Dark Side Augmentation: Supplementary

Albert Mohwald Tomas Jenicek Ondřej Chum
VRG, Faculty of Electrical Engineering, Czech Technical University in Prague
mohwaalb@fel.cvut.cz jenicto2@fel.cvut.cz chum@cmp.felk.cvut.cz

1 Appendix A

Full table versions. Method variants not shown in the main paper are emphasized by bold.

In Table A1, and Table A7, we also show results using image generators tuned by us. Specifically in CUT [9], the downsampling and upsampling convolutional layers are trainable in both the generator and discriminator, generator weights are initialized following [3] as in our HED$^{\text{N}}$GAN, the contrastive loss is applied on the output of a different set of layers (4,7,10,14), and weight λ_Y of identity loss is set to 10. We observe that learning the convolutional layers and changing the layers in the contrastive loss aids the performance of the final retrieval models, while the other changes has shown to stabilize training in our setup. In CyEDA [2], while training on the *SfM*120k dataset [12], weight λ of cycle consistency is set to 0.3. We observed that the original $\lambda = 10$ causes the night generator to synthesise images nearly identical to the input day images.

In the Table A7, we also tested augmentation with CycleGAN trained on the *Aachen* dataset [15, 14, 20], with both Aachen v1.0 and Aachen v1.1 images together. *Aachen* dataset contains much less night training examples than *SfM*120k dataset (206 compared to 10039, respectively). Despite that, the performance loss of CycleGAN Aachen augmentation is negligible, see Table A7.

VGG-16 backbone

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
GeM [12]	69.9	79.4	**60.9**	69.3
GeM N/D [4] **!**	71.1	83.5	60.0	69.8
CIConv [6]	-	83.3	-	-
CLAHE $\mathcal{C}$ [4]	71.6	84.1	60.8	69.8
CLAHE N/D $\mathcal{C}$ [4] **!**	72.4	87.0	60.2	70.0
HED$^{\mathrm{N}}$GAN $\mathcal{CD}$ (ours)	**73.4**	**88.9**	**61.1**	**70.3**
CycleGAN $\mathcal{CD}$ (ours)	**74.0**	**90.2**	60.7	**71.0**

ResNet-18 backbone

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
GeM	65.4	76.1	50.5	69.5
GeM N/D !	67.0	79.6	51.3	**70.0**
CLAHE $\mathcal{C}$	67.3	80.6	52.6	68.6
CLAHE N/D $\mathcal{C}$ **!**	68.0	82.5	52.5	69.0
RCF$^{\mathrm{N}}$GAN $\mathcal{CD}$ **(ours)**	69.4	83.5	**54.4**	**70.0**
HED$^{\mathrm{N}}$GAN $\mathcal{CD}$ **(ours)**	**69.6**	**85.1**	53.6	**70.0**
CycleGAN $\mathcal{CD}$ **(ours)**	**69.7**	**84.4**	**55.1**	69.6

ResNet-50 backbone

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
GeM	74.6	85.4	63.4	**75.1**
GeM N/D !	75.7	88.3	63.1	**75.6**
CLAHE $\mathcal{C}$	74.7	87.4	62.5	74.2
CLAHE N/D $\mathcal{C}$ **!**	75.3	89.0	62.3	74.5
RCF$^{\mathrm{N}}$GAN $\mathcal{CD}$ **(ours)**	76.8	91.4	**64.5**	74.4
HED$^{\mathrm{N}}$GAN $\mathcal{CD}$ **(ours)**	**77.0**	**91.7**	**64.4**	74.9
CycleGAN $\mathcal{CD}$ **(ours)**	**77.0**	**92.3**	64.0	74.7

ResNet-101 backbone

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
GeM [12]	75.7	85.0	65.3	76.7
GeM N/D !	77.0	88.6	65.7	76.8
CIConv [6]	75.0	88.3	62.0	74.7
CLAHE $\mathcal{C}$	76.9	88.1	66.1	76.6
CLAHE N/D $\mathcal{C}$ **!**	77.4	89.5	66.1	76.5
CUT $\mathcal{CD}$	77.9	90.2	65.7	**77.7**
CUT (tuned) $\mathcal{CD}$	78.0	90.9	65.7	77.3
CyEDA BDD $\mathcal{CD}$	77.8	90.3	65.7	77.3
CyEDA GTA $\mathcal{CD}$	78.2	91.2	65.8	**77.6**
CyEDA (tuned) $\mathcal{CD}$	78.0	90.9	65.5	77.5
RCF$^{\mathrm{N}}$GAN $\mathcal{CD}$ **(ours)**	78.2	91.5	**66.8**	76.3
HED$^{\mathrm{N}}$GAN $\mathcal{CD}$ (ours)	**78.4**	**92.2**	66.3	76.6
CycleGAN $\mathcal{CD}$ (ours)	**78.4**	**92.0**	**66.8**	76.4

(continues)

HOW ResNet-18 backbone

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
HOW [16]	80.8	87.8	**75.1**	79.4
HOW N/D **!**	**82.0**	89.2	**75.5**	**81.4**
RCF$^{\mathrm{N}}$GAN $\mathcal{CD}$ (ours)	81.8	91.5	74.6	79.4
HED$^{\mathrm{N}}$GAN $\mathcal{CD}$ (ours)	**82.0**	**91.6**	74.6	79.7
CycleGAN $\mathcal{CD}$ (ours)	**82.4**	**92.9**	74.6	**79.8**

Table A1: Comparison in terms of mAP on Tokyo 24/7, $\mathcal{R}$Oxf Medium and $\mathcal{R}$Par Medium datasets and their average on retrieval. Methods marked by **!** use paired day-night training data. Methods starting with GeM and CLAHE not marked with a reference were trained by us. The best score for each backbone architecture (in separate tables) is emphasized by red bold, second best by bold.

