StyLitGAN: Image-based Relighting via Latent Control

Supplementary Material

A. Losses Ablation

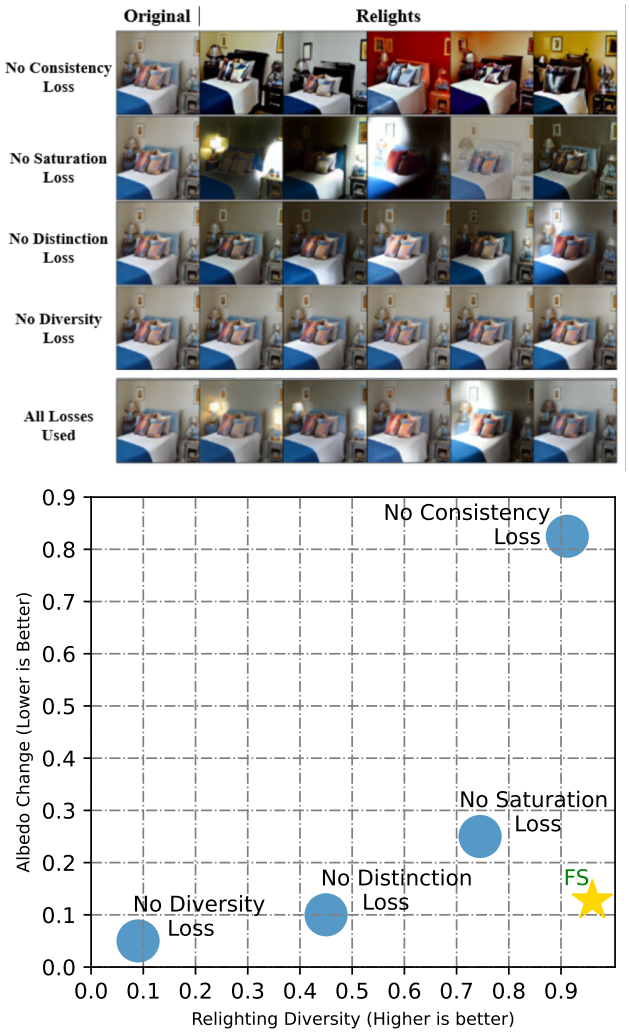

Figure 13. Ablation study of all required losses. Without consistency loss, there is a significant change in the scene's albedo. Without a saturation penalty, the relightings are either too saturated or undersaturated. Without distinction loss, the relightings are reasonable but not diverse, and without diversity loss the relights are the same as the original image. However, when we combine all our losses (gold star), we can produce realistic, plausible, and diverse relights of the same scene.

B. Choice of Decomposition

The choice of decomposition matters for relighting without change in geometry and albedo. The best-performing admissible decomposition from our experiments has been a variant decomposition that models fine edges in albedo rather than in the shading field. As we apply diversity loss on the shading field; it is practical to not model geometry (fine edges; normals). Otherwise, undesirable geometry shifts may occur, as demonstrated in the videos on our project page. Representative examples of our modified decomposition can be found in Fig. 16. Furthermore, we observed that incorporating gloss as an additional component enhances the identification of light sources and facilitates more realistic lighting alterations while maintaining diverse appearance changes.

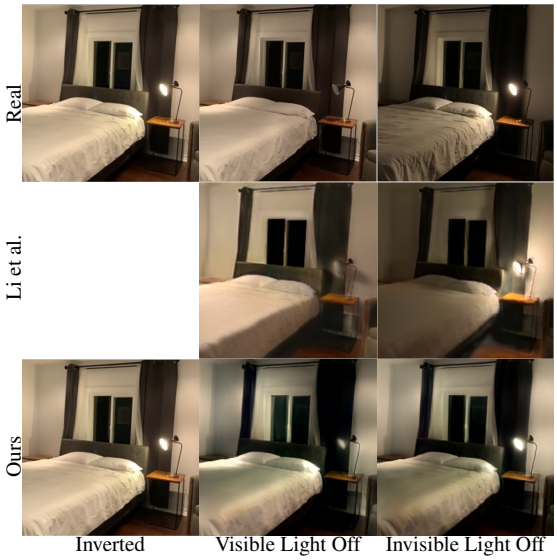

Figure 14. With an appropriate inversion method [5], we can relight real scenes. **Top row:** shows three images of the same scene in different lightings, obtained from [26]. **Middle row:** shows relightings of the first image, obtained by Li *et al.* by inverse rendering, changing luminaire parameters, then forward rendering [26]. **Bottom left:** image generated by passing latent variables from our inversion of the top left image through StyleGAN. **Bottom center** and **bottom right** show relightings obtained by adding a relighting direction to these latent variables. Recall the relighting directions are *independent of image*. The directions were selected by hand to correspond to the top center, right respectively; directions are Relit - 5 (visible light off) and Relit - 2 (invisible light off) from Figure 1. Note: our relights compare well to real images; our relights do not ring on fine edges (eg the lamp); our relights preserve high spatial frequencies in the image and do not require CGI, physical rendering, or light source annotation.

