StyleLight: HDR Panorama Generation for Lighting Estimation and Editing

Guangcong Wang ⓘ, Yinuo Yang ⓘ, Chen Change Loy ⓘ, and Ziwei Liu✉ ⓘ

S-Lab, Nanyang Technological University
{guangcong.wang,ccloy,ziwei.liu}@ntu.edu.sg   YANG0689@e.ntu.edu.sg

LDR LFOV image   **HDR Panorama Generation**   **Lighting Removal**   **Lighting Addition**

Fig. 1. StyleLight. The proposed StyleLight estimates HDR panorama lighting from an LDR LFOV image and performs lighting addition and removal.

Abstract. We present a new lighting estimation and editing framework to generate high-dynamic-range (HDR) indoor panorama lighting from a single limited field-of-view (LFOV) image captured by low-dynamic-range (LDR) cameras. Existing lighting estimation methods either directly regress lighting representation parameters or decompose this problem into LFOV-to-panorama and LDR-to-HDR lighting generation subtasks. However, due to the partial observation, the high-dynamic-range lighting, and the intrinsic ambiguity of a scene, lighting estimation remains a challenging task. To tackle this problem, we propose a coupled dual-StyleGAN panorama synthesis network (**StyleLight**) that integrates LDR and HDR panorama synthesis into a unified framework. The LDR and HDR panorama synthesis share a similar generator but have separate discriminators. During inference, given an LDR LFOV image, we propose a focal-masked GAN inversion method to find its latent code by the LDR panorama synthesis branch and then synthesize the HDR panorama by the HDR panorama synthesis branch. StyleLight takes LFOV-to-panorama and LDR-to-HDR lighting generation into a unified framework and thus greatly improves lighting estimation. Extensive experiments demonstrate that our framework achieves superior performance over state-of-the-art methods on indoor lighting estimation. Notably, StyleLight also enables intuitive lighting editing on indoor HDR panoramas, which is suitable for real-world applications. Code is available at https://style-light.github.io/.

Keywords: Lighting Estimation and Editing; Panorama Generation

1   Introduction

Environmental lighting models aim to approximate realistic lighting effects in a scene. Existing state-of-the-art methods [5,11,19,25] use HDR spherical panoramas for lighting. In many real-world applications, however, only LDR LFOV images are available. Therefore, it is important to estimate HDR panorama illumination from an LDR LFOV image, aided with the capability of lighting editing to flexibly control lighting conditions. Lighting estimation and editing have a wide range of applications in mixed reality such as object insertion and relighting in virtual meetings and games. For example, given an LDR LFOV background, a person can be inserted into the background and realistically relit by controllable panorama illumination.

There are three main challenging problems in lighting estimation. First, only a partial scene is observed in a limited FOV image. The other unobserved part of a scene is ambiguous and could significantly affect the illumination distribution in a full scene due to ambiguous lighting sources with different numbers, shapes, and directions. Second, only a partial range of lighting is observed in typical LDR images. Physical lighting has a wide range of intensity levels in real scenes, ranging from shadows to bright light sources. However, in LDR images, underexposed areas are clipped to 0, and overexposed areas are clipped to 1. Out-of-range lighting cannot be observed in LDR images. Third, lighting estimation of unobserved views depends on intrinsic attributes of observed objects (e.g., material and 3D geometry), but these annotations are often unavailable in real-world applications.

Existing methods on illumination estimation provide partial solutions to illumination estimation, which could be roughly classified into two groups, i.e., regression-based estimation methods [5,4,6,31,9,14] and generation-based estimation methods [32,23,24]. In the first group, researchers propose to directly regress panorama lighting information from a given LDR LFOV image. The main difficulty of this pipeline is to parameterize lighting representations as target labels for regression, including light representation (e.g., number, shape, directions and intensity) [4], Spherical Harmonics [6], and wavelet transformation [31]. However, these lighting labels mainly focus on strong lighting sources and thus lose the details of a scene. When a scene contains reflective materials (e.g., glass, mirror, and metal), this pipeline will fail to provide a desired rendered visual effect. In image-based rendering, the HDR map can be also applied as background. Moreover, regression-based models often suffer from unstable training of neural networks due to partial observed view, a wide range of lighting, unknown geometry, and material of objects. For example, EMlight [32] claimed that it is necessary to train a regression model by gradually increasing training examples to avoid training collapse. Second, generation-based estimation methods aim at generating full HDR panoramas from an LDR LFOV image. They typically decompose illumination estimation as several sub-tasks, such as LFOV-to-panorama completion, LDR-to-HDR regression [32], and 3D geometry prediction [23]. These methods either require additional annotations or multiple optimization stages.

Our proposed method belongs to the second group. Unlike the existing generation-based estimation methods, we propose a coupled dual-StyleGAN panorama synthesis network (**StyleLight**) that solves LFOV-to-panorama and LDR-to-HDR problems in a unified framework and does not require any 3D geometry annotation. Specifically, StyleLight is a coupled dual-StylGAN that maps noises to both HDR and LDR panoramas. The LDR and HDR panorama synthesis share a similar generator. At the output of the generator, a tone mapping is adopted for the translation between HDR and LDR. Considering the fact that an HDR panorama consists of extremely different weak lighting distribution and strong lighting distribution, we use two discriminators to distinguish real/fake HDR and LDR panorama distributions, respectively. Given an LDR LFOV image, we propose a focal-masked GAN inversion method to find its latent code through the LDR panorama synthesis branch and then generate its corresponding HDR panorama as illumination maps through the HDR synthesis branch. Moreover, we propose a structure-preserved GAN inversion method for lighting editing with a trained StyleLight model to flexibly control the environment lighting of panoramas, such as turning on/off lights, opening/closing doors and windows, making our method well-suited for various real-world applications.

