Supplementary Materials for Deep Image-based Illumination Harmonization

Zhongyun Bao[1], Chengjiang Long[2*], Gang Fu[1], Daquan Liu[1], Yuanzhen Li[1], Jiaming Wu[1], Chunxia Xiao[1*†]
[1]School of Computer Science, Wuhan University, Wuhan, Hubei, China
[2] Meta Reality Labs, Burlingame, CA, USA
clong1@fb.com, xyzgfu@gmail.com, {zhongyunbao, daquanliu, cxxiao}@whu.edu.cn

Abstract

In this supplementary material, we first introduce more detailed motivation clarification in Section 1, then provide the statistical analysis of distribution properties of our IH dataset in Section 2, and finally present more visual results for real-world images in Section 3.

1. Motivation

Integrating a foreground object into a background scene with illumination harmonization is an important but challenging task in computer vision and augmented reality community. However, to our best knowledge, existing methods related our task suffer from two issues. First, the performance of traditional methods heavily depends upon the accuracy of the estimated geometry, illumination, and light source. They usually produce poor results due to inaccurate estimation of geometry and illumination of the object or scene. Second, they generally take long time to process a single image, due to involving many intermediate processing steps. Besides, learning-based methods mainly focus on foreground and background appearance consistency or the foreground object shadow generation, which rarely consider global appearance and illumination harmonization. To overcome these issues, first, we construct a large-scale high-quality dataset, IH dataset, which contains diverse illumination situations, for the research community. Then we propose a learning-based framework named DIH-GAN, which can consider global appearance and illumination harmonization and directly infer the mapping relationship between the inserted image-based foreground object and real-world background in an image-to-image translation style without complex but unreliable inverse rendering. Also, our DIH-GAN framework is significantly faster than traditional methods.

*This work was co-supervised by Chengjiang Long and Chunxia Xiao.
†Corresponding author.

2. Dataset Analysis

To generate our synthesized images, we use Blender to render 3D object. We add a plane at the bottom of the 3D object for casting shadow, and the corresponding HDR panorama is used to render the object. Besides, we use Photoshop to manually annotate each foreground object in our dataset to obtain accurate mask. We list a part of our 3D objects used for our dataset construction in Figure 1. Note that, our rendering method is not only for the objects mentioned in the paper, researchers can use it to render other models according to their needs.

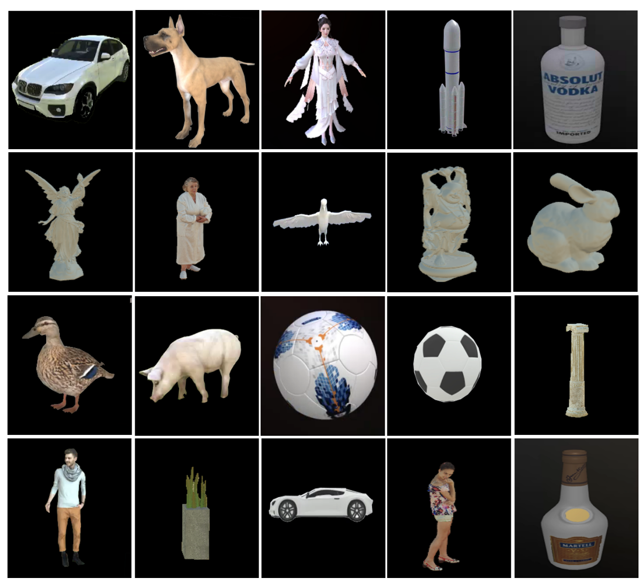

Figure 1. The part of our 3D objects in our IH dataset.

Besides we also downloaded various 3D objects on the Internet, including characters, buildings, etc. to further expand our dataset.

Figure 2 shows statistical analysis of our dataset. As we can see, we provide the statistics on the ratio of virtual objects, real objects (occluders), real world shadows, and illumination. The area distribution is expressed as the ratio between the target (shadows, occluders or virtual objects)

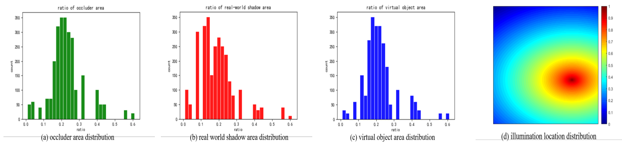

Figure 2. Statistics on the ratio of virtual objects, real object, real shadow, and illumination.

area and image area. We observe that the majority of occluders falls in range of $(0.1, 0.4]$, the majority of shadows falls in range of $(0.1, 0.3]$ and the majority of virtual objects falls in range of $(0.1, 0.3]$. Since shadow clues falling in $0.08, 0.3]$, it is difficult to insert the virtual object. Our dataset almost has no such case. In addition, we analyze the spatial distribution of scene illumination in our dataset, we compute a probability map (Figure 2 (d)) to show how likely a pixel belongs to the illumination range. As can be seen, the illumination tend to cluster around the lower center of the image, since inserted image-based objects are usually placed approximately around the human eyesight.

To clarify, although we take denoising operation in Blender by choosing the reasonable denoising parameters during the rendering process, there might still be some noise. To ensure the high quality of the dataset for our task, we further manually filter out the images without obvious or natural illumination and conspicuous shadows. What's more, the final experiment shows that our IH dataset is suitable for our task and obtains satisfactory results as expected.

To make the quality measurement more solid, besides statistical analysis, we randomly select 1000 rendered images from the IH dataset and 1000 real images from the SOBA dataset [2] to conduct a user study by asking 100 participants whether an image is realistic and illumination harmonious. The result shows that 58.1% of rendered images from the IH dataset are judged to be "Yes", while the percentage of the SOBA real images is 61.7%. This strongly demonstrates the high quality of rendered images in our IH dataset.

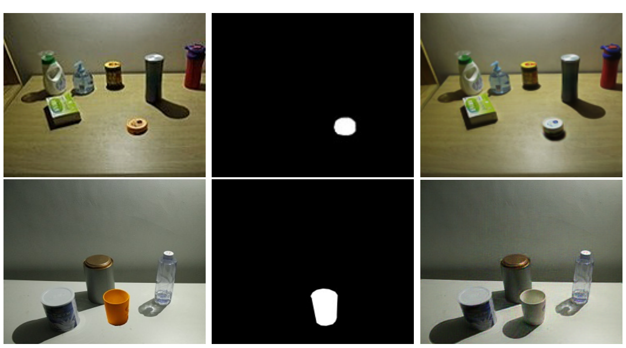

Figure 3. The visual results of DIH-GAN on three real-world images. From left to right: input images without illumination harmony, object masks, and output results with illumination harmony.

3. More Visualization Results of Real-World Images

To provide some visualization results of real-world images, we collect some real-world scene images and objects with different illumination conditions through self-photographed images, existing images on the Internet and some public images provided by Liu *et al.* [1] to test our DIH-GAN. As shown in Figure 3, our proposed DIH-GAN performs well on real-world images.

References

[1] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. Arshadowgan: Shadow generative adversarial network for augmented reality in single light scenes. In *CVPR*, pages 8139–8148, 2020. 2

[2] Tianyu Wang, Xiaowei Hu, Qiong Wang, Pheng-Ann Heng, and Chi-Wing Fu. Instance shadow detection. In *CVPR*, pages 1880–1889, 2020. 2