EMLight: Lighting Estimation via Spherical Distribution Approximation

Fangneng Zhan

Appendix Outline

In these appendices, we present more details and experimental results which includes: details of network structure and training setting, spatially-varying illumination prediction, and more qualitative results.

Neural Projector and Training setting

Lighting estimation is a classic challenge in computer vision and computer graphics, and it is critical for realistic relighting in objects insertion and image synthesis (Lalonde, Efros, and Narasimhan 2012; Barron and Malik 2013; Hold-Geoffroy et al. 2017; Murmann et al. 2019; Zhan and Zhang 2020; Boss et al. 2020; Zhan et al. 2021c,e, 2020a,b).

Empowered by the image-to-image translation (Wang et al. 2018,?; Choi et al. 2020; Park et al. 2019; Tang et al. 2019; Zhan et al. 2021f; Zhu et al. 2020; Zhan et al. 2021d; Zhan, Zhu, and Lu 2019; Zhan et al. 2021a,b; Zhan, Xue, and Lu 2019; Zhan et al. 2022b,a; Zhan, Lu, and Xue 2018), we employ the SPADE (Park et al. 2019) as the architecture of our neural projector. The detailed architectures of Generator (including the Fusion Block and SConv Block), Discriminator and Encoder are shown in Figs. 4, 5, and 6, respectively. Spherical convolution (Coors, Condurache, and Geiger 2018) is adopted in Generator and Discriminator, normal convolution is adopted in the Encoder since the input to the Encoder is a normal image.

The proposed EMLight is implemented by the PyTorch framework. The Adam is adopted as optimizer which employs a learning rate decay mechanism (initial learning rate is 0.001). The network is trained in 100 epochs with a batch size of 4. In addition, the network training is performed on two NVIDIA Tesla P100 GPUs with 16GB memory.

Anchor points

The anchor points on unit sphere are generated by the Vogel's method (Vogel 1979). We visualize 128 anchor points (default setting) as shown in Fig. 1.

Spatially-varying Illumination

It is quite an engineering problem to achieve spatially-varying illumination in the Laval Indoor dataset (Gardner et al. 2017). We therefore include the analysis of spatially-varying illumination in the Supplementary Material. The spatially-varying illumination cannot be predicted directly as there is no ground truth of spatially-varying illumination in the Laval Indoor dataset (Gardner et al. 2017). The prior research (Gardner et al. 2019) estimates depth parameters of light sources to achieve the effect of spatially-varying illumination. We followed similar idea and included a branch in the Regression Network to predict the depth of N anchor points (128 in this case) as illustrated in Fig. 2.

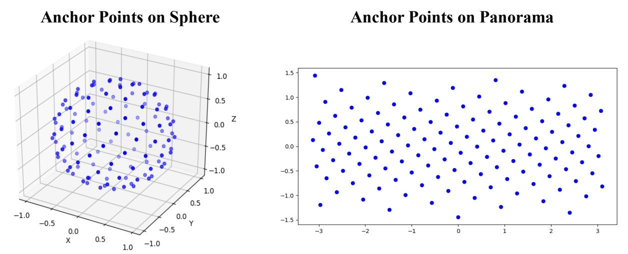

Figure 1: Visualization of the anchor points (number=128) on a unit sphere and panorama.

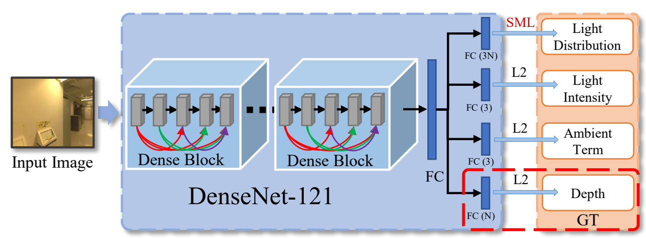

Figure 2: The Regression Network with a branch to predict the depth of N anchor points as highlighted by red box.

The Gaussian map is constructed through spherical Gaussian function as follows:

$$M = \sum_{i=1}^{N} v_i * exp\frac{d_i * u - 1}{s} + A \tag{1}$$

When we move the insertion position by ∇d, the new direction of the anchor point i can be denoted by $d_i + \nabla d$.