The main contributions of this paper can be summarized as follows: **1)** We propose a coupled dual-styleGAN synthesis network (StyleLight) that integrates HDR and LDR panorama synthesis in a unified framework. During inference, we propose a focal-masked GAN inversion method to solve both LFOV-to-panorama and LDR-to-HDR generation. **2)** We propose a structure-preserved GAN inversion method for lighting editing with the trained StyleLight model to flexibly control panorama lighting. **3)** Extensive experiments demonstrate the superiority and the effectiveness of our proposed lighting estimation method over state-of-the-art methods on indoor HDR panoramas and show promising applications with our proposed lighting editing.

2   Related Work

Lighting estimation is one of the long-standing problems in computer vision and graphics. Early work [3,29,27] typically simplifies the models by considering fixed number of lights [17], annotated light positions [13] and geometry [13,13,16,13,18,18]. Recent lighting estimation methods relax such assumptions and focus on more challenging lighting estimation. We briefly review the work most relevant to this paper.

Lighting estimation. Existing lighting estimation methods can be roughly classified into two groups, i.e., regression-based methods and generation-based methods. Regression-based models [4,14,6,31,32,9] seek for effective lighting representations of HDR panoramas as labels for lighting regression. These methods take an LDR LFOV image as input and predict lighting labels, such as direct lighting representation (e.g., numbers, shapes, directions, and intensity) [4], Spherical Harmonics (SH) [6], and wavelet transformation [31]. For example, Gardner *et al.* [5] trained a lighting classifier on an annotated LDR panorama

dataset to predict the location of lighting sources of LDR LFOV images and fine-tuned the prediction of lighting intensity on a small HDR panorama dataset. Gardner *et al.* [4] regressed the illumination conditions such as the light source positions in 3D, areas, intensities, and colors. LeGendre *et al.* [14] designed a joint camera-sphere device to collect LDR LFOV images that include three reflective spheres, and regressed HDR lighting from an LDR LFOV image. Garon *et al.* [6] regressed the 5th-order SH coefficients for the lighting at a certain location given a single image and a 2D location in an image. Regression-based models ignore the low-intensity lighting details and thus fail to provide a desired rendered quality for mirror materials. Besides, they often suffer from unstable training of neural networks due to the high dynamic range of lighting.

Generation-based methods [5,23,24] focus on decomposing illumination estimation as several sub-tasks, such as LFOV-to-panorama completion and LDR-to-HDR regression. Song *et al.* [23] decomposed illuminations prediction into several simpler differentiable sub-tasks including geometry estimation, scene completion, and LDR-to-HDR estimation. Zhan *et al.* [32] decomposed the illumination map into spherical light distribution, light intensity and the ambient term for illumination regression. This aforementioned paradigm either decomposes illumination estimation as several sub-tasks or collects expensive annotations for training.

Image editing via GAN inversion. Image editing is a task that modifies a target attribution of a given image while preserving other details [28,7,15,21,20,33,26,22]. Recent state-of-the-art image editing methods focus on generative adversarial network (GAN) inversion techniques to control the latent space of GANs for image editing. For example, Ganspace [7] used Principal Component Analysis (PCA) to identify important latent directions and create interpretable controls for image attributes, such as viewpoint, aging, lighting, and time of day. InterfaceGAN [21] proposed to edit the latent code $\mathbf{z}$ with $\mathbf{z}_{edit} = \mathbf{z} + \alpha\mathbf{n}$ for each semantic attribute where $\mathbf{n}$ is a normal vector of a linear SVM's hyper-plane. SeFa [22] proposed a closed-form factorization algorithm for latent semantic discovery by directly decomposing the pre-trained weights. StyleFlow [1] formulated conditional exploration as an instance of conditional continuous normalizing flows in the GAN latent space conditioned by attribute features. These image editing methods focus on controlling the latent code to control attributes of objects. However, in our lighting editing setting, our goal is to edit a light source that is located at a small area of a panorama given a bounding box, which is different from the typical image editing.

3   Our Approach

3.1   Overview of the Proposed Framework

An overview of our proposed framework is depicted in **Fig. 2**. The framework consists of lighting estimation and lighting editing. The goal of lighting estimation is to predict HDR panorama from an LDR LFOV image. To achieve this
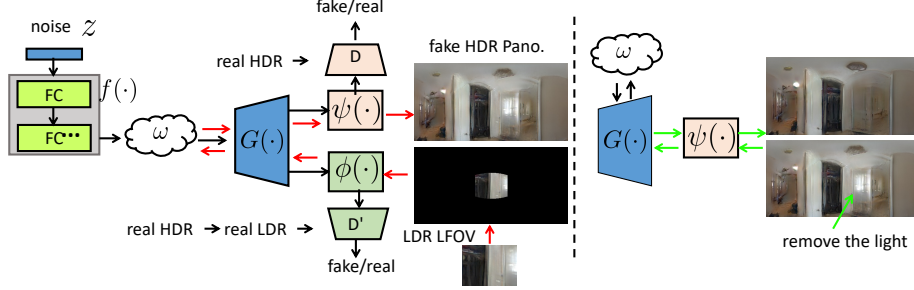
Fig. 2. Overview of the proposed framework. **Left:** Lighting estimation. We first propose a coupled dual-StyleGAN that generates HDR and LDR panoramas from noises. Given an LDR LFOV image, we map it into a panorama. We introduce a focal-masked GAN inversion method to find its latent code through the LDR branch and compute the HDR panorama through the HDR branch. **Right:** Lighting editing. With the trained dual-StyleGAN, we perform lighting editing to control panoramas lighting.

goal, we introduce a novel coupled dual-StyleGAN synthesis network (Style-Light) that learns to synthesize both HDR and LDR panoramas from noises with a shared generator and two specific discriminators. With the trained Style-Light model, we propose a focal-masked GAN inversion method to find the latent code of an LDR LFOV image and predict the HDR panorama as an illumination map (Section 3.2). To control various lighting conditions of the target HDR panorama, we introduce a structure-preserved lighting editing method to solve this problem (Section 3.3). To the best of our knowledge, this is the first attempt to perform lighting editing on HDR panoramas for controlling the scene lighting.