C. Albedo Scores and Analogy with CCA

The relevant maximum can be computed by analogy with canonical correlation analysis. Reshape each color component of the patch into an $(3MN)$ vector $\mathbf{P}$ with components $p_{w(i,j),k}$. Construct (MN) basis vectors $x_{w(i,j)} = i$ and

Figure 15. Simultaneous Joint Relighting and Resurfacing. **Columns** show different relights for a fixed resurfacing, and **rows** show different resurfacing for a fixed relight. The interactions between relighting and resurfacing are largely disentangled.

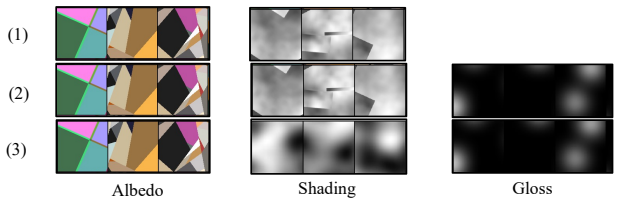

Figure 16. We compare three different decomposition models derived from [14]. We find decomposition (3) to provide significantly better results with small geometry shifts after relighting. In general, the directions obtained with this decomposition are admissible. The other two decompositions resulted in relighting directions with large albedo changes. The major differences between these decomposition methods are (a) we use an additional gloss component in (2) and (3) and (b) we also assume shading to be smooth and have all the high-frequency fine-edges information in the albedo to preserve geometric details without noticeable shifts after relighting in (3).

$y_{w(i,j)} = j$. Now construct the $(3MN) \times 3$ matrix $\mathcal{M}(\mathbf{p})$ such that

$$\mathbf{P}'(q_x, q_y, q_c) = \mathcal{M}(\mathbf{p})\mathbf{q}_c = \begin{bmatrix} x_1 p_{1,1} & y_1 p_{1,1} & p_{1,1} \\ \cdots & \cdots & \cdots \\ x_1 p_{1,2} & y_1 p_{1,2} & p_{1,2} \\ \cdots & \cdots & \cdots \\ x_1 p_{1,3} & y_1 p_{1,3} & p_{1,3} \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} q_x \\ q_y \\ q_c \end{bmatrix} \tag{7}$$

so that $d_a(\mathbf{p}, \mathbf{q})$

$$d_a(\mathbf{p}, \mathbf{q}) = 1 - \frac{\max\limits_{\mathbf{p}_c, \mathbf{q}_c} (\mathbf{q}_c \mathbf{M}'(\mathbf{p}) \mathbf{M}(\mathbf{q}) \mathbf{p}_c)}{\sqrt{(\mathbf{q}_c \mathbf{M}'(\mathbf{p}) \mathbf{M}(\mathbf{p}) \mathbf{q}_c)(\mathbf{p}_c \mathbf{M}'(\mathbf{q}) \mathbf{M}(\mathbf{q}) \mathbf{p}_c)}} \tag{8}$$

$$1 - \frac{\max\limits_{\mathbf{p}_c, \mathbf{q}_c} \mathbf{q}_c \Sigma_{xy} \mathbf{p}_c}{\sqrt{(\mathbf{q}_c \Sigma_{xx} \mathbf{q}_c)(\mathbf{p}_c \Sigma_{yy} \mathbf{p}_c)}}. \tag{9}$$

Standard results then yield that

$$d_a(\mathbf{p}, \mathbf{q}) = 1 - \sqrt{\lambda_x} \tag{10}$$

where λ_x is the largest eigenvalue of

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}^T \tag{11}$$

D. Additional Qualitative Examples and Movies

For better visualization, we provide interpolation movies on our project page. We use a simple linear interpolation between distinct relighting directions that we found. The movies show smooth continuous lighting changes with very small local geometry changes.

E. Other Experimental Details

For our Model 14 relighting, we employ the following λ coefficients: $\lambda_{const} = 750, \lambda_{per} = 0.1, \lambda_{dist} = 1, \lambda_{deco} = 0.01$. We also apply distinct λ_{div} values for different categories. For bedrooms, we use $\lambda_{div} = 0.125$; for kitchens, dining, and living rooms, $\lambda_{div} = 0.25$; for conference rooms, $\lambda_{div} = 0.4$; for faces, $\lambda_{div} = 0.1$; and for churches, $\lambda_{div} = 0.5$. It is important to note that these coefficients pertain to the selected model with albedo, shading, and gloss decomposition, and fine edges are modeled in albedo, as previously discussed.

