Zero-Shot Day-Night Domain Adaptation with a Physics Prior Supplementary Material

Attila Lengyel[1] Sourav Garg[2] Michael Milford[2] Jan C. van Gemert[1]
Delft University of Technology[1] QUT Centre for Robotics[2]
{a.lengyel, j.c.vangemert}@tudelft.nl {s.garg, michael.milford}@qut.edu.au

1. Derivation of color invariants

This section summarizes the derivation of the Kubelka-Munk [3] based color invariants by Geusebroek et al. [2].

The Kubelka-Munk model for material reflections describes the spectrum of light E reflected from an object in the viewing direction as

$$E(\lambda, \mathbf{x}) = e(\lambda, \mathbf{x}) \left((1 - \rho_f(\mathbf{x}))^2 R_\infty(\lambda, \mathbf{x}) + \rho_f(\mathbf{x}) \right) \quad (1)$$

where $\mathbf{x}$ denotes the spatial location on the image plane, λ the wavelength of the light, e the spectrum of the light source, R_∞ the material reflectivity and ρ_f the Fresnel reflectance coefficient. Partial derivatives of E with respect to x and λ are denoted by subscripts E_x and E_λ, respectively.

By exploring certain simplifying assumptions in Eq. (1) we can derive representations that are invariant to one or more of the following conditions: 1) *scene geometry*, i.e. shadows and shading; 2) *Fresnel reflections* from shiny surfaces; 3) *illumination intensity*; and 4) *illumination color*.

1.1. *E*

E is a non-invariant baseline edge detector and therefore no simplifying assumptions are made on Eq. (1). Color invariant E is simply defined as:

$$E = \sqrt{E_x^2 + E_{\lambda x}^2 + E_{\lambda\lambda x}^2 + E_y^2 + E_{\lambda y}^2 + E_{\lambda\lambda y}^2}. \quad (2)$$

1.2. *W*

Assuming spectrally and spatially uniform illumination, $e(\lambda, \mathbf{x})$ can be represented by a constant i. Moreover, assuming only matte surfaces, i.e. $\rho_f(\mathbf{x}) = 0$, Eq. (1) reduces to

$$E(\lambda, \mathbf{x}) = i R_\infty(\lambda, \mathbf{x}). \quad (3)$$

The ratio $W_x = \frac{E_x}{E}$ is then independent of the illuminant i:

$$W_x = \frac{E_x}{E} = \frac{1}{R_\infty(\lambda, \mathbf{x})} \frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \mathbf{x}} \quad (4)$$

The same holds for the ratios $W_{\lambda x} = \frac{E_{\lambda x}}{E}$ and $W_{\lambda\lambda x} = \frac{E_{\lambda\lambda x}}{E}$, and consequently the invariant W can be defined as

$$W = \sqrt{W_x^2 + W_{\lambda x}^2 + W_{\lambda\lambda x}^2 + W_y^2 + W_{\lambda y}^2 + W_{\lambda\lambda y}^2}. \quad (5)$$

W is invariant to *illumination intensity*.

1.3. *C*

We assume a spectrally uniform illuminant represented as $i(\mathbf{x})$ and matte surfaces, i.e. $\rho_f(\mathbf{x}) = 0$. Eq. (1) then reduces to

$$E(\lambda, \mathbf{x}) = i(\mathbf{x}) R_\infty(\lambda, \mathbf{x}). \quad (6)$$

The ratio $C_\lambda = \frac{E_\lambda}{E}$ is then independent of the illuminant i:

$$C_\lambda = \frac{E_\lambda}{E} = \frac{1}{R_\infty(\lambda, \mathbf{x})} \frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \lambda}. \quad (7)$$

The same holds for the ratios $C_{\lambda\lambda} = \frac{E_{\lambda\lambda}}{E}$, $C_{\lambda x} = \frac{E_{\lambda x} E - E_\lambda E_x}{E^2}$ and $C_{\lambda\lambda x} = \frac{E_{\lambda\lambda x} E - E_{\lambda\lambda} E_x}{E^2}$. The color invariant C is defined as

$$C = \sqrt{C_{\lambda x}^2 + C_{\lambda\lambda x}^2 + C_{\lambda y}^2 + C_{\lambda\lambda y}^2}. \quad (8)$$

C is invariant to *scene geometry* and *illumination intensity*.