3.2   Lighting Estimation

Generation-based lighting estimation methods typically decompose illumination estimation into several sub-tasks, such as LFOV-to-panorama completion and LDR-to-HDR regression. These methods optimize sub-tasks independently and could lead to sub-optimal estimation results. Moreover, these methods can be hardly scalable to lighting editing. Different from these methods, we propose a coupled dual-StyleGAN that integrates LFOV-to-panorama, LDR-to-HDR, and lighting editing into a unified framework.

Coupled dual-StyleGAN for joint LDR and HDR panorama synthesis.
HDR panoramas are more difficult to synthesize than LDR panoramas due to the complex characteristics of distributions. **Figure 3** shows the distributions of two HDR panorama examples (others are similar). An HDR panorama consists of two components, i.e., low-intensity pixels (blue, denoted as LDR, from 0 to 1) and high-intensity pixels that are out of range LDR pixels (red, denoted as HDR^-, larger than 1). The distributions suggest three key points. First, LDR distributions are similar to a Gaussian distribution. Most of the pixels are LDR pixels and they are limited in the narrow range. This suggests that LDR
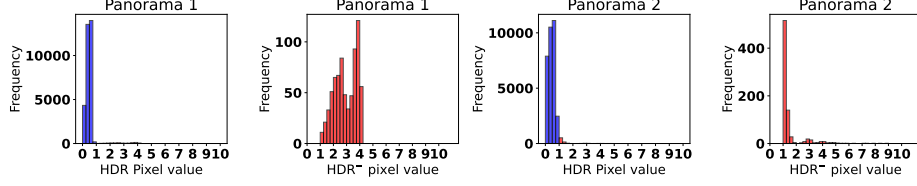
Fig. 3. Pixel distributions of two HDR panoramas. The first and third histograms cover the full range of pixel values. The second and fourth histograms visualize high-intensity values of the first and third histograms. An HDR panorama consists of two components, i.e., low-intensity pixels (blue, from 0 to 1, denoted as LDR) and high-intensity pixels that are out of range LDR pixels (red, larger than 1, denoted as HDR^-).

distributions are easy to generate. Second, HDR^- distributions are dissimilar and cover a few pixels in the high dynamic range (much larger than 1). This implies that HDR^- distributions are difficult to generate and could lead to unstable training of HDR synthesis. If we only use an HDR discriminator to distinguish entire HDR panoramas, the dynamic HDR^- pixels (e.g., >100) largely affect the discriminator and other regions (0∼1) are often ignored through convolution. $D_\prime$ can be regarded as a regularization term to preserve image semantics in LDR regions. We analyze the distributions of 1401 training images. We fit a Gaussian distribution for both LDR and HDR^- of each image, respectively. We find that most of LDR distributions (99.86% LDR images) follow Gaussian distributions and most of HDR^- distributions (70.38% HDR^- images) fail to converge to Gaussian distributions. Note that LDR panoramas play a key role in the synthesis from LDR LFOV images to HDR panoramas in the next section. We expect HDR and LDR images to be well synthesized. It is noteworthy that although HDR^- distributions are difficult to generate, HDR^- and LDR have a strong semantic relation. LDR pixels provide the location and shape contextual information of lighting sources while HDR pixels provide intensity information.

Motivated by the fact that LDR panoramas are easy to generate and they provide location and shape contextual information of lighting for HDR synthesis, we introduce a coupled dual-StyleGAN for joint HDR and LDR panorama synthesis. With the help of LDR panorama synthesis, dual-StyleGAN is able to generate accurate HDR panoramas. Specifically, the dual-StyleGAN takes a noise vector $z \in z$ as input and maps the noise z into a intermediate latent space ω that is much less entangled than the z space, as analyzed in StyleGAN2 [12]. The mapping function $w = f(z)$ consists of eight fully-connected layers with activation layers. The latent code $w \in \omega$ is then fed into the generator G. After that, two transformations $\psi(\cdot)$ and $\phi(\cdot)$ are used to transform the output of G into HDR and LDR panoramas, respectively. That is, $x_{hdr} = \psi(G(w))$ and $x_{ldr} = \phi(G(w))$, as illustrated in **Fig. 2**. The HDR output $\psi(\cdot)$ is an inverse tone-mapping plus clamping (positive pixels) and the LDR output $\phi(\cdot)$ is clampped from 0 to 1. To ease the stable training, we pre-process the the HDR training panoramas by $x^{'} = \alpha x^{1/\gamma}$, where $\gamma = 2.4$ and α is set to 0.5 divided by 50-th per-

centile of all pixel values. We use $clamp(x^{'}, 0, +\infty))$ for HDR and $clamp(x^{'}, 0, 1)$ for LDR. During inference, we perform inverse tone-mapping $clamp(x^{'}, 0, +\infty))$ into x. Other tone-mappings are also suitable for our framework. The idea behind the shared generator network is that HDR and LDR panoramas share the same contextual semantics. Formally, the proposed coupled dual-StyleGAN is formulated as

$$
\begin{aligned}
\min_{G} \max_{D,D^{'}} \mathcal{J}(G, D, D^{'}) = {} & E_{x \backsim P_{data}(x)}[\log(D(x))] \\
& + E_{x \backsim P^{'}_{data}(x)}[\log(D^{'}(x))] \\
& + E_{w \backsim P_{\boldsymbol{w}}}[\log(1 - D(\psi(G(\boldsymbol{w}))))] \\
& + E_{w \backsim P_{\boldsymbol{w}}}[\log(1 - D^{'}(\phi(G(\boldsymbol{w}))))],
\end{aligned}
\tag{1}
$$

where $D(\cdot)$ and $D^{'}(\cdot)$ denote two discriminators for HDR synthesis and LDR synthesis, respectively. In Eq. 1, we optimize G by minimizing the third and fourth terms. We optimize D and $D^{'}$ simultaneously by maximizing the probability of predicting real and fake panoramas. Note that we do not use the same discriminator because the distributions of HDR and LDR panoramas are different, as illustrated in **Fig. 3**.

