GMLight: Lighting Estimation via Geometric Distribution Approximation: Supplementary Material

Fangneng Zhan

I. Appendix Outline

In these appendices, we present more details of the network structure.

II. Neural Projector

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 [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

Empowered by the image-to-image translation [11], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], we employ the SPADE [13] 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. 1, 2, and 3, respectively. Spherical convolution [25] is adopted in Generator and Discriminator, normal convolution is adopted in the Encoder since the input to the Encoder is a normal image. The detailed architectures of Generator, Discriminator and Encoder are shown in Figs. 1, 2, and 3, respectively.

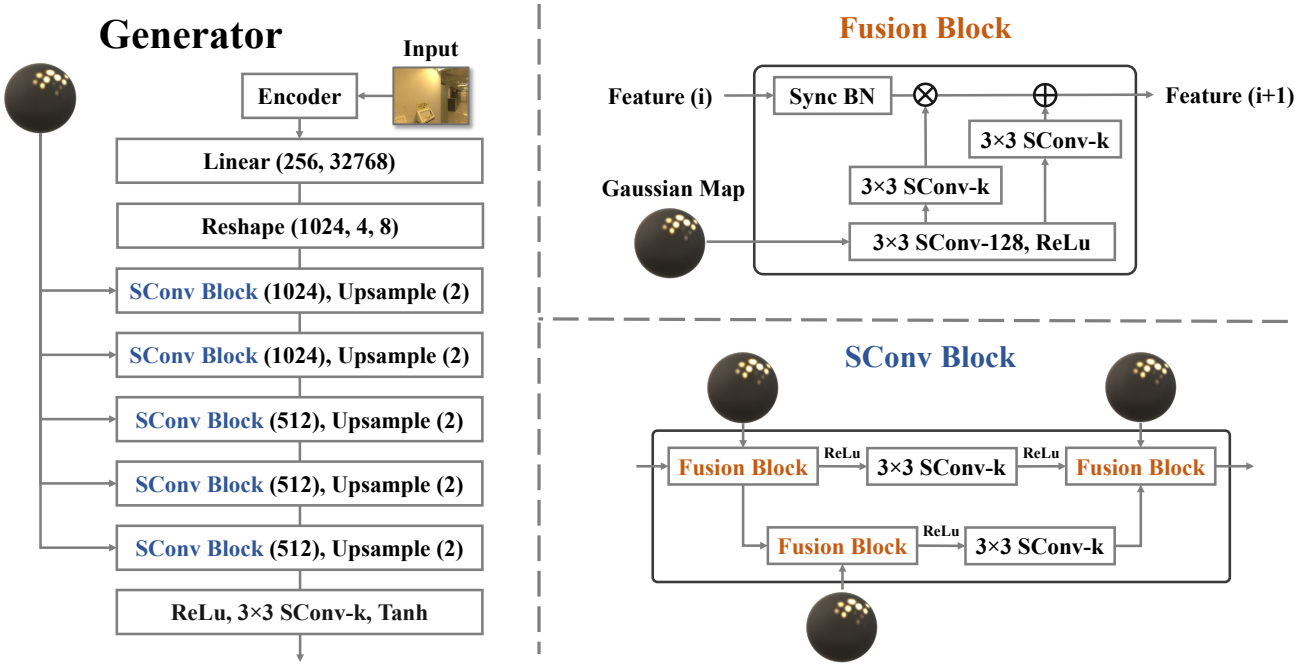

Fig. 1. Detailed structures of the Fusion Block, SConv Block, and the full structure of generator: 'SConv' denotes spherical convolution; 'Sync BN' denotes synchronized batch normalization.

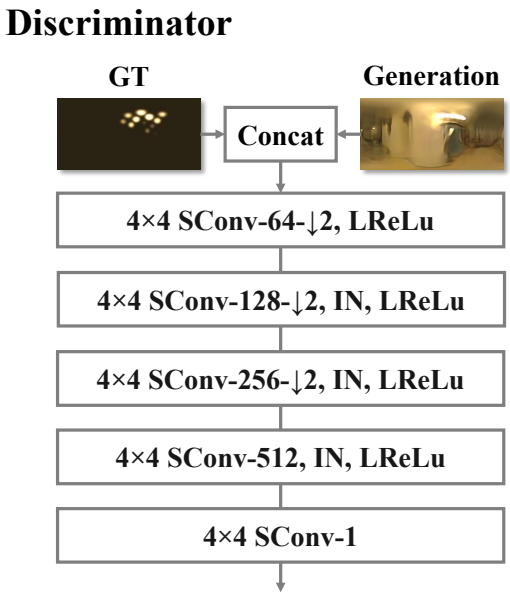

Fig. 2. The architecture of the Discriminator in EMLight.