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
EdgeMAC [11]	45.6	75.9	17.3	43.5
HEDMAC	56.8	79.5	38.3	52.5
HED$^{\mathrm{N}}$MAC	**59.2**	81.9	**38.4**	**57.2**
RCF$^{\mathrm{N}}$MAC	58.5	**88.9**	35.1	51.4
HEDMAC+GeM **‡**	72.0	84.8	60.9	70.3
HED$^{\mathrm{N}}$MAC+GeM **‡**	72.6	85.7	**61.1**	70.9
HEDMAC+$^{\mathrm{N}}$GAN ‡	73.8	90.9	60.1	70.5
HED$^{\mathrm{N}}$MAC+$^{\mathrm{N}}$GAN **‡**	74.4	91.4	60.6	71.3
HEDMAC+GAN ‡	74.2	91.5	60.0	71.2
HED$^{\mathrm{N}}$MAC+GAN **‡**	**74.7**	**91.8**	60.4	**71.9**

Table A4: The effect of our trained HED$^{\mathrm{N}}$ detector (from HED$^{\mathrm{N}}$GAN) on the EdgeMAC [11] method. HEDMAC and HED$^{\mathrm{N}}$MAC is a variant of EdgeMAC method with the HED [18] edge detector with either original or our weights, respectively. RCF$^{\mathrm{N}}$MAC is a variant of EdgeMAC with the RCF [7] edge detector with our weights. In the bottom block, ensembles of EdgeMAC variants with chosen methods from Table A1 are reported. GeM is from [12], $^{\mathrm{N}}$GAN corresponds to HED$^{\mathrm{N}}$GAN $\mathcal{CD}$, and GAN to CycleGAN $\mathcal{CD}$, all from Table A1. Ensembles have double the dimensionality (1024) and are marked with **‡**. The best scores for each dimensionality are in bold.

Method	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
GeM	70.0	80.4	59.9	69.8
GeM $\mathcal{D}$	70.3	81.2	59.8	69.9
CLAHE $\mathcal{C}$	71.9	85.4	60.0	70.1
CLAHE $\mathcal{CD}$	**72.2**	**85.9**	**60.3**	**70.5**
GeM N/D !	71.5	84.0	60.4	70.0
GeM N/D $\mathcal{D}$ **!**	71.5	84.1	60.3	70.1
CLAHE N/D $\mathcal{C}$ **!**	72.5	87.5	59.9	70.1
CLAHE N/D $\mathcal{CD}$ **!**	**73.0**	**87.7**	**60.8**	**70.7**
RCF$^{\text{N}}$GAN	72.3	86.8	59.8	70.2
RCF$^{\text{N}}$GAN $\mathcal{D}$	72.7	87.6	60.1	70.4
RCF$^{\text{N}}$GAN $\mathcal{C}$	72.9	87.8	**60.4**	70.6
RCF$^{\text{N}}$GAN $\mathcal{CD}$	**73.2**	**88.3**	**60.4**	**70.8**
HED$^{\text{N}}$GAN	72.7	88.0	60.2	70.0
HED$^{\text{N}}$GAN $\mathcal{D}$	73.0	88.7	60.2	70.1
HED$^{\text{N}}$GAN $\mathcal{C}$	73.2	88.7	60.5	70.4
HED$^{\text{N}}$GAN $\mathcal{CD}$	**73.4**	**88.8**	**60.7**	**70.6**
CycleGAN	73.0	88.8	59.6	70.5
CycleGAN $\mathcal{D}$	73.3	89.1	59.9	70.7
CycleGAN $\mathcal{C}$	73.6	89.6	60.5	70.9
CycleGAN $\mathcal{CD}$	**74.0**	**90.2**	**60.7**	**71.0**

Table A5: The effect of diverse anchors ($\mathcal{D}$). Methods GeM [12], GeM N/D, CLAHE $\mathcal{C}$ [4], and CLAHE N/D $\mathcal{C}$ [4] from Table A1 are reported in the top two blocks. Please note that we re-train the models for this ablation, so we obtain a slightly higher performance. In the last three blocks, the effect of CLAHE ($\mathcal{C}$) and diverse anchors ($\mathcal{D}$) is reported on the RCF$^{\text{N}}$GAN, HED$^{\text{N}}$GAN and CycleGAN methods. The best score for each dataset in each block is in bold.