Focal-Masked GAN inversion for LDR-to-HDR transformation and LFOV-to-panorama completion. After obtaining the coupled dual-StyleGAN, we can perform LDR-to-HDR transformation and LFOV-to-panorama completion. In the coupled dual-StyleGAN, HDR synthesis and LDR synthesis share the same generator G and thus guarantee pixel-level alignment between the LDR-to-HDR transformation. Thanks to the coupled dual-StyleGAN that integrates HDR synthesis and LDR synthesis into a unified framework, we can find a shared latent code of an HDR panorama and an LDR LFOV image in the ω space, as illustrated in **Fig. 2**(left). Given an LDR LFOV image, we map it into a spherical panorama as a masked panorama. We then extend a GAN inversion method [20] into a masked GAN inversion to tailor for our coupled dual-StyleGAN. We find the latent code of the masked panorama in the ω space. We forward the latent code through the HDR branch to obtain an HDR panorama. Let $x_{\text{lfov}}^{\text{ldr}}$ denote a masked panorama of the LDR LFOV image and $\boldsymbol{n}$ denote noise maps of all layers. We first optimize latent code $\boldsymbol{w} \in \omega$ such that synthesized LDR LFOV matches the target LDR LFOV as much as possible, which can be formulated as

$$
\boldsymbol{w}_*, \boldsymbol{n}_* = \underset{\boldsymbol{w}, \boldsymbol{n}}{\operatorname{argmin}} \, \mathcal{L}_{\text{LPIPS}}(M \odot \phi(G(\boldsymbol{w}, \boldsymbol{n}; \theta)), x_{\text{lfov}}^{\text{ldr}}) + \lambda_{\boldsymbol{n}} \mathcal{L}_{\boldsymbol{n}}(\boldsymbol{n}),
\tag{2}
$$

where $\mathcal{L}_{\text{LPIPS}}$ represents the VGGNet based Perceptual loss [10] for distance measure and $\mathcal{L}_{\boldsymbol{n}}(\boldsymbol{n})$ represents a noise regularization term to constrain noise maps $\boldsymbol{n}$ and $\lambda_{\boldsymbol{n}}$ controls its importance. We re-use the symbol G with latent code $\boldsymbol{w} \in \omega$, noise maps $\boldsymbol{n}$ and weights θ as input. The variable M denotes a mask that preserves the pixels corresponding to the LFOV image $x_{\text{lfov}}^{\text{ldr}}$. Because it is hard to directly reconstruct the target LDR LFOV image by optimizing $\boldsymbol{w}$

and $\boldsymbol{n}$, we then fine-tune the weights θ of G by

$$
\begin{aligned}
\mathcal{L}_G(\theta) =& \mathcal{L}_{\mathrm{LPIPS}}(M \odot \phi(G(\boldsymbol{w}_*, \boldsymbol{n}_*; \theta)), x_{\mathrm{lfov}}^{\mathrm{ldr}}) \\
& + \lambda_{L2}^R \mathcal{L}_{L2}(M \odot \phi(G(\boldsymbol{w}_*, \boldsymbol{n}_*; \theta)), x_{\mathrm{lfov}}^{\mathrm{ldr}}),
\end{aligned}
\tag{3}
$$

where $\mathcal{L}_G$ represents the similarity loss between a masked generated LDR image and the LFOV HDR image $x_{\mathrm{lfov}}^{\mathrm{ldr}}$. We use the Mean-Squared-Error loss and the VGGNet based Perceptual loss for distance measure. To further improve visual quality, we also use a regularization term, which is given by

$$
\begin{aligned}
\mathcal{L}_R(\theta) =& \mathcal{L}_{\mathrm{LPIPS}}(\phi(G(\boldsymbol{w}_r, \boldsymbol{n}_*; \theta)), \phi(G(\boldsymbol{w}_r, \boldsymbol{n}_*; \theta_0))) \\
& + \lambda_{L2}^{R'} \mathcal{L}_{L2}^R(\phi(G(\boldsymbol{w}_r, \boldsymbol{n}_*; \theta)), \phi(G(\boldsymbol{w}_r, \boldsymbol{n}_*; \theta_0))),
\end{aligned}
\tag{4}
$$

where $\boldsymbol{w}_r = \boldsymbol{w}_* + \alpha(\boldsymbol{w}_z - \boldsymbol{w}_*)/\|\boldsymbol{w}_z - \boldsymbol{w}_*\|_2$ and $\boldsymbol{w}_z$ is a latent code that is generated by a random $\boldsymbol{z}$. The θ_0 is the old weights of G before finetuning. The α is a interpolation parameter. Finally, we combine Equations 3 and 4, and fine-tune G by

$$
\theta_* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}_G(\theta) + \eta \mathcal{L}_R(\theta).
\tag{5}
$$