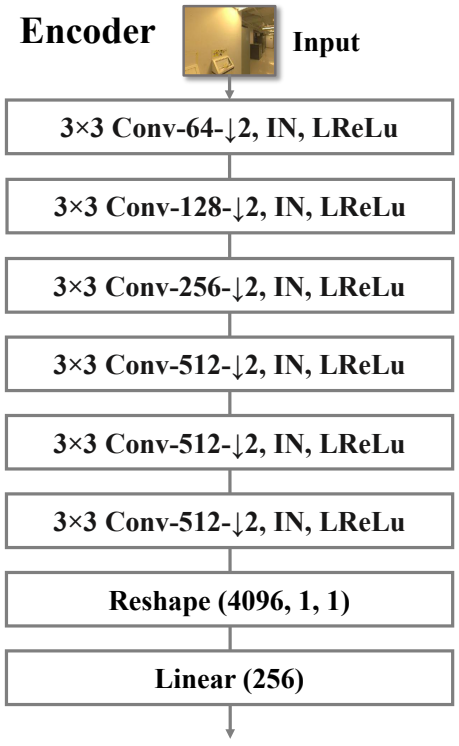

Fig. 3. The architecture of the Encoder to produce latent feature vector from the input image.

REFERENCES

[1] J. F. Lalonde, A. A. Efros, and S. G. Narasimhan, "Estimating the natural illumination conditions from a single outdoor image," *IJCV*, 2012.

[2] J. T. Barron and J. Malik, "Intrinsic scene properties from a single rgb-d image," *TPAMI*, 2015.

[3] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde, "Deep outdoor illumination estimation," in *CVPR*, 2017.

[4] L. Murmann, M. Gharbi, M. Aittala, and F. Durand, "A dataset of multi-illumination images in the wild," in *ICCV*, 2019.

[5] F. Zhan and C. Zhang, "Spatial-aware gan for unsupervised person re-identification," *Proceedings of the International Conference on Pattern Recognition*, 2020.

[6] M. Boss, V. Jampani, K. Kim, H. P. Lensch, and J. Kautz, "Two-shot spatially-varying brdf and shape estimation," in *CVPR*, 2020.

[7] F. Zhan, Y. Yu, C. Zhang, R. Wu, W. Hu, S. Lu, F. Ma, X. Xie, and L. Shao, "Gmlight: Lighting estimation via geometric distribution approximation," *IEEE Transactions on Image Processing*, vol. 31, pp. 2268–2278, 2022.

[8] F. Zhan, C. Zhang, W. Hu, S. Lu, F. Ma, X. Xie, and L. Shao, "Sparse needlets for lighting estimation with spherical transport loss," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 830–12 839.

[9] F. Zhan, S. Lu, C. Zhang, F. Ma, and X. Xie, "Adversarial image composition with auxiliary illumination," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[10] ——, "Towards realistic 3d embedding via view alignment," *arXiv preprint arXiv:2007.07066*, 2020.

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

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

[13] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *CVPR*, 2019.

[14] H. Tang, D. Xu, N. Sebe, Y. Wang, J. J. Corso, and Y. Yan, "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*, 2019, pp. 2417–2426.

[15] F. Zhan, C. Zhang, Y. Yu, Y. Chang, S. Lu, F. Ma, and X. Xie, "Emlight: Lighting estimation via spherical distribution approximation," *arXiv preprint arXiv:2012.11116*, 2020.

[16] Z. Zhu, Z. Xu, A. You, and X. Bai, "Semantically multi-modal image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5467–5476.

[17] F. Zhan, Y. Yu, R. Wu, J. Zhang, and S. Lu, "Multimodal image synthesis and editing: A survey," *arXiv preprint arXiv:2112.13592*, 2021.

[18] F. Zhan, H. Zhu, and S. Lu, "Spatial fusion gan for image synthesis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 3653–3662.

[19] F. Zhan, Y. Yu, K. Cui, G. Zhang, S. Lu, J. Pan, C. Zhang, F. Ma, X. Xie, and C. Miao, "Unbalanced feature transport for exemplar-based image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

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

[21] F. Zhan, C. Xue, and S. Lu, "Ga-dan: Geometry-aware domain adaptation network for scene text detection and recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 9105–9115.

[22] F. Zhan, J. Zhang, Y. Yu, R. Wu, and S. Lu, "Modulated contrast for versatile image synthesis," *arXiv preprint arXiv:2203.09333*, 2022.

[23] F. Zhan, Y. Yu, R. Wu, J. Zhang, S. Lu, and C. Zhang, "Marginal contrastive correspondence for guided image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 663–10 672.

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

[25] B. Coors, A. P. Condurache, and A. Geiger, "Spherenet: Learning spherical representations for detection and classification in omnidirectional images," in *ECCV*, 2018.