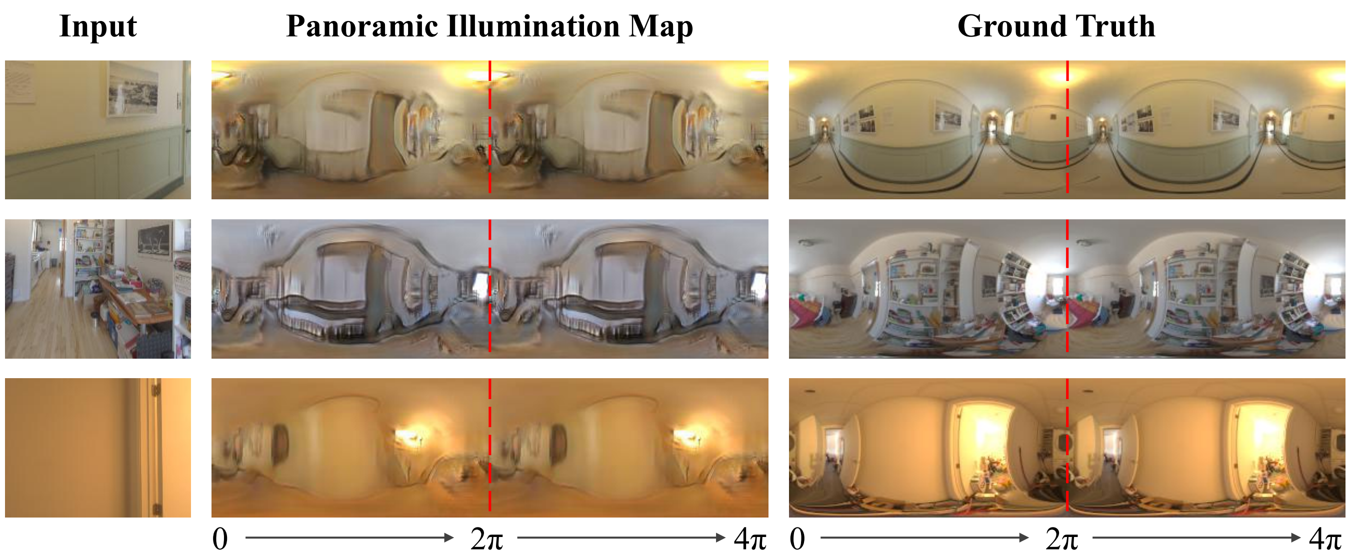

Figure 3: Illustration of the generated Panoramic Illumination Maps by our proposed neural projector.

The depth of original insertion position and the new position are l_i and $l_i + \nabla l$, which can be obtained from the predicted depth value of N anchor points. The light intensity in the new insertion position can thus be approximated by $v_i * \frac{l_i}{l_i+\nabla l}$, and the Gaussian map M_∇ of new insertion position can be constructed as follows:

$$M_\nabla = \sum_{i=1}^{N} v_i(\frac{l_i}{l_i + \nabla l}) * exp\frac{(d_i + \nabla d) * u - 1}{s} + A \quad (2)$$

The Gaussian map is then fed to the following generation network.

Qualitative results

With spherical convolution adopted for image generation, the proposed neural projector is able to synthesize panoramic illumination map as illustrated in Fig. 3.