Simply extending the GAN inversion method [20] into a masked GAN inversion for the coupled dual-StyleGAN cannot guarantee that the pixels of lights in LDR panoramas will be transformed into high-intensity HDR^-. Strong lighting sources are the main factors for rendering environment, but high-intensity pixels are much less than low-intensity pixels, as shown in **Fig. 3**. Without any constraint, the model tends to fit the major LDR pixels and ignore the few HDR^- pixels. Therefore, we propose a focal-masked to highlight the strong lighting sources. In our setting, we highlight top 10% strongest pixels. Given an LDR LFOV image $x_{\mathrm{lfov}}^{\mathrm{ldr}}$, we compute the top 10% strongest pixels and obtain a focal mask, which denoted as $FOCAL(x_{\mathrm{lfov}}^{\mathrm{ldr}})$. We re-write Equation 2 as

$$
\begin{aligned}
\boldsymbol{w}_*, \boldsymbol{n}_* =& \underset{\boldsymbol{w}, \boldsymbol{n}}{\operatorname{argmin}}\, \mathcal{L}_{\mathrm{LPIPS}}(M \odot \phi(G(\boldsymbol{w}, \boldsymbol{n}; \theta)), x_{\mathrm{lfov}}^{\mathrm{ldr}}) + \lambda_{\boldsymbol{n}} \mathcal{L}_{\boldsymbol{n}}(\boldsymbol{n}) \\
& + \beta_{L2} \mathcal{L}_{L2}(FOCAL(\phi(G(\boldsymbol{w}, \boldsymbol{n}; \theta))), FOCAL(x_{\mathrm{lfov}}^{\mathrm{ldr}})).
\end{aligned}
\tag{6}
$$

Note that $FOCAL(\phi(G(\boldsymbol{w}, \boldsymbol{n}; \theta)))$ is consistent with $FOCAL(x_{\mathrm{lfov}}^{\mathrm{ldr}})$ in Equation 6. Similarly, we re-write Equation 3 as

$$
\begin{aligned}
\mathcal{L}_G(\theta) =& \mathcal{L}_{\mathrm{LPIPS}}(M \odot \phi(G(\boldsymbol{w}_*, \boldsymbol{n}_*; \theta)), x_{\mathrm{lfov}}^{\mathrm{ldr}}) \\
& + \lambda_{L2}^R \mathcal{L}_{L2}(M \odot \phi(G(\boldsymbol{w}_*, \boldsymbol{n}_*; \theta)), x_{\mathrm{lfov}}^{\mathrm{ldr}}) \\
& + \beta_{L2} \mathcal{L}_{L2}(FOCAL(\phi(G(\boldsymbol{w}_*, \boldsymbol{n}; \theta))), FOCAL(\boldsymbol{w}_{\mathrm{lfov}}^{\mathrm{ldr}})).
\end{aligned}
\tag{7}
$$

3.3   Lighting Editing

Lighting estimation provides a way to replicate real-world lighting conditions of a scene from a single LDR LFOV image and can be used to re-render inserted objects. In some occasions, however, the estimated lighting conditions do not

meet our design/modeling requirements. For example, we would like to turn on an indoor light and "turn off" the light from windows at night given daytime HDR panoramas. Although light sources are located in a local area, it affects the global panorama lighting. Existing approaches do not consider this problem. Here, we introduce a lighting editing method to control panorama lighting through exploiting the powerful StyleLight.

In general, lighting editing includes three operations, i.e, adding a new light, removing a light, and controlling lighting intensity. In essence, the first two operations can be accomplished by the third one. In light estimation, we obtain a latent code w_0 that recovers an HDR panorama from an LDR LFOV image. Here we assume that $\boldsymbol{w}_0$ is known. Our goal is to find a editable direction of $\boldsymbol{w}_0$ to adjust the lighting intensity such that the lighting intensity at a candidate location changes while the geometry structure of the scene is preserved. Given a bounding box of a candidate lighting location as a mask M, we introduce a structure-preserved lighting editing loss to find the editable direction of $\boldsymbol{w}$, which is formulated as

$$
\begin{aligned}
\boldsymbol{w}_*, \boldsymbol{n}_* = \underset{\boldsymbol{w}, \boldsymbol{n}}{\operatorname{argmin}}\, & \mathcal{L}_{\mathrm{LPIPS}}((1 - M) \odot \phi(G(\boldsymbol{w}, \boldsymbol{n}; \theta)), (1 - M) \odot \phi(G(\boldsymbol{w}_0, \boldsymbol{n}; \theta))) \\
& + \mathcal{L}_{L2}((1 - M) \odot \phi(G(\boldsymbol{w}, \boldsymbol{n})), (1 - M) \odot \phi(G(\boldsymbol{w}_0, \boldsymbol{n}; \theta))) \\
& + \frac{\delta}{N} \sum M \odot \phi(G(\boldsymbol{w}, \boldsymbol{n}; \theta)),
\end{aligned}
$$

(8)

where δ denotes a lighting adjusting factor. When δ is positive, the lighting intensity decreases; when δ is negative, the lighting intensity increases. N is the number of masked pixels. The third term aims to adjust lighting intensity given a bounding box, and the first and second terms preserve other regions.

4   Experiments

Datasets and experimental settings. We evaluate our StyleLight model on the Laval Indoor HDR dataset [5] that contains 2,100 HDR panoramas. Following the train/test split of [5], we use 1,719 panoramas for training and 289 panoramas for testing. For the test set, we crop one HDR LFOV image from each panorama and transform it into an LDR LFOV image by tone mapping [32]. Since lighting estimation is a challenging task, some compared methods use extra annotations. For example, Gardner2017 [5] used annotated lighting sources of SUN360 [30] to learn to predict the locations and shapes of lighting sources. Gardner2019 [4] manually annotated the Laval Indoor dataset using EnvyDepth [2] and used per-pixel depth maps to improve lighting estimation.

Implementation details. We use StyleGAN2 [12] as the backbone. The output of the generator is a 256×512 image. We resize it into 128×256 for evaluation, which is consistent with other methods. The resolution of feature maps begins with a 4×8 constant map and increases by $\times 2$ upsampling each block, yielding 256×512 patch-based feature maps. In the full

training of StyleLight, 4000k real images (kimg) are shown to the discriminators. We train StyleLight with 8 v100 GPUs for 7 hours. The batch size is set to 32. We use Adam optimizer for the generators and two discriminators with a learning rate of 0.0025 and set $\beta_1 = 0$, $\beta_2 = 0.99$.