Method	{day, sunset}		{sunset, night}		{day, night}	
	D→S	S→D	S→N	N→S	D→N	N→D
CLAHE $\mathcal{C}$	97.7	98.2	80.1	81.3	70.9	76.1
CLAHE $\mathcal{CD}$	97.5	98.4	80.2	81.3	72.4	77.0
CLAHE N/D $\mathcal{C}$ **!**	97.5	98.2	80.3	86.2	73.0	81.3
CLAHE N/D $\mathcal{CD}$ **!**	97.4	97.8	80.8	86.5	74.0	82.2
HED$^{\text{N}}$GAN $\mathcal{C}$	97.1	98.2	84.5	86.9	77.1	80.3
HED$^{\text{N}}$GAN $\mathcal{CD}$	97.5	98.0	84.3	88.3	77.9	81.1
CycleGAN $\mathcal{C}$	97.6	98.2	85.7	88.5	78.8	81.4
CycleGAN $\mathcal{CD}$	97.8	98.2	86.8	89.3	80.0	82.6

Table A6: Retrieval performance (mAP) on *Tokyo* for a combination of three different subsets of the dataset – day (D), sunset (S), and night (N). Images from the first class are always queries and from the second class are positives (query→positive); the last image of the scene from the unused class is excluded from the evaluation. Scores for selected methods from Table A5 are reported.

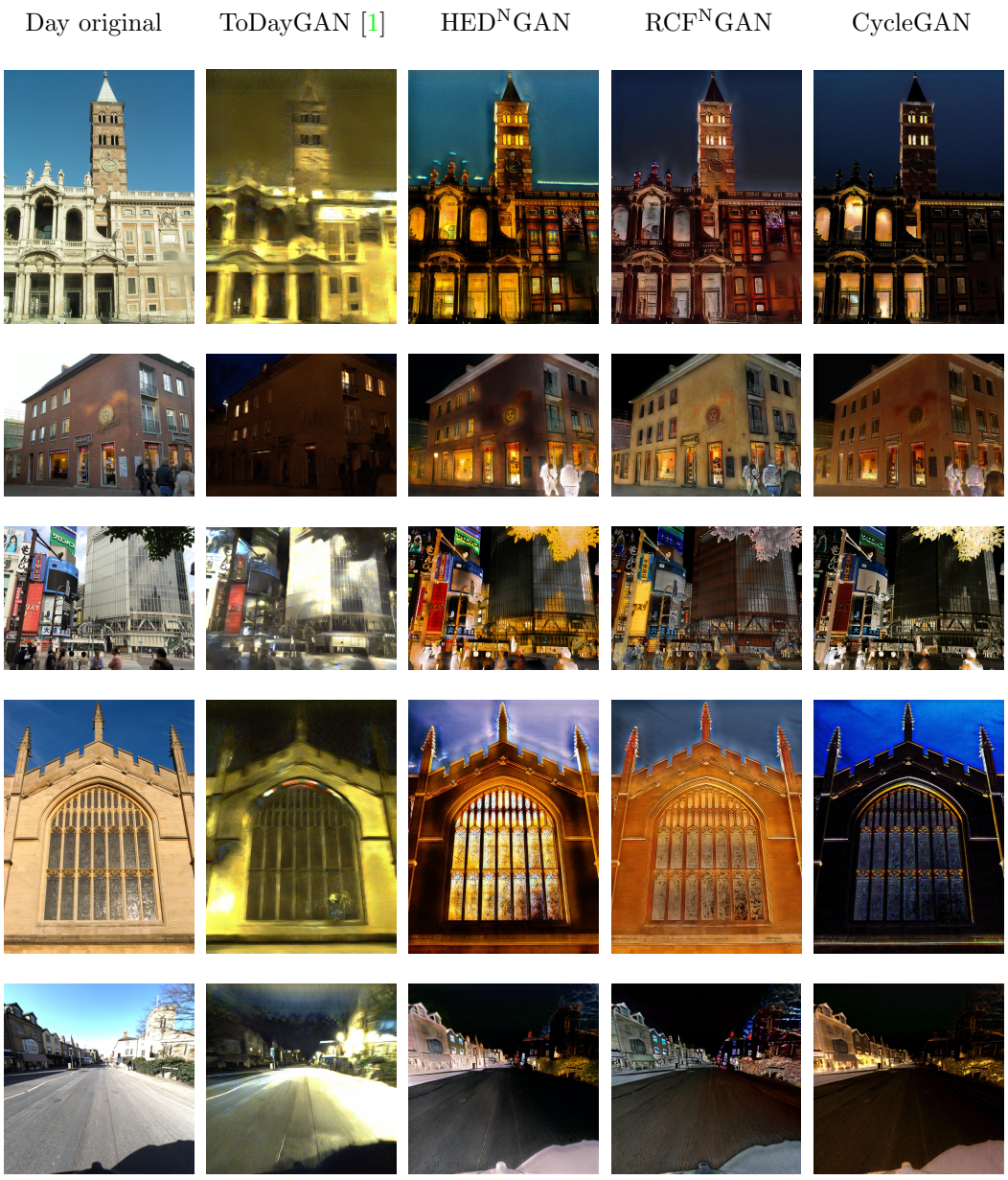
Night Data	Avg	Tokyo	$\mathcal{R}$Oxf	$\mathcal{R}$Par
DRIT $\mathcal{CD}$	73.5	90.2	59.8	70.5
CUT $\mathcal{CD}$	73.0	87.7	60.3	**71.1**
CUT (tuned) $\mathcal{CD}$	73.4	88.6	60.8	70.8
CyEDA [2] BDD $\mathcal{CD}$	72.9	87.9	60.4	70.3
CyEDA [2] GTA $\mathcal{CD}$	72.9	87.9	60.3	70.4
CyEDA $\mathcal{CD}$	70.9	82.0	60.1	70.5
CyEDA (tuned) $\mathcal{CD}$	73.3	88.2	60.5	71.2
RCF$^{\mathrm{N}}$GAN $\mathcal{CD}$	73.2	88.3	60.4	70.8
RCF$^{\mathrm{N}}$GAN 50% $\mathcal{CD}$	73.4	90.0	59.7	70.4
HEDGAN $\mathcal{CD}$	73.2	88.1	61.0	70.5
HED$^{\mathrm{N}}$GAN $\mathcal{CD}$	73.4	88.9	**61.1**	70.3
HED$^{\mathrm{N}}$GAN 50% $\mathcal{CD}$	73.4	90.3	60.0	70.0
CycleGAN $\mathcal{CD}$	**74.0**	90.2	60.7	71.0
CycleGAN 50% $\mathcal{CD}$	73.9	**91.4**	60.0	70.4
CycleGAN Aachen $\mathcal{CD}$	73.8	89.9	60.7	70.8
CycleGAN + N/D $\mathcal{CD}$ **!**	73.5	88.6	60.8	**71.1**
CycleGAN + N/D 50% $\mathcal{CD}$ **!**	73.9	90.1	**61.1**	70.6
HED$^{\mathrm{N}}$GAN + N/D $\mathcal{CD}$ **!**	73.3	87.9	**61.1**	70.8
HED$^{\mathrm{N}}$GAN + N/D 50% $\mathcal{CD}$ **!**	73.6	89.1	**61.1**	70.6
HED$^{\mathrm{N}}$GAN + HEDGAN $\mathcal{CD}$	73.4	88.0	60.7	70.6
HED$^{\mathrm{N}}$GAN + HEDGAN 50% $\mathcal{CD}$	73.4	90.0	60.5	69.8
CycleGAN + CycleGAN Aachen $\mathcal{CD}$	74.0	90.4	60.5	**71.1**
CycleGAN + CycleGAN Aachen 50% $\mathcal{CD}$	74.0	**91.4**	60.3	70.4
HED$^{\mathrm{N}}$GAN + CycleGAN $\mathcal{CD}$	74.0	90.0	61.0	70.9
HED$^{\mathrm{N}}$GAN + CycleGAN 50% $\mathcal{CD}$	**74.1**	**91.4**	60.5	70.5