For our recoloring or resurfacing, we use the following λ coefficients: $\lambda_{const} = 1000, \lambda_{per} = 0.1, \lambda_{dist} = 1, \lambda_{deco} = 0$. We also employ different λ_{div} values for various categories. For bedrooms, we use $\lambda_{div} = 0.3$; for kitchens, dining, and living rooms, $\lambda_{div} = 1$; for conference rooms, $\lambda_{div} = 0.5$; for faces, $\lambda_{div} = 0.2$; and for churches, $\lambda_{div} = 0.6$.

For all categories, we employ 2000 search iterations; however, effective relighting directions become apparent after only a few hundred iterations. In addition, we utilize the Adam optimizer for searching the latent directions with a learning rate of 0.001 and for updating the classifier with a learning rate of 0.0001.

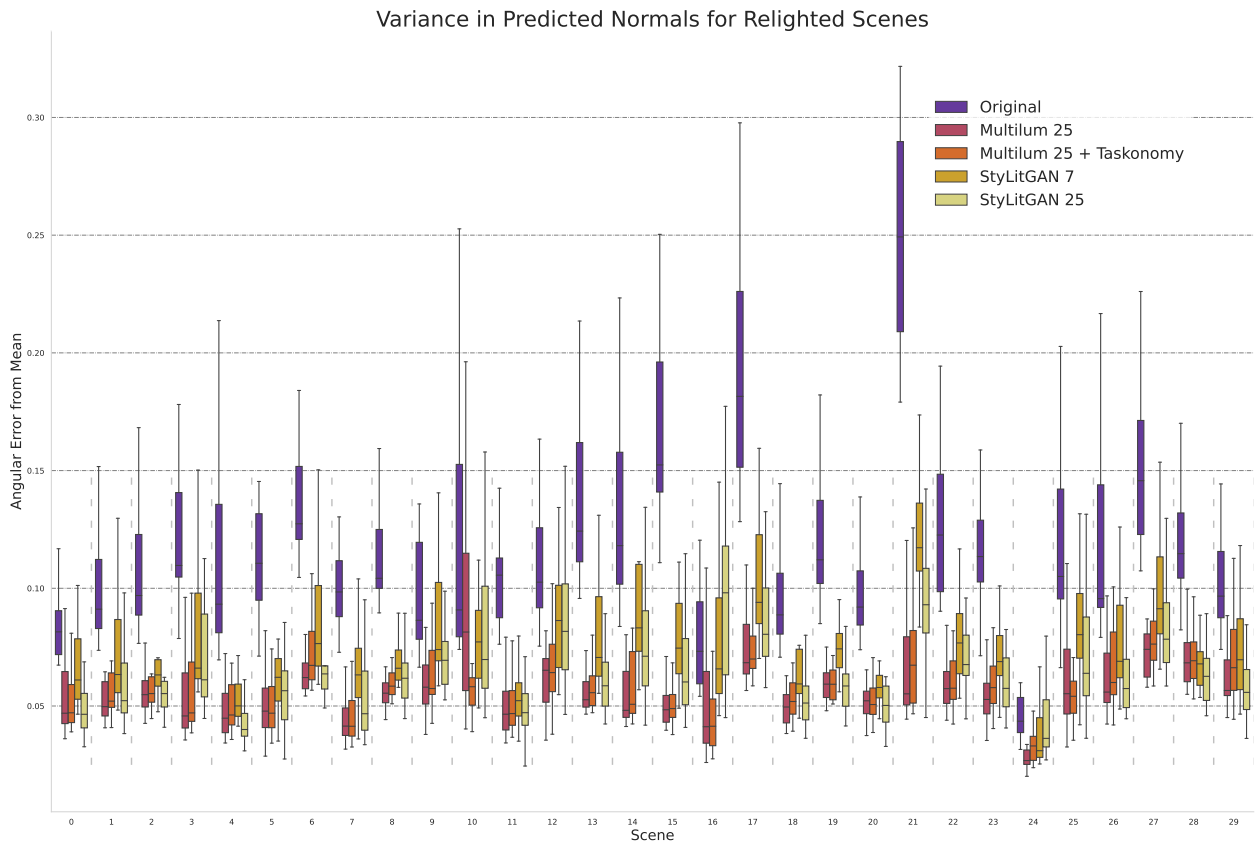

Figure 17. Scene-specific performance of a finetuned normal predictor on full Multilum test set, a comprehensive version of Figure 11. Finetuning with real multi-illumination images and synthetic StyLitGAN relightings both reduce variance in predicted normals.