Evaluation metrics. Similar to Gardner2019 [4] and EMlighting [32], we use three spheres with different materials for evaluation, including mirror, matte silver, and gray diffuse, as illustrated in **Fig. 4**. We render the three spheres by ground-truth HDR panoramas and the estimated illumination maps. Similar to [32], we use Root Mean Square Error (RMSE), scale-invariant RMSE (si-RMSE), and Angular Error for evaluation.

Fig. 4. Three spheres with different materials: mirrored silver, matte silver, and diffuse grey.

RMSE and si-RMSE measure pixel-level errors of lighting intensity. Angular Error measures the ratio of R, G, and B components. Different from [32], we compute the metrics on the three rendered spheres and ignore the area of the rendered black background, which is invariant to the area of the background. Because there are no public codes for evaluation, we design our own hyper-parameters for the three spheres. We also use Fréchet Inception Distance (FID) [8] to measure the distribution similarity between the ground-truth and the estimated illumination maps.

4.1   Quantitative Evaluation

We compare StyleLight with three representative state-of-the-art illumination estimation methods, including a generation-based method [32] and two regression-based methods [5,4]. Note that Gardner2017 [5] used extra lighting annotations and Gardner2019 [4] used extra depth annotations for training. StyleLight and EMlight [32] models do not require any extra annotation. For each method, we render three typical spheres (**Fig. 4**) by the predicted illumination maps and ground-truth HDR panoramas on the test set and evaluate RMSE, si-RMSE, and Angular Error. We also evaluate FID to measure the distribution distance of ground-truth and predicted illumination maps.

As shown in **Table 1**, our StyleLight outperforms the compared methods on all metrics and materials. Among them, Gardner2017 and Gardner2019 focus on strong lighting sources and ignore low-intensity lighting, so they fail to achieve a desired RMSE and si-RMSE on the mirror sphere and do not preserve components of colors. EMlight first regresses a Gaussian map and fuses it with the LFOV image for illumination generation. The drawback is that the multi-step optimization could lead to sub-optimal performance. Since EMlight contains regression learning, it is difficult to train. EMlight[1] claimed that the model has to be trained on subsets of 100, 1000, 2500,... and the full set gradually and could

---

[1] https://github.com/fnzhan/Illumination-Estimation

Fig. 5. Visual comparisons of our method with several SOTA methods.

collapse occasionally during training. We also see that our StyleLight largely outperforms other methods on FID, indicating that our predicted lighting distribution is closest to ground truth on a wide range of lighting. The results also suggest that StyleLight can synthesize plausible scenes and can achieve a promising result when rendering mirror materials. Our StyleLight also provides controllable lighting editing.

Table 1. Comparison with previous work. Gardner2017 and Gardner2019 are regression-based methods. EMlight is a generation-based method. Annotation denotes required manual annotations. The evaluation metrics include Angular Error, RMSE, si-RMSE and FID. M, S, and D denote mirror, matte silver, and diffuse material spheres, respectively. ↓ denotes that smaller is better.

Methods	Gardner2017 [5]			Gardner2019 [4]			EMlight [32]			StyleLight		
	M	S	D	M	S	D	M	S	D	M	S	D
Annotation	lighting sources			depth			No			No		
Angular Error↓	6.68	6.03	5.47	4.63	3.91	3.24	4.67	3.47	2.59	**4.30**	**2.96**	**2.41**
RMSE↓	0.70	0.38	0.19	0.85	0.49	-	0.59	0.33	0.17	**0.56**	**0.30**	**0.15**
si-RMSE↓	0.65	0.35	0.15	0.70	0.40	0.14	0.58	0.31	0.12	**0.55**	**0.29**	**0.11**
FID↓	307.5			344.1			263.8			**137.7**		

4.2 Qualitative Evaluation

In this section, we provide visual results to show the effectiveness of the proposed StyleLight model, including visual comparison with state-of-the-art lighting estimation methods and lighting editing.

Visual comparison with the state-of-the-art lighting estimation methods. As depicted in **Fig. 5**, the predicted illumination maps by StyleLight are similar to the ground truth, and the proposed method obtains high-quality panoramas. Compared with the ground truth, Gardner2017 and Gardner2019
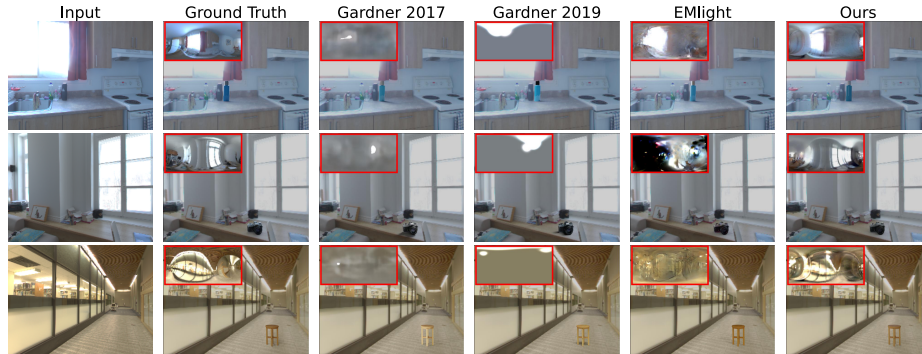
Fig. 6. Visual comparisons on object insertion and lighting. For each input LDR LFOV image, the four different methods estimate illumination maps (at the top-left of LFOV images) for rendering newly inserted objects.

highlight strong lighting sources, which are not suitable to render mirror materials. EMlight considers both low-intensity and high-intensity lighting, but it obtains blurred images and it is difficult to train because of the unstable regression. Our StyleLight can synthesize panoramas of better quality, which can be further used for mirror rendering and lighting editing.