Table A7: The impact of retrieval training data. In the top block, generator architectures DRIT [5], CUT [9] (original and our tuned variation), CyEDA [2] (pretrained models from [2], trained by us on *SfM* dataset, and our tuned variation trained on *SfM*120k), RCF$^{\mathrm{N}}$GAN (trained RCF), HEDGAN, HED$^{\mathrm{N}}$GAN (trained HED), and CycleGAN [21] are tested. In the bottom block, the HED$^{\mathrm{N}}$GAN or CycleGAN generator architecture is further combined with the *SfM-N/D* dataset (*e.g.* HED$^{\mathrm{N}}$GAN + N/D) or a different generator (*e.g.* HED$^{\mathrm{N}}$GAN + CycleGAN) with ratio 1:1; scores for 25% (default in experiments) and 50% of night images in the training data are reported. Methods marked by **!** use paired day-night training data. The best score for each dataset in each block is in bold.

2 Appendix B

We provide example outputs for all generative models referred in the paper. Generators for translation from day to night as well as from night to day are examined in Figure B1, and Figure B2, respectively. Please note that HED$^{\text{N}}$GAN, RCF$^{\text{N}}$GAN, and CUT architectures do not contain a night $\rightarrow$ day generator.

We observe that for night to day translation (Figure B2), the ToDayGAN [1] method achieves satisfactory results only on the *RobotCar* dataset [8]. Also, these results are visually more appealing than other results on this dataset, even though a car is hallucinated close to the middle of the image. This can be explained by the fact that the ToDayGAN method is specifically trained for this dataset.



