moving_front	(.95, 1)
	moving_rear	(.95, 1)
	parked_front	(.1, .3)
	parked_rear	(.1, .3)
bicycle	moving_front	(.95, 1)
	moving_rear	(.95, 1)
	parked_front	(.1, .2)
	parked_rear	(.1, .2)

Table 2. **Empirically set intervals for uniform distributions** from which the Bernoulli parameters in light source activation are sampled.