1.4. *N*

We assume a colored illuminant where the power spectrum remains constant over the scene and only varies in intensity, such that the illuminant can be decomposed into a separate spectral and spatial term as $e(\lambda, \mathbf{x}) = e(\lambda) i(\mathbf{x})$. Furthermore, we again assume matte surfaces, i.e. $\rho_f(\mathbf{x}) = 0$. Eq. (1) is then defined as

$$E(\lambda, \mathbf{x}) = e(\lambda) i(\mathbf{x}) R_\infty(\lambda, \mathbf{x}). \quad (9)$$

Differentiating Eq. (9) with respect to λ yields

$$E_\lambda = i(\mathbf{x}) R_\infty(\lambda, \mathbf{x}) \frac{\partial e(\lambda)}{\partial \lambda} + e(\lambda) i(\mathbf{x}) \frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \lambda}. \quad (10)$$

Dividing Eq. (10) by Eq. (9) results in a representation that is invariant to the spatial illuminant term i:

$$N_\lambda = \frac{E_\lambda}{E} = \frac{1}{e(\lambda)}\frac{\partial e(\lambda)}{\partial \lambda} + \frac{1}{R_\infty(\lambda, \mathbf{x})}\frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \lambda}. \quad (11)$$

Additionally differentiating with respect to $\mathbf{x}$ results in the left term dropping out, yielding the color invariant $N_{\lambda x}$ which is only dependent on the material property R_∞:

$$N_{\lambda x} = \frac{\partial}{\partial \mathbf{x}}\left\{\frac{E_\lambda}{E}\right\} = \frac{\partial}{\partial \mathbf{x}}\left\{\frac{1}{R_\infty(\lambda, \mathbf{x})}\frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \lambda}\right\}, \quad (12)$$

$$= \frac{E_{\lambda x}E - E_\lambda E_x}{E^2}. \quad (13)$$

The same holds for higher order derivatives, e.g.

$$N_{\lambda\lambda x} = \frac{E_{\lambda\lambda x}E^2 - E_{\lambda\lambda}E_x E - 2E_{\lambda x}E_\lambda E + 2E_\lambda^2 E_x}{E^3}. \quad (14)$$

The color invariant N is defined as

$$N = \sqrt{N_{\lambda x}^2 + N_{\lambda\lambda x}^2 + N_{\lambda y}^2 + N_{\lambda\lambda y}^2} \quad (15)$$

and is invariant to *scene geometry*, *illumination intensity* and *illumination color*.

1.5. *H*

We again assume an illuminant with uniform power spectrum such that $e(\lambda, \mathbf{x}) = i(\mathbf{x})$. Eq. (1), including Fresnel reflections, then simplifies to

$$E(\lambda, \mathbf{x}) = i(\mathbf{x})\left((1 - \rho_f(\mathbf{x}))^2 R_\infty(\lambda, \mathbf{x}) + \rho_f(\mathbf{x})\right) \quad (16)$$

The first and second order derivatives with respect to λ are defined as

$$E_\lambda = i(x)\,(1 - \rho_f(\mathbf{x}))^2\,\frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \lambda}, \quad (17)$$

$$E_{\lambda\lambda} = i(x)\,(1 - \rho_f(\mathbf{x}))^2\,\frac{\partial^2 R_\infty(\lambda, \mathbf{x})}{\partial \lambda^2}. \quad (18)$$

The ratio $H = \frac{E_\lambda}{E_{\lambda\lambda}}$ then only depends on the material property R_∞ and is thus an invariant to *scene geometry*, *illumination intensity* and *Fresnel reflections*. Since the spatial derivative $H_x = \frac{\partial}{\partial x}\frac{E_\lambda}{E_{\lambda\lambda}}$ is ill-defined for $E_{\lambda\lambda} = 0$, H is instead defined as $H = \arctan\frac{E_\lambda}{E_{\lambda\lambda}}$, for which the spatial derivative is

$$H_x = \frac{1}{1 + \left(\frac{E_\lambda}{E_{\lambda\lambda}}\right)^2}\frac{E_{\lambda\lambda}E_{\lambda x} - E_\lambda E_{\lambda\lambda x}}{E_{\lambda\lambda}^2} \quad (19)$$

$$= \frac{E_{\lambda\lambda}E_{\lambda x} - E_\lambda E_{\lambda\lambda x}}{E_\lambda^2 + E_{\lambda\lambda}^2}. \quad (20)$$

Color invariant H is defined as

$$H = \sqrt{H_x^2 + H_y^2}. \quad (21)$$

2. Distribution alignment by CIConv

Fig. 1 shows the feature map activations of a baseline ResNet-18 model and each of the different color invariant models, as described in section 4.1 of the paper. The intensity change between the "Normal" (daytime) and "Darker" (nighttime) test set causes a clear distribution shift throughout all network layers of the baseline model. In contrast, the CIConv layer produces a domain invariant feature representation and consequently the distributions in the color invariant networks are more aligned between the two domains. This is the case for each of the color invariants, although the "Normal" and "Darker" distributions in the final layer appear to be most aligned for W, which may explain its generally better performance compared to the other invariants.