(continues)

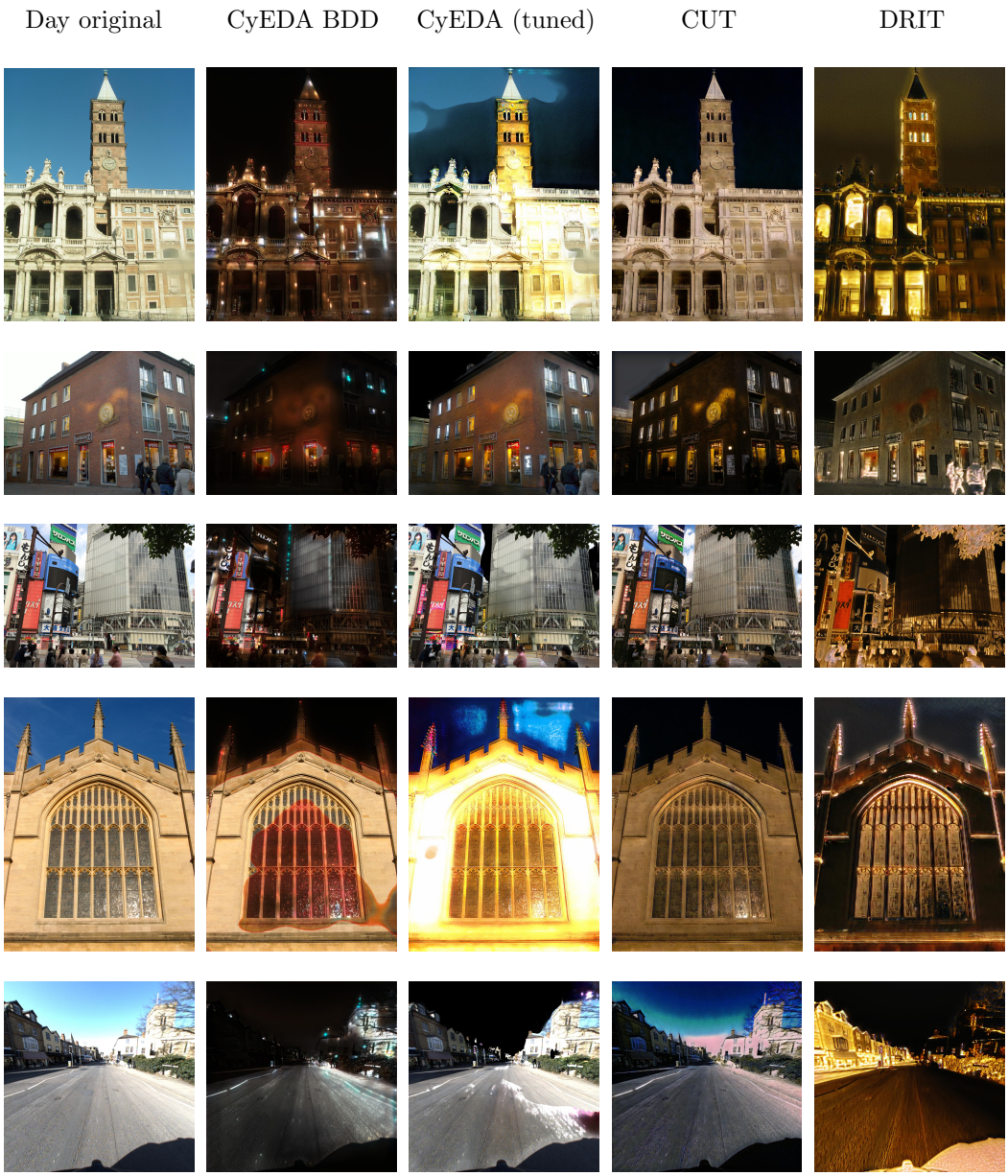
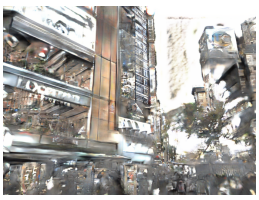
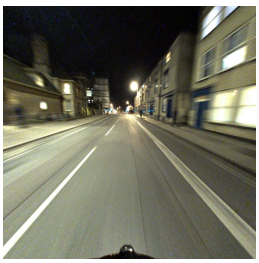
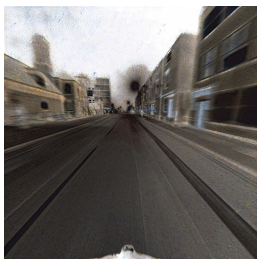


Figure B1: Examples of day images translated into the night domain by different day → night generators. The columns correspond to (left-to-right): the original image, ToDayGAN generator [1], our generators HED$^{\text{N}}$GAN, RCF$^{\text{N}}$GAN, and CycleGAN, the original image, CyEDA [2] generator trained on BDD dataset [19], CyEDA generator trained on *SfM*120k [12] and tuned by us, and our CUT, and DRIT generators. The rows show example images from different datasets (top-to-bottom): *SfM*120k [12], *Aachen* [14, 15], *Tokyo* [17], Oxford [10], and *RobotCar* [8], respectively.