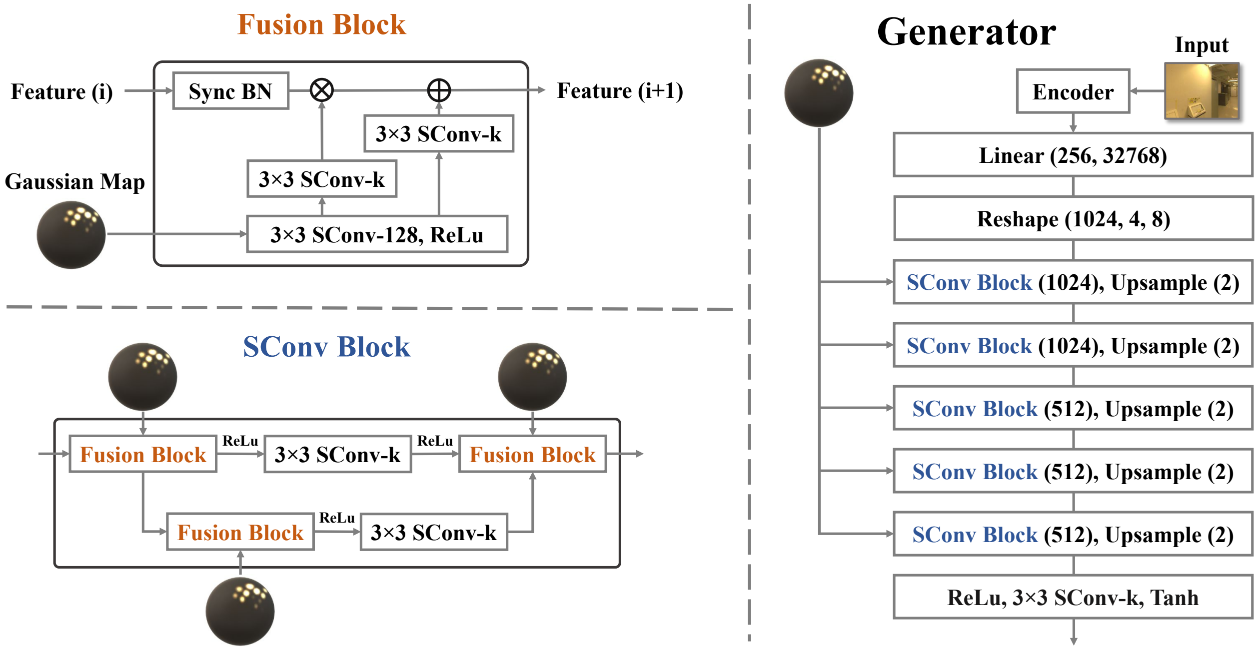

Figure 4: Detailed structures of the Fusion Block, SConv Block, and the full structure of generator: 'SConv' denotes spherical convolution; 'Sync BN' denotes synchronized batch normalization.

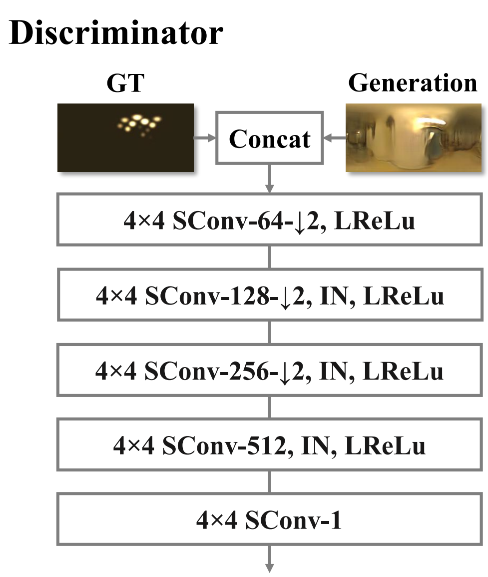

Figure 5: The architecture of the Discriminator in EMLight.

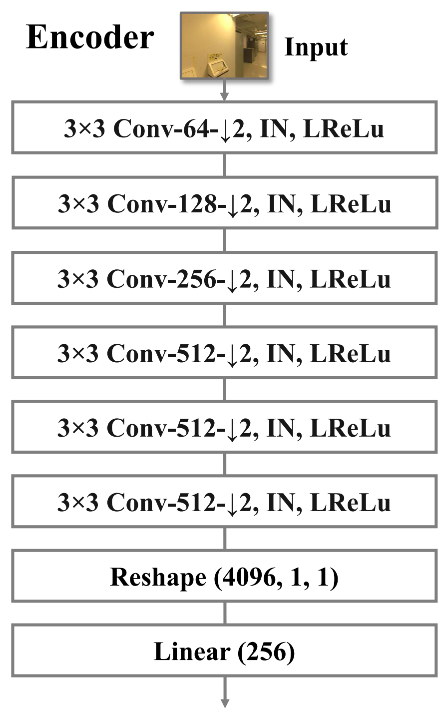

Figure 6: The architecture of the Encoder to produce latent feature vector from the input image.

References

Barron, J. T.; and Malik, J. 2013. Intrinsic scene properties from a single rgb-d image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 17–24.

Boss, M.; Jampani, V.; Kim, K.; Lensch, H.; and Kautz, J. 2020. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3982–3991.

Choi, Y.; Uh, Y.; Yoo, J.; and Ha, J.-W. 2020. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8188–8197.

Coors, B.; Condurache, A. P.; and Geiger, A. 2018. Spherenet: Learning spherical representations for detection and classification in omnidirectional images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 518–533.

Gardner, M.-A.; Hold-Geoffroy, Y.; Sunkavalli, K.; Gagné, C.; and Lalonde, J.-F. 2019. Deep parametric indoor lighting estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7175–7183.

