Ambient Lighting Generation for Flash Images with Conditional Adversarial Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

Sometimes the level of light in a scene is very low and insufficient to get a properly digitization of an image. To cope with this situation, photographers usually use the camera flash. However, in flash images, we have very dark and very bright areas, and side-way shadows. To give flash images a more natural illumination, the present work proposes an end-to-end architecture based on a conditional adversarial network for generating synthetic ambient lighting. Our proposed network not only normalizes the illumination of flash images but also removes the side-way shadows of them simultaneously generating synthetic ambient shadows, thus, giving a more natural appearance to the image.

1 Introduction

We can handle a low illumination in a scene with an external device such as a camera flash, thus, creating flash images, which probably do not have uniform illumination. In contrast, in an ambient image, the illumination is more natural and uniform, because the available light can be more evenly distributed. Therefore, researchers have studied the conversion from flash images to ambient images. Capece et al. [2] proposed a network for translating flash images to image with studio lighting. This network normalizes the illumination of portraits images with a green background, but without replicating the natural skin tone of people. However, in a real scenario, objects away from the camera will have very low illumination, this creates dark areas in the image, considering that there is only the illumination of the camera flash. Furthermore, objects very close to the camera present more brightness. In this situation, the model should fill this area using the information of the closest pixels. The side-way shadows of an object in a flash image should be removed because they sometimes cover significant areas behind it. Finally, the model should be able to generate synthetic shadows, as a part of the simulation of a natural light source.

One approach is to perform a guided image enhancement. Guo et al. [4] introduced an enhancement by estimating an illumination map. More specific, the illumination map of each pixel is first estimated individually by finding the maximum value in the R, G and B channels, then the illumination map is refined by imposing a structure prior. The structure of the illumination map can be defined based on the flash image, so that illuminate the dark areas more than the bright areas. However, the model was not designed for removing shadows or reconstruct overexposed areas.

2 Proposal

In this work, we proposed a cGAN [6], where the discriminator is based on Isola et al. [5]. The generator has as an encoder all the convolutional layers of the VGG-16[7] architecture pre-trained on the ImageNet dataset[3]. The generator models the translation from flash images to ambient images. While the discriminator distinguishes which images contain natural lighting and which do not.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

3 Experiments

In order to validate our proposal, we use images from FAID(Flash and Ambient Illuminations Dataset)[1], which is a collection of 2775 pairs of flash and ambient images. Inspecting each image in this dataset, some ambient images have problems such as low illumination, shadows from external objects, and even artifacts. Therefore, we manually select from the FAID dataset, ambient images that did not present these kinds of problems. Table 1 shows the distribution of the dataset for training and testing.

Table 1: Distribution of the training and test dataset.

	Training data	Test data	Total data
Pairs of images	969	116	1085

The results for the LIME [4] method are estimated by applying it to each flash image in the test set. Table 2 reports the quantitative comparison using the mean PSNR(Peak Signal-to-Noise Ratio) and the mean SSIM(The Structural Similarity).

Table 2: Comparing the mean PSNR/SSIM with LIME [4].

	LIME	Ours
PSNR/SSIM	12.36/0.626	15.37/0.692

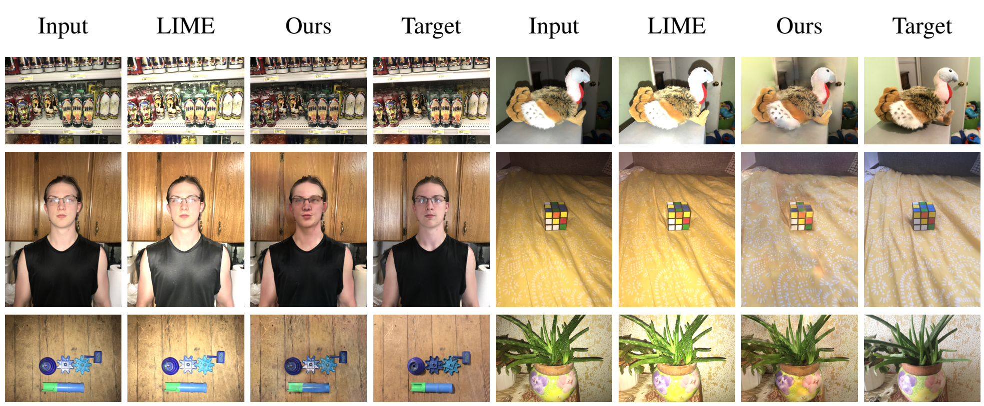

Figure 1: Qualitative comparison. Our results present more natural illumination.

Figure 1 shows that our model can distinguish between an object with a natural dark color and an area with low illumination. The illumination map created by LIME can not distinguish the natural color of an object. For example, the sleeveless t-shirt of the man in Figure 1 is well illuminated, but the LIME method can not detect it, so it illuminates, even more, all this area. Moreover, this method tends to overexpose the flash image instead of normalizing the illumination. All results of the LIME method still present the shadows and the very bright areas of the flash images.

4 Conclusions

Comparing different network architectures, and training them from scratch. The pre-trained model is better than the model trained from scratch, due to the size of the training data. The proposed architecture, although it is not able to completely remove the shadows or reconstruct the very bright areas, out model generates synthetic shadows on e.g., shelves and also generates and also generates a more natural and uniform illumination.

References

[1] Yagiz Aksoy, Changil Kim, Petr Kellnhofer, Sylvain Paris, Mohamed Elgharib, Marc Pollefeys, and Wojciech Matusik. A dataset of flash and ambient illumination pairs from the crowd. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 634–649, 2018.

[2] Nicola Capece, Francesco Banterle, Paolo Cignoni, Fabio Ganovelli, Roberto Scopigno, and Ugo Erra. Deepflash: Turning a flash selfie into a studio portrait. *Signal Processing: Image Communication*, 77:28 – 39, 2019. ISSN 0923-5965. doi: https://doi.org/10.1016/j.image.2019.05.013. URL `http://www.sciencedirect.com/science/article/pii/S0923596519300451`.

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[4] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, Feb 2017. ISSN 1057-7149. doi: 10.1109/TIP.2016.2639450.

[5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[6] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.