Night original	ToDayGAN [1]	CycleGAN
		
		
		
		

(continues)

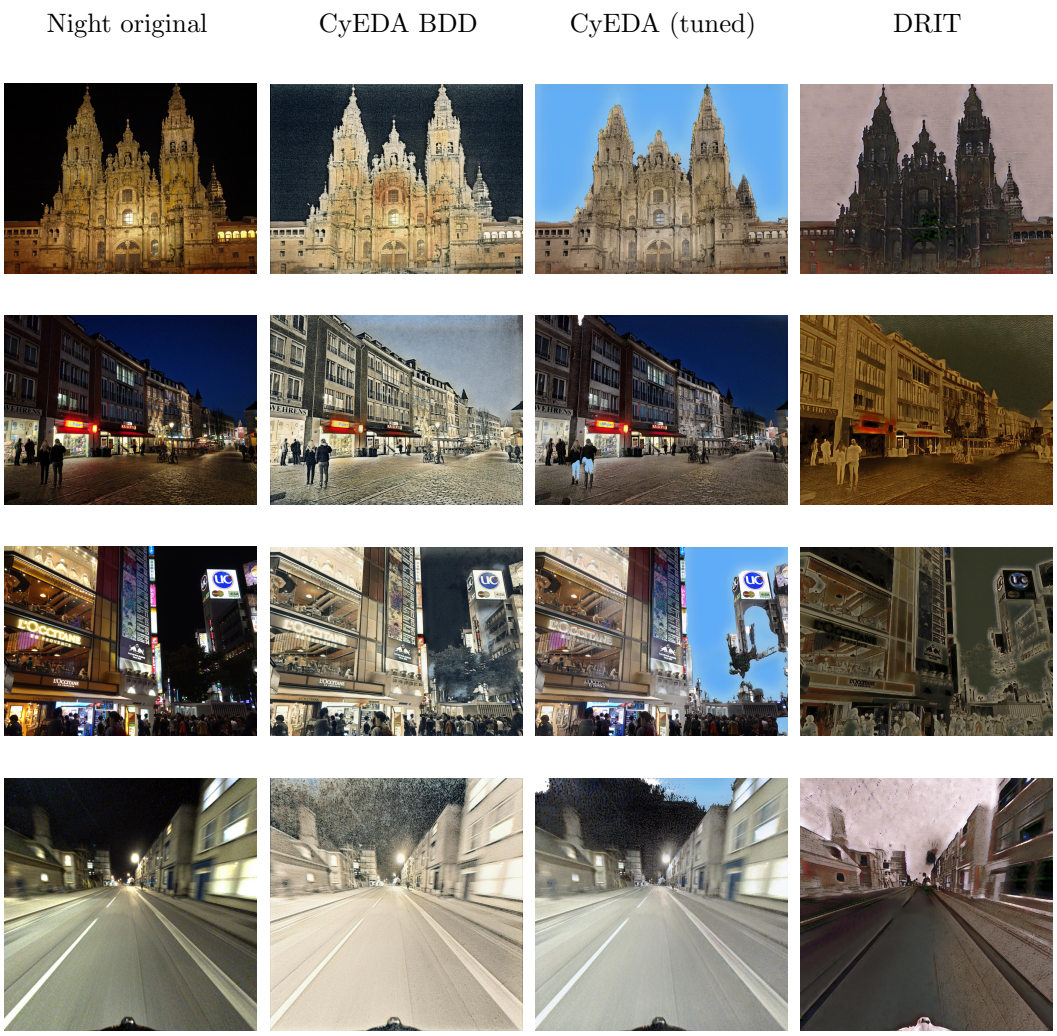


Figure B2: Examples of night images translated into the day domain by different night → day generators. The columns correspond to (left-to-right): the original image, ToDayGAN generator [1], our CycleGAN generator, the original image, CyEDA [2] generator trained on BDD dataset [19], CyEDA generator trained on *SfM*120k [12] and tuned by us, and our DRIT generator. The rows show example images from datasets (top-to-bottom) *SfM-N/D* [4], *Aachen* [14, 15], *Tokyo* [17], and *RobotCar* [8], respectively.

3 Appendix C

Examples of the evolution of light condition invariance.

In Figure C1, a training data tuple is visualized. The negatives are images of a different landmark that are the most similar to the translated anchor. Negatives are mined (re-computed) in the beginning of each epoch, so as the network trains, the mined negatives change - mined negatives shift from appearance-similar dark images to content-similar images of any domain.

In Figure C2, in each example, there are three images from the *SfM-N/D* [4] validation set: night anchor, day positive, and mined negative. The distance between the anchor and positive as well as the anchor and negative is plotted - increasing the gap between these two distances translates to easier distinction of the day positive from the night negative.

In epoch 0, the embedding network starts from ImageNet weights [13] and has never seen a night image. Therefore the two dark images are deemed similar, measured in L2 distance of their embeddings, while the images of the same objects under different light conditions are less similar. Later in the training with HED$^{\text{N}}$GAN augmentation, the distance between the images starts reflecting actual visual similarity rather than the similarity of image illumination.