Gardner, M.-A.; Sunkavalli, K.; Yumer, E.; Shen, X.; Gambaretto, E.; Gagné, C.; and Lalonde, J.-F. 2017. Learning to predict indoor illumination from a single image. *arXiv preprint arXiv:1704.00090* .

Hold-Geoffroy, Y.; Sunkavalli, K.; Hadap, S.; Gambaretto, E.; and Lalonde, J.-F. 2017. Deep outdoor illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7312–7321.

Lalonde, J.-F.; Efros, A. A.; and Narasimhan, S. G. 2012. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision* 98(2): 123–145.

Murmann, L.; Gharbi, M.; Aittala, M.; and Durand, F. 2019. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4080–4089.

Park, T.; Liu, M.-Y.; Wang, T.-C.; and Zhu, J.-Y. 2019. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2337–2346.

Tang, H.; Xu, D.; Sebe, N.; Wang, Y.; Corso, J. J.; and Yan, Y. 2019. Multi-channel attention selection gan with cascaded semantic guidance for cross-view image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2417–2426.

Vogel, H. 1979. A better way to construct the sunflower head. *Mathematical biosciences* 44(3-4): 179–189.

Wang, T.-C.; Liu, M.-Y.; Zhu, J.-Y.; Tao, A.; Kautz, J.; and Catanzaro, B. 2018. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8798–8807.

Zhan, F.; Lu, S.; and Xue, C. 2018. Verisimilar image synthesis for accurate detection and recognition of texts in scenes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 249–266.

Zhan, F.; Lu, S.; Zhang, C.; Ma, F.; and Xie, X. 2020a. Adversarial image composition with auxiliary illumination. In *Proceedings of the Asian Conference on Computer Vision*.

Zhan, F.; Lu, S.; Zhang, C.; Ma, F.; and Xie, X. 2020b. Towards realistic 3d embedding via view alignment. *arXiv preprint arXiv:2007.07066* .

Zhan, F.; Xue, C.; and Lu, S. 2019. GA-DAN: Geometry-aware domain adaptation network for scene text detection and recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, 9105–9115.

Zhan, F.; Yu, Y.; Cui, K.; Zhang, G.; Lu, S.; Pan, J.; Zhang, C.; Ma, F.; Xie, X.; and Miao, C. 2021a. Unbalanced Feature Transport for Exemplar-based Image Translation. In *Proceedings of the IEEE International Conference on Computer Vision*.

Zhan, F.; Yu, Y.; Wu, R.; Cui, K.; Xiao, A.; Lu, S.; and Shao, L. 2021b. Bi-level feature alignment for versatile image translation and manipulation. *arXiv preprint arXiv:2107.03021* .

Zhan, F.; Yu, Y.; Wu, R.; Zhang, C.; Lu, S.; Shao, L.; Ma, F.; and Xie, X. 2021c. GMLight: Lighting Estimation via Geometric Distribution Approximation. *arXiv preprint arXiv:2102.10244* .

Zhan, F.; Yu, Y.; Wu, R.; Zhang, J.; and Lu, S. 2021d. Multimodal Image Synthesis and Editing: A Survey. *arXiv preprint arXiv:2112.13592* .

Zhan, F.; Yu, Y.; Wu, R.; Zhang, J.; Lu, S.; and Zhang, C. 2022a. Marginal Contrastive Correspondence for Guided Image Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10663–10672.

Zhan, F.; and Zhang, C. 2020. Spatial-Aware GAN for Unsupervised Person Re-identification. *Proceedings of the International Conference on Pattern Recognition* .

Zhan, F.; Zhang, C.; Hu, W.; Lu, S.; Ma, F.; Xie, X.; and Shao, L. 2021e. Sparse needlets for lighting estimation with spherical transport loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12830–12839.

Zhan, F.; Zhang, C.; Yu, Y.; Chang, Y.; Lu, S.; Ma, F.; and Xie, X. 2021f. EMLight: Lighting Estimation via Spherical Distribution Approximation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3287–3295.

Zhan, F.; Zhang, J.; Yu, Y.; Wu, R.; and Lu, S. 2022b. Modulated Contrast for Versatile Image Synthesis. *arXiv preprint arXiv:2203.09333* .

Zhan, F.; Zhu, H.; and Lu, S. 2019. Spatial fusion gan for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3653–3662.

Zhu, Z.; Xu, Z.; You, A.; and Bai, X. 2020. Semantically Multi-modal Image Synthesis. In *Proceedings of* *the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5467–5476.