Visual comparison with the state of the art on object insertion and lighting. We further compare StyleLight with state-of-the-art methods on object insertion and lighting. The predicted HDR panorama is a sphere. The object is placed at the center of the panorama and a warping operator [5] is used when inserted objects are not at the center of a panorama. Finally, image-based rendering is applied. The object insertion technique is similar to EMlight. The four different methods estimate illumination maps for rendering newly inserted objects. As shown in **Fig. 6**, compared with nearby objects in LFOV images, the inserted objects by Gardner2017 and Gardner2019 tend to lose low-intensity lighting information and generate inaccurate shadows. Thanks to the unique design of coupled dual-StyleGAN that integrates both HDR and LDR panoramas into a unified framework, StyleLight achieves a better visual outputs.

Lighting editing with StyleLight. As shown in **Fig. 7**, given a panorama (the first column), we can remove lights in the second column and add new lights in the third column with the method presented in Sec. 3.3.

Effectiveness of the Coupled Dual-StyleGAN structure. We conduct an ablation study to analyze the effectiveness of the coupled dual-StyleGAN framework. As shown in **Fig. 8**, the first row shows the quality of StyleLight that uses two discriminators to learn LDR and HDR distributions. The second row shows the LDR panorama quality by using one discriminator to train HDR synthesis (denoted as HDR-only) and the LDR panoramas are computed by tone mapping. It is observed that LDR panoramas from the 'HDR-only' suffer from severe distortion. This ablation demonstrates the effectiveness of the coupled dual-StyleGAN structure.

Fig. 7. Lighting editing with StyleLight. For each original panorama (Row 1), Style-Light removes a light (red box, Row 2) and then adds a new light (green box, Row 3).
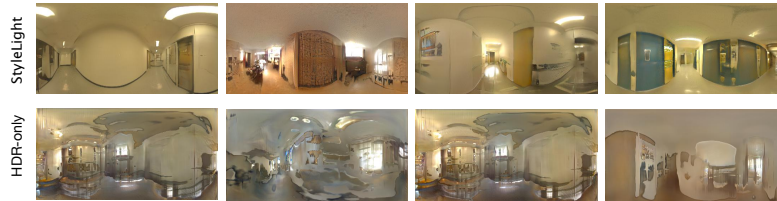

Fig. 8. Effectiveness of the coupled dual-StyleGAN structure. The dual-StyleGAN generates both HDR and LDR panoramas. HDR-only directly generates HDR panoramas.

Effectiveness of the focal-masked loss function. We conduct an ablation study for the new loss in Eq. 7. We analyze RMSE and si-RMSE as shown in Table 2. The existing GAN inversion is unavailable to LFOV images. We first develop a masked version and then extend it to a focal-masked version. The table shows the improvement of the proposed focal-masked loss function.

User Study. We conduct a user study to analyze predicted illumination maps. **Fig. 10** shows two examples of the user study. For each exam-

Fig. 9. User preference.

ple, we insert an object into an image. We render the inserted object with illumination maps predicted by four methods. Users were asked to pick the images that were more closer to the images rendered by ground-truth HDR panorama. We rendered different objects under 10 scenes, and 100 users participated in the study. Results are shown in **Fig. 9**. We can observe that 52.1% prefer our rendered results, which significantly outperform other methods.
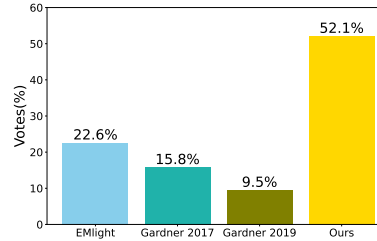
Table 2. Effectiveness of the focal-masked loss function. M, S, and D denote mirror, matte silver, and diffuse material spheres, respectively. ↓ denotes that smaller is better.

metrics	baseline [20]	masked			focal-masked		
	N/A	M	S	D	M	S	D
RMSE↓	N/A	0.57	0.31	0.16	**0.56**	**0.30**	**0.15**
si-RMSE↓	N/A	0.56	0.30	0.12	**0.55**	**0.29**	**0.11**

Ground Truth   (A)   (B)   (C)   (D)

Fig. 10. For each row, the inserted objects are rendered by HDR illumination maps estimated by four methods. Which one do you think is more closer to the object rendered by the ground-truth HDR illumination map?

First row: (A)EMlight, (B)Gardner 2017, (C)Ours, (D)Gardner 2019; Second row: (A)Gardner 2019, (B)Ours, (C)Gardner 2017, (D)EMlight

5   Conclusion

We have presented a coupled dual-StyleGAN panorama synthesis network (StyleLight) for lighting estimation and editing. StyleLight offers high-quality lighting estimation that generates the full indoor panorama lighting from a single LFOV image. With the trained StyleLight, we propose a structure-preserved lighting editing method that enables flexible lighting editing. Extensive experiments demonstrate that StyleLight achieves superior performance over state-of-the-art methods on indoor lighting estimation and also enables promising lighting editing on indoor HDR panoramas.

Limitation. StyleLight only considers indoor scenes in this paper. When applied to outdoor scenes, StyleLight needs a new set of hyper-parameters because lighting sources (e.g., area and intensity) are different between outdoor scenes and indoor scenes. An intuitive idea is to develop a generalizable method that adaptively tailors for various scenes, which is left to future research.