Anchor
Translated anchor
Positive
Negatives mined at:
epoch 0
epoch 10
epoch 20

(continues)

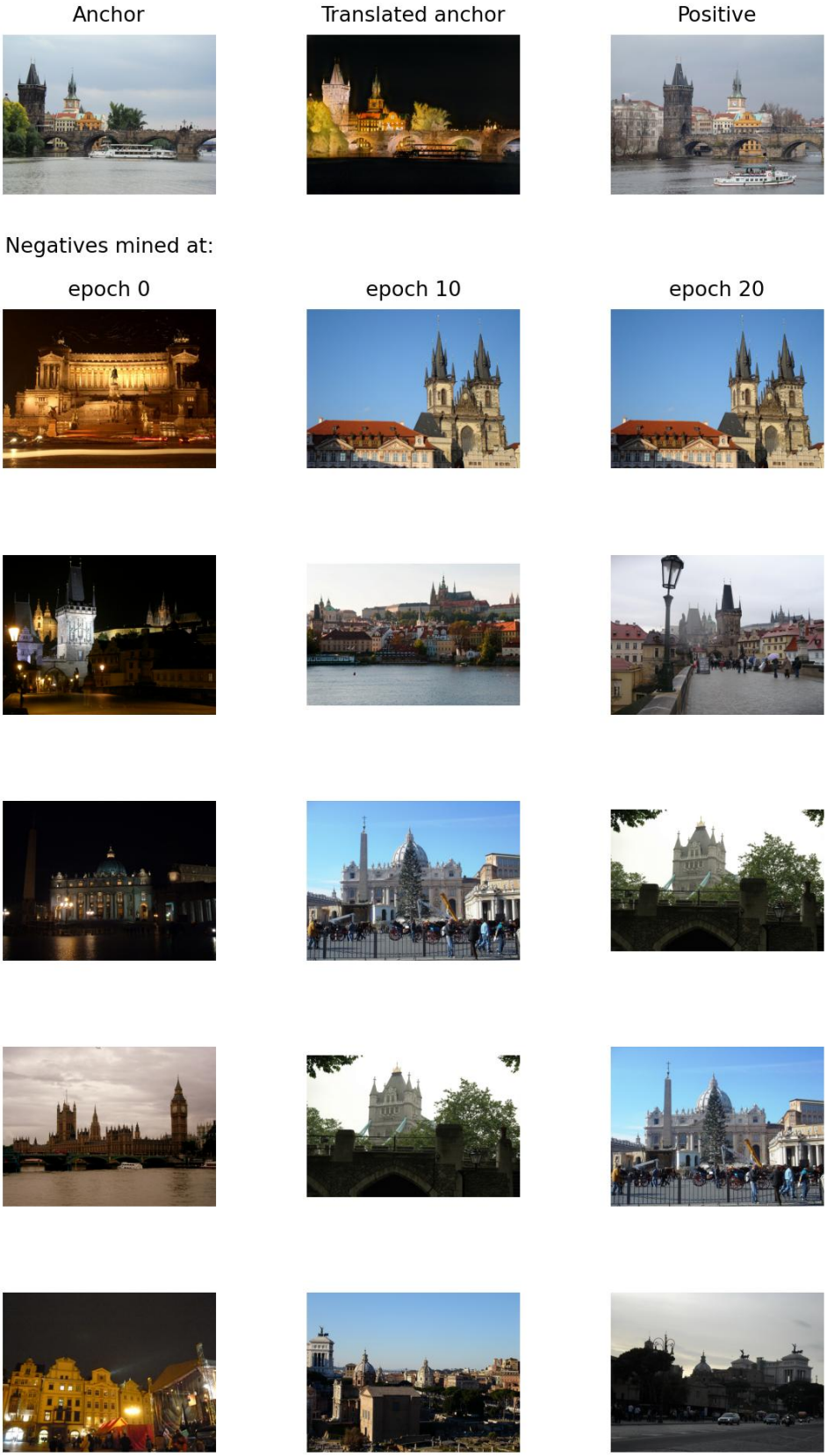


Figure C1: The visualization of hard-negative mining during the training of our method. The anchor image is translated with our HED$^{\text{N}}$GAN generator into the night domain. Five hardest negative examples are mined for each translated anchor at the beginning of each epoch. We provide mined examples at epoch 0, 10, and 20 for two different anchors. Notice the shift from negatives that are similar to the translated anchor in appearance to negatvies that are similar in content.