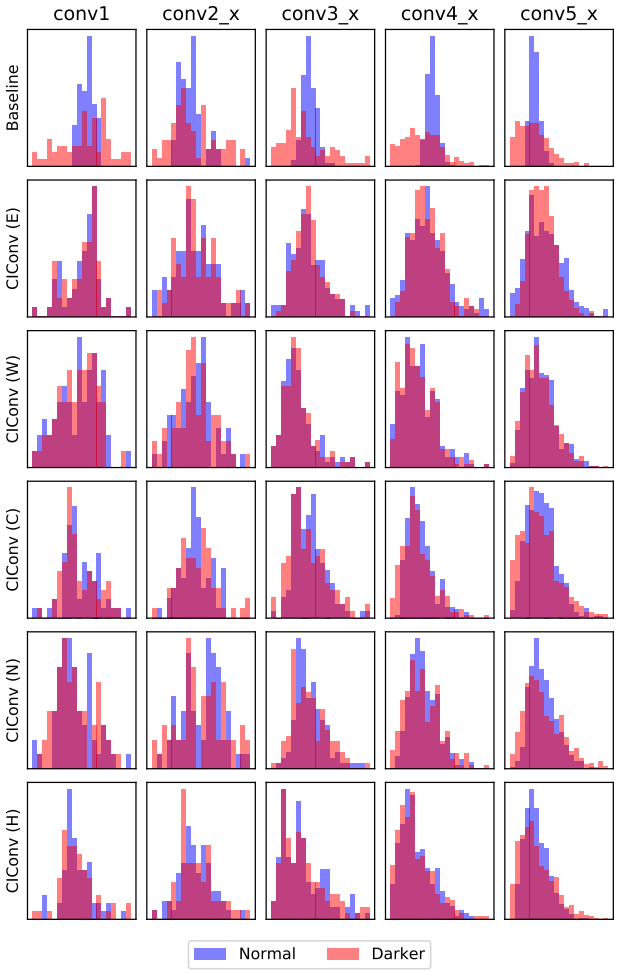

Figure 1: Histogram of ResNet-18 feature map activations for "Normal" (daytime) and "Darker" (nighttime) test sets of synthetic dataset. Baseline network shows clear distribution shift between test sets, which is greatly reduced in color invariant networks.

Table 1: Feature map activation similarities of ResNet-18 feature maps for "Normal" and "Darker" test sets of synthetic dataset, measured by L2 distance. W has most constant feature maps.

	conv1	conv2	conv3	conv4	conv5
Baseline	25.77	2.25	2.32	2.96	2.71
E	0.02	0.43	0.5	0.58	0.58
W	0.02	**0.36**	**0.38**	**0.55**	**0.46**
C	0.02	0.95	0.91	1.33	1.14
N	0.02	1.1	1.06	1.33	1.14
H	**0.01**	0.8	0.88	0.98	1.19

Table 2: Classification accuracy (%) on CODaN. Combining W with other input modalities does not improve performance.

$W+$	None	RGB	E	C	N	H
Day	**81.49**	66.08	69.72	66.00	66.48	68.56
Night	**59.67**	43.52	46.65	46.44	45.19	47.65

To quantify the distribution shift we computed the L2 distance between the feature map activations for the "Normal" and "Darker" test sets. As shown in Table 1, W has indeed the most constant feature map activations.

3. Combining color invariants

We investigated the use of multiple input modalities by concatenating the output of W with either RGB, E, C, N or H in the input layer. Results on the CODaN classification dataset in Table 2 show that performance deteriorates compared to only W (None), likely due to overfitting on a combination of input modalities rather than using them in a complementary fashion. This again shows the need for developing an adaptive fusion mechanism as mentioned in the Discussion.

4. Semantic Segmentation per-class scores

The per-class Intersection-over-Union (IoU) scores of the semantic segmentation experiment are shown in Table 3 for Nighttime Driving [1] and Table 4 for Dark Zurich [8]. Our W-RefineNet improves segmentation over the baseline performance across nearly all classes.