References

1. Abdal, R., Zhu, P., Mitra, N.J., Wonka, P.: Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. ACM Transactions on Graphics (TOG) **40**(3), 1–21 (2021) 4

2. Banterle, F., Callieri, M., Dellepiane, M., Corsini, M., Pellacini, F., Scopigno, R.: Envydepth: An interface for recovering local natural illumination from environment maps. In: Computer Graphics Forum. vol. 32, pp. 411–420. Wiley Online Library (2013) 9

3. Barron, J.T., Malik, J.: Intrinsic scene properties from a single rgb-d image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 17–24 (2013) 3

4. Gardner, M.A., Hold-Geoffroy, Y., Sunkavalli, K., Gagné, C., Lalonde, J.F.: Deep parametric indoor lighting estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7175–7183 (2019) 2, 3, 4, 9, 10, 11

5. Gardner, M.A., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C., Lalonde, J.F.: Learning to predict indoor illumination from a single image. arXiv preprint arXiv:1704.00090 (2017) 2, 3, 4, 9, 10, 11, 12

6. Garon, M., Sunkavalli, K., Hadap, S., Carr, N., Lalonde, J.F.: Fast spatially-varying indoor lighting estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6908–6917 (2019) 2, 3, 4

7. Härkönen, E., Hertzmann, A., Lehtinen, J., Paris, S.: Ganspace: Discovering interpretable gan controls. Advances in Neural Information Processing Systems **33**, 9841–9850 (2020) 4

8. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017) 10

9. Hold-Geoffroy, Y., Sunkavalli, K., Hadap, S., Gambaretto, E., Lalonde, J.F.: Deep outdoor illumination estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7312–7321 (2017) 2, 3

10. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016) 7

11. Kanamori, Y., Endo, Y.: Relighting humans: occlusion-aware inverse rendering for full-body human images. arXiv preprint arXiv:1908.02714 (2019) 2

12. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8110–8119 (2020) 6, 9

13. Karsch, K., Hedau, V., Forsyth, D., Hoiem, D.: Rendering synthetic objects into legacy photographs. ACM Transactions on Graphics (TOG) **30**(6), 1–12 (2011) 3

14. LeGendre, C., Ma, W.C., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J., Debevec, P.: Deeplight: Learning illumination for unconstrained mobile mixed reality. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5918–5928 (2019) 2, 3, 4

15. Li, H., Liu, J., Bai, Y., Wang, H., Mueller, K.: Transforming the latent space of stylegan for real face editing. arXiv preprint arXiv:2105.14230 (2021) 4

16. Lombardi, S., Nishino, K.: Reflectance and illumination recovery in the wild. IEEE transactions on pattern analysis and machine intelligence **38**(1), 129–141 (2015) 3

17. Lopez-Moreno, J., Hadap, S., Reinhard, E., Gutierrez, D.: Compositing images through light source detection. Computers & Graphics **34**(6), 698–707 (2010) 3

18. Maier, R., Kim, K., Cremers, D., Kautz, J., Nießner, M.: Intrinsic3d: High-quality 3d reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In: Proceedings of the IEEE international conference on computer vision. pp. 3114–3122 (2017) 3

19. Pandey, R., Escolano, S.O., Legendre, C., Haene, C., Bouaziz, S., Rhemann, C., Debevec, P., Fanello, S.: Total relighting: learning to relight portraits for background replacement. ACM Transactions on Graphics (TOG) **40**(4), 1–21 (2021) 2

20. Roich, D., Mokady, R., Bermano, A.H., Cohen-Or, D.: Pivotal tuning for latent-based editing of real images. arXiv preprint arXiv:2106.05744 (2021) 4, 7, 8

21. Shen, Y., Gu, J., Tang, X., Zhou, B.: Interpreting the latent space of gans for semantic face editing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9243–9252 (2020) 4

22. Shen, Y., Zhou, B.: Closed-form factorization of latent semantics in gans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1532–1540 (2021) 4

23. Song, S., Funkhouser, T.: Neural illumination: Lighting prediction for indoor environments. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6918–6926 (2019) 2, 4

24. Srinivasan, P.P., Mildenhall, B., Tancik, M., Barron, J.T., Tucker, R., Snavely, N.: Lighthouse: Predicting lighting volumes for spatially-coherent illumination. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8080–8089 (2020) 2, 4

25. Sun, T., Barron, J.T., Tsai, Y.T., Xu, Z., Yu, X., Fyffe, G., Rhemann, C., Busch, J., Debevec, P.E., Ramamoorthi, R.: Single image portrait relighting. ACM Trans. Graph. **38**(4), 79–1 (2019) 2

26. Tewari, A., Elgharib, M., Bharaj, G., Bernard, F., Seidel, H.P., Pérez, P., Zollhofer, M., Theobalt, C.: Stylerig: Rigging stylegan for 3d control over portrait images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6142–6151 (2020) 4

27. Valgaerts, L., Wu, C., Bruhn, A., Seidel, H.P., Theobalt, C.: Lightweight binocular facial performance capture under uncontrolled lighting. ACM Trans. Graph. **31**(6), 187–1 (2012) 3

28. Wang, T., Zhang, Y., Fan, Y., Wang, J., Chen, Q.: High-fidelity gan inversion for image attribute editing. arXiv preprint arXiv:2109.06590 (2021) 4

29. Wu, C., Wilburn, B., Matsushita, Y., Theobalt, C.: High-quality shape from multiview stereo and shading under general illumination. In: CVPR 2011. pp. 969–976. IEEE (2011) 3

30. Xiao, J., Ehinger, K.A., Oliva, A., Torralba, A.: Recognizing scene viewpoint using panoramic place representation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2695–2702. IEEE (2012) 9

31. Zhan, F., Zhang, C., Hu, W., Lu, S., Ma, F., Xie, X., Shao, L.: Sparse needlets for lighting estimation with spherical transport loss. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12830–12839 (2021) 2, 3

32. Zhan, F., Zhang, C., Yu, Y., Chang, Y., Lu, S., Ma, F., Xie, X.: Emlight: Lighting estimation via spherical distribution approximation. arXiv preprint arXiv:2012.11116 **2** (2020) 2, 3, 4, 9, 10, 11

33. Zhuang, P., Koyejo, O., Schwing, A.G.: Enjoy your editing: Controllable gans for image editing via latent space navigation. arXiv preprint arXiv:2102.01187 (2021) 4