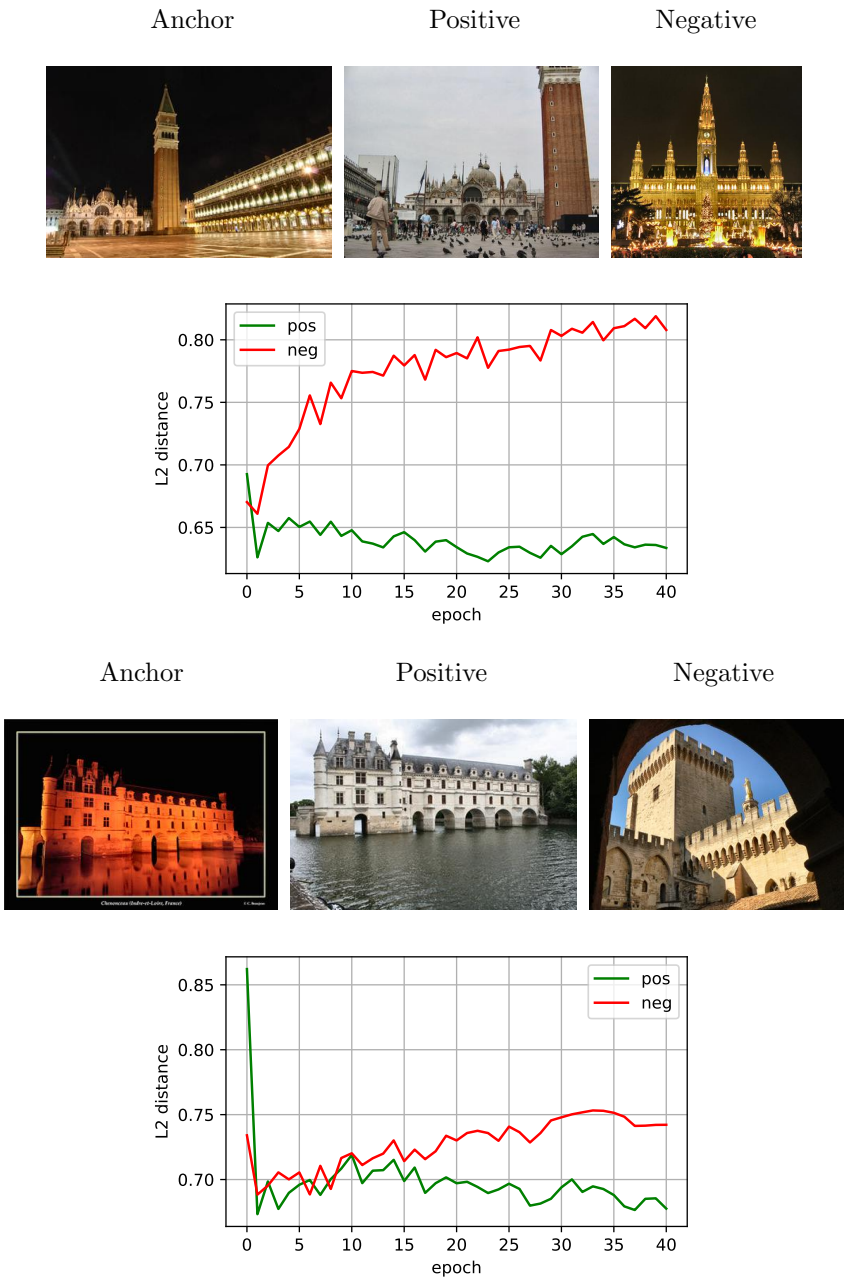


Figure C2: Two examples of image triplets: night anchor, day positive and mined negative (from left to right). Under each triplet, the L2 distance of anchor-positive (green) and anchor-negative (red) is plotted against the current training epoch. Notice the expanding gap between these two distances as the training progresses.

References

[1] Asha Anoosheh, Torsten Sattler, Radu Timofte, Marc Pollefeys, and Luc Van Gool. Night-to-day image translation for retrieval-based localization. In *ICRA*, 2019.

[2] Jing Chong Beh, Kam Woh Ng, Jie Long Kew, Che-Tsung Lin, Chee Seng Chan, Shang-Hong Lai, and Christopher Zach. Cyeda: Cycle-object edge consistency domain adaptation. In *ICIP*. IEEE, 2022.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015.

[4] Tomas Jenicek and Ondrej Chum. No fear of the dark: Image retrieval under varying illumination conditions. In *ICCV*, 2019.

[5] Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. Drit++: Diverse image-to-image translation via disentangled representations. *IJCV*, 2020.

[6] Attila Lengyel, Sourav Garg, Michael Milford, and Jan C. van Gemert. Zero-shot day-night domain adaptation with a physics prior. In *ICCV*, 2021.

[7] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Jia-Wang Bian, Le Zhang, Xiang Bai, and Jinhui Tang. Richer convolutional features for edge detection. *IEEE TPAMI*, 2019.

[8] Will Maddern, Geoff Pascoe, Chris Linegar, and Paul Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research (IJRR)*, 2017.

[9] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *ECCV*. Springer, 2020.

[10] Filip Radenović, Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondřej Chum. Revisiting Oxford and Paris: Large-scale image retrieval benchmarking. In *CVPR*, 2018.

[11] Filip Radenovic, Giorgos Tolias, and Ondřej Chum. Deep shape matching. In *ECCV*, 2018.

[12] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning cnn image retrieval with no human annotation. *IEEE TPAMI*, 2018.

[13] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 2015.

[14] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Fredrik Kahl, and Tomas Pajdla. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. In *CVPR*, 2018.

[15] Torsten Sattler, Tobias Weyand, Bastian Leibe, and Leif Kobbelt. Image Retrieval for Image-Based Localization Revisited. In *BMVC*, 2012.

[16] Giorgos Tolias, Tomas Jenicek, and Ondřej Chum. Learning and aggregating deep local descriptors for instance-level recognition. In *ECCV*. Springer, 2020.

[17] Akihiko Torii, Relja Arandjelović, Josef Sivic, Masatoshi Okutomi, and Tomas Pajdla. 24/7 place recognition by view synthesis. In *CVPR*, 2015.

[18] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *ICCV*, 2015.

[19] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *CVPR*, 2020.

[20] Zichao Zhang, Torsten Sattler, and Davide Scaramuzza. Reference pose generation for long-term visual localization via learned features and view synthesis. *IJCV*, 2021.

[21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.