References

[1] D. Dai and L. V. Gool. Dark model adaptation: Semantic image segmentation from daytime to nighttime. In *ITSC*, pages 3819–3824, Nov 2018. 3, 4

[2] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts. Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12):1338–1350, 2001. 1

[3] P. Kubelka and F. Munk. Ein beitrag zur optik der farbanstriche. In *Zeitung fur Technische Physik*, volume 12, page 593, 1999. 1

[4] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *CoRR*, abs/1603.04779, 2016. 4

[5] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 4

[6] G. Lin, A. Milan, C. Shen, and I. Reid. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. In *CVPR*, July 2017. 4

[7] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 4

[8] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 3, 4

[9] Y. Tsai, W. Hung, S. Schulter, K. Sohn, M. Yang, and M. Chandraker. Learning to adapt structured output space for semantic segmentation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7472–7481, 2018. 4

[10] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Mathieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *CVPR*, 2019. 4

Table 3: Per-class semantic segmentation results on the Nighttime Driving [1] dataset, reported as IoU.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	**mIoU**
Trained on source data																				
RefineNet [6]	83.2	32.9	82.0	18.6	0.0	35.5	22.5	39.4	**45.8**	0.0	29.0	53.0	0.0	57.7	0.0	67.7	63.1	0.0	16.5	34.1
W-RefineNet [ours]	**89.6**	52.7	**82.7**	16.2	0.0	**39.6**	**52.2**	**60.6**	43.9	0.0	**38.6**	**55.1**	**24.3**	**72.0**	0.0	**73.2**	**66.8**	0.0	23.6	**41.6**
RefineNet-AdaBN [4]	88.9	**58.2**	75.5	**22.5**	0.0	39.0	21.3	50.9	36.4	0.0	25.7	53.4	0.0	68.0	0.0	63.3	62.7	0.0	**24.4**	36.3

Table 4: Per-class semantic segmentation results on the Dark Zurich [7] dataset, reported as IoU. Results of methods trained on source and target data taken from [8].

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	**mIoU**
Trained on source data																				
RefineNet [6]	86.2	34.8	**62.0**	26.0	12.8	30.9	14.4	27.7	38.4	10.0	3.1	**38.3**	34.5	49.1	6.0	0.0	**55.4**	31.1	**20.4**	30.6
W-RefineNet [ours]	**90.3**	48.3	57.8	**29.3**	11.1	**36.3**	**24.4**	30.2	**45.8**	7.6	**8.0**	37.6	**40.1**	**69.7**	10.1	0.0	55.0	**37.4**	16.0	**34.5**
RefineNet-AdaBN [4]	87.0	**51.8**	53.1	28.4	**14.7**	32.8	11.3	**31.9**	33.8	**18.4**	2.4	32.4	39.6	59.7	**10.5**	0.0	32.9	34.2	20.0	31.3
Trained on source and target data																				
AdaptSegNet [9]	**86.1**	44.2	55.1	22.2	4.8	21.1	5.6	16.7	37.2	8.4	1.2	35.9	26.7	68.2	45.1	0.0	50.1	33.9	15.6	30.4
ADVENT [10]	85.8	37.9	55.5	27.7	14.5	23.1	14.0	21.1	32.1	8.7	2.0	39.9	16.6	64.0	13.8	0.0	**58.8**	28.5	20.7	29.7
BDL [5]	85.3	41.1	61.9	**32.7**	17.4	20.6	11.4	21.3	29.4	8.9	1.1	37.4	22.1	63.2	28.2	0.0	47.7	**39.4**	15.7	30.8
DMAda [1]	75.5	29.1	48.6	21.3	14.3	34.3	36.8	29.9	49.4	13.8	0.4	43.3	50.2	69.4	18.4	0.0	27.6	34.9	11.9	32.1
GCMA [7]	81.7	46.9	58.8	22.0	**20.0**	41.2	**40.5**	**41.6**	**64.8**	**31.0**	32.1	**53.5**	47.5	**75.5**	39.2	0.0	49.6	30.7	21.0	42.0
MGCDA [8]	80.3	**49.3**	**66.2**	7.8	11.0	**41.4**	38.9	39.0	64.1	18.0	**55.8**	52.1	**53.5**	74.7	**66.0**	0.0	37.5	29.1	**22.7**	**42.5**